CLASSIFICATION BY NEURAL NETWORK AND STATISTICAL MODELS IN TANDEM: DOES INTEGRATION ENHANCE PERFORMANCE?

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

David Mitchell, B.S., M.S.

Denton, Texas

December, 1998

Mitchell, David, Classification by neural network and statistical models in tandem: Does integration enhance performance?. Doctor of Philosophy (Management Science), December, 1998, 171 pp., 64 tables, 5 figures, references, 152 titles.

Much research and practical effort has gone into the classification of various natural phenomena. Classification is, after all, a requisite step in the process of explaining the complex relationships among such phenomena. A complete understanding of the efficacy of each of the myriad statistical and non-statistical classification models across all possible problem domains does not currently exist.

The current research uses a live transportation planning data set from Dallas/Fort Worth and five simulated data sets to compare the classificatory performance of six classification models: logit, linear discriminant analysis, quadratic discriminant analysis, backpropagation neural network, modular neural network, and radial basis function neural network. The five simulated data sets are all two-group problems with two independent variables. The two groups range from both being bivariate normal and linearly separable to being contaminated such that they are neither linearly nor quadratically separable.

In addition, a new method for potentially enhancing the performance of the models is introduced. This approach entails using outputs from matched statistical and neural network model pairs in an iterative fashion to increase the classificatory performance of the combined model pair over the performance of the separate models.

The study also examines the impact of proportional mix of observations on the relative classificatory performance of the six models.

The results indicate that the included classification models can be improved upon absolutely, and often significantly, by incorporating the results of a complementary model in the model estimation process. The results also indicate that among the statistical and neural network models included, the quadratic discriminant analysis and modular neural network models, respectively, tend to be the most efficacious. This is true whether they are used alone or are paired with complementary models using the iterative procedure. Lastly, the neural network models appear to be less sensitive than the statistical models to changes in the proportional mix of observations in each group.

CLASSIFICATION BY NEURAL NETWORK AND STATISTICAL MODELS IN TANDEM: DOES INTEGRATION ENHANCE PERFORMANCE?

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

David Mitchell, B.S., M.S.
Denton, Texas

December, 1998

ACKOWLEDGMENTS

I wish to express a debt of gratitude to my committee chair, Dr. Robert Pavur, for his academic guidance, extensive support, and friendship prior to and during this effort. Additionally, I appreciate the effort and constructive comments of my committee, Drs. Maliyakal Jayakumar, Alan Kvanli, and Ian Parberry.

I wish also to extend my thanks to my family for their support and love throughout my life and their understanding and patience during the undertaking of my doctoral studies and this dissertation.

I acknowledge the University of North Texas for the environment and resources that made this work possible.

Finally, special thanks to my best friend and confidant, Tanya Vazquez, for her untiring support, patience, and many kindnesses throughout the execution of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Much research and practical effort has gone into the classification of various natural phenomena. Classification is, after all, a requisite step in the process of explaining the complex relationships among such phenomena. Many methods have been devised to assist in this classification effort. These range from the development of trained human experts (e.g., medical doctors trained to classify sets of symptoms into disease categories) to a variety of statistical and computer-based methods such as discriminant analysis, logit, decision trees, expert systems, genetic algorithms, and neural networks. A complete understanding of the efficacy of each of these myriad classification methods across all possible problem domains does not currently exist. This is often true even within problem domains for which a particular classification method has become the *de facto* standard, such as the use of logit in the transportation mode choice domain. Much work, therefore, remains in establishing the utility of and understanding of these sundry methods in each new and existing problem domain for which they are potential solutions.

The current research uses a live transportation planning data set from Dallas/Fort Worth and five simulated data sets to compare the classificatory performance of six classification models: logit (LOG), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), backpropagation neural network (BPNN), modular neural network (MNN), and radial basis function neural network (RBFNN). In addition, a new method for potentially enhancing the performance of the models is introduced. This approach entails using outputs from statistical and neural network model pairs in an iterative fashion to increase the classificatory performance of the combined model pair over the performance of the separate models. The study also examines the impact of training set configuration on the relative classificatory performance of the six models.

Purpose of the Study

The major purposes of the current research are twofold. The first purpose is to present a composite approach to the general classification problem by using outputs from various parametric statistical procedures and neural networks. Statistical outputs are used as inputs to the neural networks and vice versa in an attempt to enhance the classificatory performance of both methods beyond what either method can produce alone. This technique is in stark contrast to the generally used methods for developing classification models (see, for example, Patuwo, Hu, and Hung 1993 and Subramanian, Hung, and Hu 1993). It is an extension of the work begun by Markham and Ragsdale (1995) in which linear discriminant outputs were input into a BPNN in order to enhance its classificatory performance. The advantage of the method introduced in this study is that it attempts to enhance the performance of both the parametric and neural network models regardless of which performs best in the particular domain of interest.

The second purpose is to compare several parametric and neural network models on a transportation planning related classification problem and five simulated classification problems. The transportation related problem entails using live survey data from the Dallas/Fort Worth metroplex to develop transportation mode choice models. The traditional method used for modeling mode choice decisions in nearly all major metropolitan areas around the world is the multinomial logit model (Ortuzar and Willumsen 1994). This model's performance is compared to linear and quadratic discriminant analysis as well as several neural network models including not only the most widely used BPNN model, but also MNN and RBFNN models. The structure of MNNs and RBFNNs allow them to overcome some of the limitations of BPNNs such as slow convergence, getting stuck in local minima, (Lee et al. 1996), and temporal crosstalk (Jacobs and Jordan 1993). To date, no work has systematically evaluated the efficacy of these various models in the transportation mode choice domain nor have they all been compared in any simulation study. The comparisons of these classification methods is made using simulated bivariate normal and contaminated data sets. These comparisons are intended to shed light on how the degree of data overlap, operationally defined as the distance between group means coupled with the group variances, and data contamination affect the performance of the various classification models.

The Classification Problem

Classification is a fundamental function without which survival of our species would not be possible. From the time we are born, we begin to identify and group characteristics of our environment into classes that allow us to separate the pleasurable from the painful, the safe from the dangerous, and the friendly from the hostile. Classification is the most fundamental process used to turn seeming randomness into intelligible, predictable phenomena. This process of observation and classification was systematized by ancients, both primitive and civilized, in order that they might better understand and regulate their environment. For example, in order for people to regulate the planting and harvesting of crops, it was necessary that they classify the days of the year into the appropriate season. The formalization of the taxonomy of natural phenomena is generally ascribed to the classical Greeks and constitutes an important component of the scientific method (Asmis 1984).

In addition to being critical to the survival of the individual and fundamental to the scientific method, classification is a process basic to the flourishing of all modern organizations, both public and private. Classification is important for both efficiency of operation and prediction of future states. For example, in order to more efficiently manufacture parts, it is sometimes important to classify them into groups according to their processing similarities. This part classification process is fundamental to the implementation of group technology, a technique that leads to more flexible manufacturing of smaller batches of parts, while reducing setup time, wait time, and part movement (Hyer and Wemmerlov 1989; Knox 1980). In the marketing arena, efficient buyer targeting requires the segmentation of potential customers into groups according to their proclivity to purchase the product of interest (Venugopal and Baets 1994). This reduces the total resources expended soliciting customers and at the same time yields the highest return on the marketing dollar. A final example of the necessity of classification by organizations, one most relevant to the current research, is the grouping of individuals according to the mode of transportation they will select for a specific trip. By determining the mode choice for each trip, transportation planners are better able to design transportation networks and policies so as to improve the flow of people and goods from their origins to their destinations. Improved traffic flow has significant implications for productivity improvement, environmental quality, economic growth, and the quality of urban/suburban life (Owen 1992).

In all of these and thousands of similar classification problem domains, it is imperative that the objects of interest be placed in the appropriate categories. To this end, many techniques have evolved, improving our ability to classify accurately. Under some limited set of known conditions for a given classification problem, one or another of these methods are able to identify the underlying features of the objects needing classification and place them almost perfectly into the appropriate categories. For example, if the objects to be classified come from a mutivariate normal population, have equal covariance structures, and are linearly separable, then Fisher's LDA model (Fisher 1936) will yield optimal object classification (Subramanian, Hung, and Hu 1993). If the objects are spatially related in such a way that each member of a particular class is closer to the centroid of its class than to the centroid of any other class, then k-nearest neighbor clustering will yield perfect object classification (Manly 1994). Even under conditions where one or another of the underlying assumptions for use of a classification method have been violated, it is still possible for the method to yield a high correct classification rate (Huang and Lippmann 1987). This is true, for example, of the application of LDA to objects that are very linearly separable, are multivariate normal, but have unequal covariance structures (see, for example, Clarke, Lachenbruch, and Broffitt 1979).

The real world, however, is generally not benevolent enough to provide objects possessing the requisite characteristics for any given classification model. This often leads to less than stellar results and provides motivation for the development of improved classification methods (see, for example, Shepherd 1997). The search for improved methods goes on continuously as can be seen by a perusal of the literature (see, for example, Anand et al. 1993; Fritzke 1994). The overwhelming majority of the studies focus on a very narrow aspect of a particular classification problem (see, for example, McLachlan 1974; Tubbs, Coberly, and Young 1984; Kandil et al. 1996) . They generally demonstrate either the superiority of one existing method over another (Liao and Gen 1993) or the superiority of a new enhancement of an existing method over the corresponding non-enhanced method (Poechmueller, Glesner, and Juergs 1996) or a more traditionally used method (Swarup and Chandrasekharaiah 1992). Each study makes an incremental contribution to the classification research body of knowledge but leaves unanswered many questions regarding generalization to other classification domains It is into one of these domains that the current research ventures with the hope of adding another significant piece to the classification puzzle.

Significance of the Study

The following statements outline an argument indicating the importance of viable classification methods that lend support to the search for incrementally better methods.

1. Classification is often a vital step in the process of scientific inquiry.

2. Many business applications require the classification of states.

3. Current parametric and nonparametric classification methods fall short of desired performance in many application domains.

4. Transportation mode choice is an important classification problem domain for which any improvement in classificatory performance would yield large dollar benefits.

The current research more thoroughly integrates what have hitherto been generally isolated methods of classification. This integrated approach makes more complete use of the information content of the classification data. That is, both the parametric and functional relationships among the data elements are exploited more thoroughly than if statistical or nonstatistical (e.g., neural networks) methods were used in isolation.

The composite approach to classification, interchanging information between statistical and neural network models, provides an alternative procedure for researchers and practitioners who might be unsure as to which stand-alone model to use. This integrated approach is an extension of previous research (Markham and Ragsdale 1995). This study directly addresses the question of the usefulness of information from statistical classification models to neural network classification models and vice versa. Also, the investigation of the various neural network (BPNN, MNN, and RBFNN) and statistical (LOG, LDA, and QDA) models provides a direct comparison of their classificatory performance and a demonstration of their relative strengths and weaknesses under several data conditions. While these techniques have all been examined previously they have not been studied simultaneously or been given equal attention.

CHAPTER 2

LITERATURE REVIEW

Classification is the process of assigning objects to groups based on a vector of measured values (Rencher 1995). It is an activity that is an integral part of many disciplines including science, medicine, education, government, and business. The ongoing effort in the research community to gain a better understanding of the behavior of extant classification methods and to develop new and enhanced classification methods has yielded important results. This chapter presents the results of the most relevant of these studies where relevance is defined as presenting theoretical and/or practical support for the current research. The chapter begins with a brief look at the breadth of research related to classification and the development and use of the classification methods used in this study. It then turns to an examination of the efficacy of each of the proposed methods as measured by their performance in various classification domains and their relative performance compared to other classification methods. Attention is then turned to the utility of combining different types of models to improve overall performance on a given task. Next, a description of the transportation planning process is presented with special attention paid to mode choice modeling, the primary application domain for the proposed research. The chapter ends with a presentation of the research questions investigated.

Statistical Models

Classification of sensory information is a task that humans generally perform more effectively and efficiently than computing machines (McClelland, Rumelhart, and Hinton 1986). We quickly process millions of bits of information and accurately place into the appropriate category the face of a person we know, the species of an animal we have never seen before, and the period and composer of a piece of music we have not heard in years (Levine 1991). When it comes to structuring large amounts of non-sensory multidimensional data characterizing some object to be classified (e.g., financial ratios and other organizational characteristics to be used for classifying a bank's bankruptcy likelihood), humans do not fare well (Kosko 1992). Since this type of classification task is important to the functioning and flourishing of organizations, it has been a catalyst for the development and/or improvement of a variety of artificial classification methods. For example, it was the need to understand survival characteristics that led to the development of the logit (LOG) model (Mendenhall and Sincich 1993). Logit is used to model discrete choice behavior and finds application in, for example, transportation mode choice modeling (Ortuzar and Willumsen 1994). Fisher developed the linear discriminant function in order to solve certain physical anthropology and biology classification problems (Fisher 1936). Quadratic discriminant functions were developed to overcome the sensitivity of Fisher's method to unequal covariance structures within the different classes (Smith 1946).

Neural Network Models

The history of the development of neural network classification begins with attempts to mimic the most basic elements of the human central nervous system (CNS). These elements were then organized into more complex structures that possessed some of the fundamental characteristics of the CNS and were able to perform some of its basic functions, such as, sensory classification. For example, McCulloch and Pitts (1943) developed a mathematical formalism of a human neuron (known today as a McCulloch-Pitts neuron) which when combined into networks are able to duplicate any logical function (Levine 1991). McCulloch and Pitts present networks of these primitive all-or-none threshold logic neurons that demonstrate a number of sensory processes, such as the sensing of hot and cold. Other notable early researchers contributed to the understanding of neural learning mechanisms (Hebb 1949), memory (Lashley 1950; von Neumann 1958), and the simulation of neural processes (Rochester et al. 1956).

In 1958, Rosenblatt published his landmark paper describing the perceptron. The perceptron consisted of sensory, associative, and response units designed to couple classification responses to stimuli. The process of learning the desired relationship between a given classification response and its corresponding stimuli, required that the system reorganize itself so as to reflect the newly learned association. Networks comprised of perceptrons were capable of learning to classify N binary stimuli into any of their possible classifications (Rosenblatt 1962). Although much controversy resulted from Rosenblatt's claims regarding the capabilities of perceptrons (Minsky and Papert

1969) and broad-based work on neural networks nearly disappeared, variants of these simple computational devices are still used in many types of neural networks.

Some serious work on neural networks did continue after the perceptron controversy. Kohonen (1972) and Anderson (1972) independently and simultaneously developed the same neural network model for associative learning. These models were a class of memory models that learned and stored associations between given input and output vectors in a connection matrix. Also during this time, work continued on the development of feature detector neural networks (Grossberg 1976; Malsburg 1973), categorical perception and probability learning (Anderson et al. 1977), and machine vision (Fukushima 1975). The publication of John Hopfield's discrete associative memory model (Hopfield 1982) with modified Hebbian learning that once again legitimized the formal study of neural networks.

Although Hopfield lent legitimacy to the discipline, the true renaissance in neural network research began with the publication of the work of Rumelhart, Hinton, and Williams (1986) on a modification of the Widrow-Hoff learning rule (Widrow and Hoff 1960) called backpropagation (Rumelhart, Hinton, and Williams), as it turns out, were not the only ones nor even the first to develop the back propagation algorithm. Others, such as, Werbos (1974) and Parker (1982) had both derived the algorithm prior to Rumelhart, Hinton, and Williams and Le Cun (1986) arrived at it concurrently. In addition, White (1989) demonstrated that backpropagation was not a new form of learning but was merely a new way of implementing learning by stochastic approximation which had been around since the 1950s.). The Widrow-Hoff learning rule is a least mean square procedure based on gradient descent applied to discrete inputs and single output (-1, +1) adaptive linear elements. The backpropagation learning algorithm implements a generalized Widrow-Hoff learning rule. It is based on a gradient descent procedure that minimizes a global error function that is comprised of the differences between the actual output from the network and the desired output (A complete description of the backpropagation learning algorithm can be found in the methodology chapter.).

With the advent of the backpropagation algorithm came a renewed stream of neural network research. The breadth and volume of this research has grown steadily into a torrent. The backpropagation algorithm has been applied to a vast array of problem types including forecasting (Mitchell, Yi, and Govind 1996; Yi, Mitchell, and Prybutok 1996), function approximation (Widrow, Rumelhart, and Lehr 1994), and classification (Pavur et al. 1997). In addition to backpropagation neural networks, a number of other neural network types have been studied to determine their efficacy for various classification tasks. These networks include fuzzy ARTMAPs (Carpenter, Grossberg, and Rosen 1991; Lee and Lai 1993), learning vector quantization networks (Poechmueller, Glesner, and Juergs 1993), modular neural networks (Jacobs et al. 1991; Jordan and Jacobs 1992), genetic algorithm trained neural networks (Kitano 1994; van Rooij, Jain, and Johnson 1996), probabilistic neural networks (Specht 1990), and radial basis function neural networks (Moody and Darken 1989; Musavi et al. 1992). Of these many types of neural networks, the current research focuses on backpropagation neural networks (BPNN), modular neural networks (MNN), and radial basis function neural networks (RBFNN) because of their superior classificatory performance during preliminary tests.

Support for the Included Classification Models

Each of the statistical and neural network classification models used in the current research has been studied extensively and shown to be a viable method for classification (Demaris 1992; Lippmann 1987; NeuralWare 1995; Rencher 1995; Zhao and Bao 1996). For example, LOG models have been effective in modeling public transportation demand by elderly and disabled people (Stern 1993) and farmers' stated level of agricultural risk (Kastens and Featherstone 1996). Transportation planners use LOG models almost exclusively for transportation mode choice modeling (U. S. Department of Transportation 1986). An example of this use of the LOG model is presented by Greene (1990). He describes a study done by Hensher in Sydney, Australia in which commuters' mode choices were modeled. Ten independent variables (such as, in-vehicle time and parking cost) were used to classify commuters into one of four transportation modes: 1) car/driver, 2) car/passenger, 3) train, and 4) bus. In another discrete choice study, a multinomial LOG model was used to analyze the stated choice of tourist destination (Morley 1994). Three price factors, airfares, hotel tariffs, and exchange rates were included in the LOG utility function to determine their impact on the stated decision to travel from Kuala Lumpur to one of eight potential destinations. The LOG model was used to classify subjects into destination categories. The results indicated that airfare was the most significant price factor influencing destination choice. LOG models have also been employed to model a court's decision in frequently litigated areas of tax law (Madeo

1979; Burns and Groomer 1983), to classify workers as having been dismissed, laid off, or having quit within six months of starting work (Campbell 1997), and using state-level data concerning potential job gains and loses, immigration, exports, and unemployment, to predict votes for and against the North American Free Trade Agreement by House and Senate members (Kahane 1996).

Discriminant analysis is another of the traditional statistical techniques commonly used by researchers and practitioners alike for classification purposes. The literature is replete with examples of linear discriminant analysis (LDA) applied to business classification problems but rather sparse with regard to examples using quadratic discriminant analysis (QDA). This appears to be another example, like that of linear regression, of the use of a simpler linear model in almost all research regardless of its appropriateness. An example of the use of LDA for classification purposes is provided in a study by Khan, Chawla, and Cianciolo (1995) that identifies potential marketing targets for small accounting services firms. In this study, small businesses were classified by LDA as either using personal computers or not. The authors suggested that firms misclassified as personal computer users be investigated as potential marketing targets for computerized accounting support because of their relative similarity to computer using firms. In another study focusing on voting behavior in a multi-union representation election (Smith, Hindman, and Havlovic 1997), employees were classified into one of three categories: 1) voting for the incumbent union, 2) voting for the raiding union, or 3) not voting. Correct classification of more than 70% across all three categories was obtained using a variety of employee demographic, union perception, and degree of union involvement variables. Other applications of LDA include bankruptcy prediction (Hughes 1993), credit scoring (Leonard 1993), and discriminating between new service businesses and new manufacturing concerns (Birley and Westhead 1994).

The only use found of QDA was in studies making comparisons of two or more classification methods. This indicates that researchers find QDA useful as either a benchmark against which other models should be compared or as a viable classification model outright. In one such study, QDA was compared to LDA and two kernel classification models (Shah 1997). The task involved classifying a given time series as belonging to one of three time series prediction models that it most closely matched based upon various of its structural characteristics. The results indicated that QDA generally outperformed LDA across a variety of classification performance measures. Other comparison studies tend to use QDA, in conjunction with LDA, as a benchmark statistical model against which to compare the classificatory performance of neural networks. While most of these studies have used simulated data exclusively (see, for example, Subramanian, Hung, and Hu 1993), some have used empirical data as well (see, for example, Huang and Lippmann 1987). Regarding the viability of QDA for classification use, Huang and Lippmann (1987) found that under conditions ideal for its use, QDA was competitive with both LDA and BPNN models. Patuwo, Hu, and Hung (1993) found that under numerous experimental conditions, QDA performed as expected based on Bayesian classifier theory.

Backpropagation and its variants (see, for example, Jacobs 1988) have been used in a myriad of classification studies in a broad array of problem domains. These include many studies in the science, engineering, and medical domains. For example, Cawley and Dorling (1996) used a multi-layer perceptron (a synonym for a feedforward network) to classify atmospheric circulation weather patterns using Rprop, a variant of the backpropagation algorithm. The results showed that the neural network classifier outperformed the traditionally used rule-based system. Studies such as those by Gorman and Sejnowski (1988) and Dror, Zagaeski, and Moss (1995) have demonstrated the utility of BPNNs for classifying sonar targets. In both of these studies, classification rates in excess of 90% were obtained on never before seen validation data. In a study on classification of an area of interest in a road-vehicle scene image, Hollis et al. (1996) used two BPNNs to identify objects in the image. The first network was used to distinguish true objects from anomalies. The second network then classified the objects into car, pedestrian, and bus categories. The classification rate on the validation data set was 89.7%. In another type of engineering study, Dougherty, Kirby, and Boyle (1994) found that a BPNN could successfully determine whether a given roadway network link was congested or not. In the medical domain, BPNNs have been used to detect the presence of heart disease (Shen et al. 1992), to classify patients into disease categories (fatty liver; non-A, non-B hepatitis; chronic aggressive non-A, non-B hepatitis) based on clinical laboratory data (Reibnegger et al. 1991), and to identify the location of primary tumors (Bonelli and Parodi 1991).

Classification tasks from the business domain have not been neglected by BPNN researchers. In the realm of real-time industrial process monitoring/control, for example, BPNNs have shown promise in monitoring on-line tool wear of automated machine tools

(Das, Chattopadhyay, and Murthy 1996), classifying temperature profiles in a blast furnace (Saxon, et al. 1995), and detecting faults in nuclear power plants (Jalel et al. 1991). Security applications include authenticating computer users using keystroke dynamics (Rostar and Olejar 1995) and recognizing and classifying computer viruses (Doumas et al. 1995). In the realm of quality control, BPNNs have been successful at recognizing control chart patterns and classifying a process as in or out of control (Velasco and Rowe 1993) as well as monitoring hard-disk drive quality (Sieger and Badiru 1993).

The largest number of BPNN business studies falls in the finance domain. Deployed finance applications have been in operation for more than 14 years (Li 1994). For example, Longo and Long (1997) used differences between various characteristics of notable stock market 'winners' and 'losers' to classify stocks as future 'winners' or 'losers'. Other finance applications include stock-price classification (Yoon and Swales 1991), bankruptcy prediction (Raghupathi, Schkade, and Raju 1991; Fletcher and Goss 1993), thrift-failure prediction (Salchenberger, Cinar, and Lash 1992; Tam and Kiang 1992), bond rating (Dutta and Shekhar 1988), corporate-distress classification (Altman, Marco, and Varetto 1994), and credit scoring (Tana and Koh 1997).

The studies investigating RBFNNs are sparse and those investigating MNNs nearly nonexistent when compared to the BPNN literature. There are, however, studies for both of these models that demonstrate their efficacy for classification problems. Modular neural networks (also known as hierarchical mixture-of-expert networks) were used by Jacobs, Peng, and Tanner (1997) to classify breast-cancer tumors as either benign or malignant and to perform a speech recognition task involving the classification of ten vowel sounds. Modular neural networks have also been applied to a credit scoring task (Desai, Crook, and Overstreet 1996), classification of simulated data (Jordan and Xu 1995), and non-classification tasks such as real-time control of a robotic arm (Jacobs and Jordan 1993). Although relatively little work has been done with MNNs, they can be applied to any classification problem for which a BPNN is appropriate (NeuralWare 1995).

Radial basis function neural networks have seen somewhat more widespread use than MNNs. A pair of RBFNNs operating in tandem has been successfully used to model a nonlinear hydraulic test rig and identify test-rig faults (Yu, Shields, and Daley 1995). In another fault detection study (Ribeiro, Costa, and Dourado 1995), RBFNNs were trained to identify industrial lime-kiln malfunctions and proved able to generalize fault detection to novel inputs. In the medical domain, RBFNNs were used by Bounds and Lloyd (1990) to classify low back disorders. The RBFNNs were able to outperform three groups of medical experts on this classification task. Classification of transformed visual images has been attempted with RBFNNs for both radar profiles of military aircraft (Zhao and Bao 1996) and hand gestures with missing and uncertain inputs (Ahmad and Tresp 1996). In a finance related application (Williamson and Munson 1995), an RBFNN was used to classify good and bad loan cases in a credit application vetting system. These studies and numerous others (e.g., Roth 1990), demonstrate the utility of RBFNNs in a variety of classification problem domains.

Comparative Studies

Many studies have compared various combinations of the included statistical and neural network models from both theoretical and applied perspectives. As regards the relative performance of these various models, there has been a general lack of consistent results (Jain and Nag 1997) which might be partially due to the unpredictability of the solutions of the widely used BPNN (Wang 1995). In addition, many of the results reported are merely anecdotal, failing to sufficiently validate the models (Glorfeld and Ching 1992), using only a single data set (real or artificial) for the comparisons, and/or not comparing classification results against even one other competing model (Prechelt 1996). A brief survey of a variety of these classification model comparison studies follows.

A common type of comparison study uses data simulated from known probability distributions to test the classificatory performance of one or more parametric techniques against one or more neural network models. In many of these studies, the simulation examples lack the complexity necessary to provide reasonable tests of the various classification methods (Hergert et al. 1992). In an early study, Huang and Lippmann (1987) performed several experiments that compared BPNNs to QDA models. On simple problems such as one continuous Gaussian input with two classes, the QDA model performed best although the BPNN models' performances were nearly equivalent. On perfectly separable nonnormal data, QDA and BPNN models performed nearly identically. When outliers were included in a mix of normal and nonnormal data, the performance of the QDA models deteriorated significantly while the BPNN models exhibited a great deal of robustness. Curram and Mingers (1994) compared BPNNs and LDA models on seven different classification problems, three of which involved simulated data. One of the simulated data sets included equal numbers of observations inside and outside of a sphere, a condition specifically designed to favor the BPNN. As expected, the BPNN performed significantly better (94.4% correct classification) than the LDA model (46.6% correct classification). On the other two simulated data sets, that did not involve significant nonlinearities, the classificatory performance of the BPNN and LDA models was indistinguishable.

Radial basis function neural networks have been shown to give better classificatory performance than BPNNs in a number of studies. For example, in a credit application vetting system study (Williamson and Munson 1995), the RBFNN classified 84% of the loan applications correctly compared to 77% for the BPNN. Similar results were obtained by Ceccarelli and Hounsou (1996) in their speech recognition study. Finally, in two studies investigating different aspects of lime-kiln operation in the paper and pulp industry, RBFNNs were inconsistent in their classificatory performance relative to BFNNs. They performed better when tasked with classifying kiln faults (Ribeiro, Costa, and Dourado 1995) but worse on a control task (Ribeiro, Dourado, and Costa 1995).

Combining Models

It is often the case that the complexities of the world leave us with data that cannot readily be optimally accommodated by a single model. As a result, researchers have attempted to combine models, both statistical and nonstatistical, to overcome an individual model's shortcomings. As an example, genetic algorithms have been combined with neural networks to optimize their topology and learning parameter settings (see, for example, Carse et al. 1995). The genetic algorithm replaces the normal trial-and-error method of settings these parameters. Genetic algorithms have also been used as a training algorithm for neural networks (see, for example, Huang, Dorsey, and Boose 1994). This method is useful when non-differentiable transfer functions are required, thereby eliminating backpropagation training as an option (Whitley 1995), and overcomes the BPNN's propensity for getting stuck in local minima (Petridis et al. 1992).

Using one model to preprocess data for a second model is another method used for combining techniques. Factor analysis has been used to preprocess data for both LOG (see, for example, Richard, LeMay, and Taylor 1995) and LDA (see, for example, Deal and Edgett 1997) models. In these studies, factor analysis is used to reduce the dimensionality of the inputs before classification is attempted. In another example, Hernandez et al. (1992) used a BPNN to preprocess satellite image data that was then fed into a learning vector quantization neural network for classification. The preprocessing was intended to both filter noise and encode the satellite image features. Toulson, Boyce, and Hinton (1992) developed a statistical classifier that maps gray level information from magnetic resonance images into a set of fuzzy pixel classifications that were then fed into a neural network for classification into anatomical types.

Finally, using statistically preprocessed data as additional input to a neural network model has been used to enhance the network's classificatory performance. In a classification study involving road vehicle scenes, Hollis et al. (1996) used quartile statistics from scene images as additional input to a BPNN. Using the quartile statistics in addition to Maitra's moment invariants feature vectors reduced the BPNN training time and improved the classification rate from 86.13% to 96.35% over the Maitra's moment invariants alone. Markham and Ragsdale (1995) used three classification approaches on two different real-world tasks. The first approach was statistical and used Mahalanobis distances, which is equivalent to using Fisher's linear discriminant function (i.e., LDA), for classification purposes. The second approach utilized a BPNN with the same input as was used to calculate the Mahalanobis distances. The third approach, a hybrid of the first two approaches, also made use of a BPNN. In this case, the Mahalanobis distances calculated using the first approach were used as additional input to a BPNN. The hybrid model performed statistically better than the other two models ($p < .005$), thus indicating the possible utility of using statistical classification information for enhancing the performance of a BPNN.

Transportation Planning

No matter in what large metropolitan city one finds oneself, a person is likely to be confronted with automobile clogged highways. From Los Angeles to Tokyo, commuters face traffic congestion that causes loss of precious time, a great deal of frustration, increases in ambient air pollution, and generally reduces the quality of life in the world's urban centers (Owen 1992). Although theory and tools exist to deal with this ubiquitous problem, one afternoon in rush hour traffic is testament to their inadequacy.

In the early 1950's, the federal government developed a four-step planning process called the urban transportation planning system (UTPS) as a tool for use by what are now called metropolitan planning organizations (The following discussion of the UTPS transportation planning process is based on Shunk (1992)). Over the intervening years, the UTPS has been refined and enhanced, but still remains the primary tool for urban transportation planning.

The first step of the UTPS process, called trip generation, uses various demographic and land use data to predict the number of trips that will be generated in a specific geographical region on an average day. A trip is any one-way movement from an origin to a destination. The purpose of the trip generation phase of the planning process is to estimate the average number of daily trips produced in and attracted to each travel zone. A travel zone is the areal unit of interest in a metropolitan transportation study. These values, when summed, represent the travel demand for the metropolitan area.

The trips generated in the first step of the UTPS process are used in the second step of the transportation planning process to determine trip origins and destinations. It is in this step that trip origin zones and destination zones are linked for each of the generated trips. These linkages are generally made using some form of gravity model which pairs trip ends in proportion to the relative attractiveness of potential trip end zones.

The trips between origin and destination zones must be made by one of a number of possible transportation modes, among which travelers must choose. This transportation mode choice decision is modeled in the third step of the UTPS process. The most common formulation for the mode choice decision is the LOG model. Mode choice models are generally the most complex of the UTPS models. These models are typically used to estimate the number of persons that will choose to ride some form of transit versus those that will choose to use some form of private vehicle for trips made between a given pair of origin and destination zones. The LOG model mode choice decisions are based on a utility function. The utility function is usually composed of various level-of-service measures for each of the transportation modes and a set of variables describing the socioeconomic characteristics of travelers with respect to each possible mode. Level-of-service measures might include such characteristics as travel time, transportation costs (e.g., fuel, tolls, and parking), walking distance to and from transit stops or parking lots, and wait time for transit service. Socioeconomic characteristics include household income levels, number of household vehicles, and number of people in the household. The determination of the relevant utility function variables is what makes the mode choice model difficult to develop. It is this classification of person trips into personal vehicle trips (highway trips) and transit trips that is the focus of the real-life data portion of the current study.

The last step of the UTPS process is traffic assignment. The zone-to-zone interchanges occurring via each mode, as specified in step three of the process, are assigned to transportation network links in this step. This network-loading step completes the picture of transportation related movements in the given metropolitan area. The overall accuracy of the transportation planning process bears directly on the ability of our urban centers to improve the efficiency with which people and goods are moved.

This efficiency is directly related to transportation costs and the general quality of life in our metropolitan centers.

Research Questions

Research Question 1

How effective is an integrative approach that uses information from both neural network and statistical models for data classification, compared to the corresponding stand-alone procedures? For which data configuration(s) is this approach effective and likely to show the greatest potential for improved classificatory performance?

Motivation for Research Question 1

There is no *a priori* method for guaranteeing that a given classification method will perform well on a particular classification problem. In addition, the complex nature of real-world data generally obfuscates the model selection process. These facts notwithstanding, a number of statistical and neural network models present themselves as possible alternatives for classifying a given set of observations. These various models approach the classification task in different ways, each designed to exploit the structure embedded in the data in a slightly different fashion. The results of the different classification approaches provide synopses of the underlying structures that might be exploited by other classification methods. Indeed, a number of researchers have recognized that such structural reductions might provide advantages to methods able to interpret them.

For example, Damodaran, Kolli, and Alexander (1993) used autocorrelations and partial autocorrelations extracted from time series data as inputs to a BPNN. The BPNN was able to correctly classify the patterns as ascending, descending, and cyclic. Wang (1995) has incorporated knowledge concerning inflection points to enhance the performance of a monotonic neural network model and indicated that, in general, neural network models might demonstrate improved classificatory performance if allowed to incorporate knowledge gained from other classification modelss. This notion of integrating information from different types of models (such as statistical and neural network models) was echoed by Clemen (1989). After reviewing 25 years of combined model time-series forecasting research, Clemen states that combining forecasts should be part of the mainstream of forecasting approaches. Given these successes and suggestions for combining the outputs of various classification and/or forecasting models, a strategy for applying the proposed iterative technique appears reasonable. This study is intended to determine if the composite classification technique is more accurate than either the parametric or neural network procedures in isolation. Even if a given neural network procedure is already better than a given parametric discriminant procedure for a specific classification problem, the composite approach may enhance its classificatory ability.

Markham and Ragsdale (1995) demonstrated that statistical information, Mahalanobis distances, can enhance the classificatory performance of a BPNN. Their approach uses statistical classification results as inputs to a BPNN model but does not investigate the use of neural network output as a means of enhancing the classificatory performance of a statistical model. Our study differs in that it also considers using results from different neural networks as inputs to the three included statistical models. While Markham and Ragsdale illustrated their approach using two real-world data sets, their study suffers from relatively small sample sizes and no Monte Carlo simulation in which various data configurations could be examined. While their method showed improvement over Fisher's LDA, no comparison was made with a QDA model, which may be more appropriate for data in which violations of equal population covariance structures is present.

Research Question 2

What is the relationship between the relative proportion of training, testing, and validation exemplars in each of the output classes and the performance of each of the classification methods? At what proportional mix (25/25/25/25 and 33/8/38/21 for the 4-class D/FW transportation mode choice configuration and 50/50 and 70/30 for the simulation configurations) does a given classification method perform best?

Motivation for Research Question 2

Many neural network classification studies have used only equal proportion classes in their investigations (see, for example, de Villiers and Barnard 1992; Subramanian, Hung, and Hu 1993). The real world, however, often presents us with very disproportionate classes to be separated (see, for example, Dutta and Shekhar 1988; Fletcher and Goss 1993). Having a class with a very small proportion of the sample observations might lead to poor classificatory performance of a given method because of its inability to discern the features necessary for classification (Sharda and Wilson 1996).

How the classification method learns to recognize relevant classification features is separate from the issue of feature coverage. For example, the backpropagation algorithm tends to converge slowly when one class contains most of the exemplars (Anand et al. 1996). This is due to the fact that the algorithm uses the net error gradient, which is dominated by the larger class, to adjust its feature detection parameters. Jain and Nag (1997) investigated class proportions as part of an experiment and found that LOG and BPNN models respond differently to data with equal proportions versus data with highly disproportionate class membership. Given these results and the general lack of research and understanding concerning disproportionate data classes, it seems reasonable to investigate this issue.

Research Question 3

How will the six classification methods perform relative to each other under each of the six data configurations (DFW and five simulation configurations) within each of the two class proportion conditions?

Motivation for Question 3

It is well known that LDA and QDA are optimal Bayesian classifiers when all the relevant parametric assumptions are met (Rencher 1995). In addition, a BPNN with sufficient hidden layers and hidden layer neurons can perform any nonlinear mapping of inputs to outputs (Lippmann 1987). Even with this level of understanding, currently no method exists for determining which of the many available classification models will perform best in a novel problem domain that is ill-defined. For each new problem domain, the myriad solution options must be evaluated to determine their efficacy. The transportation mode choice domain fits this profile and therefore, even though the LOG model has come to be the most widely accepted technique for this classification problem, the other five proposed models cannot, *a priori*, be disqualified. The Dallas/Fort Worth transportation data used in this study has been exposed to only one of the six included models, the LOG model. In addition, to our knowledge no previous Monte Carlo simulation study has simultaneously investigated the six classification models or the three neural network models investigated in this study.

CHAPTER 3

RESEARCH METHODOLOGY

Six different classification models, three statistical and three neural network, are used to solve each of twelve classification problems. This chapter describes the six classification models used in this study, the twelve classification problems addressed, and the experimental designs and procedures employed. The six classification models are:

1. Linear discriminant analysis (LDA)
2. Quadratic discriminant analysis (QDA)
3. Logit (LOG)
4. Backpropagation neural network (BPNN)
5. Modular neural network (MNN)
6. Radial basis function neural network (RBFNN).

Linear Discriminant Analysis

Linear discriminant analysis was first described by Fisher (1936). The material described below (for both LDA and QDA) is based primarily on the discussion in Rencher (1995). Fisher's procedure classifies a vector of measurements, $\mathbf{y}$, representing an observation sampled from one of two populations into one of two groups (G_1 or G_2). The only requirement is that the two populations, from which the vectors are sampled, have equal covariance structures ($\Sigma_1 = \Sigma_2$). If, however, the two populations are also multivariate normal, then Fisher's procedure results in optimal classification of the sampled vectors. In order to classify the sampled vectors, it is first necessary to calculate $\overline{\mathbf{y}}_1$, $\overline{\mathbf{y}}_2$, and $\mathbf{S}_{pl}$, where $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$ are the population 1 and population 2 centroids respectively, and $\mathbf{S}_{pl}$ is the pooled covariance matrix. Each sample vector is then transformed using the discriminant function

$$z = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} \mathbf{y} \cdot$$

If there are n_i, i = 1, 2, sample observations from populations 1 and 2 respectively, then for each observation, $\mathbf{y}_{ij}$, j = 1, 2, . . ., n_i, from samples i = 1, 2, there will result $\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{in_{ii}}$ discriminant scores that can be used to calculate $\bar{z}_i = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} \overline{\mathbf{y}}_i$. Each observation, $\mathbf{y}$, is then classified into group G_1 if its corresponding z is closer to $\bar{z}_1$. Otherwise, the observation is classified into G_2. The observation, $\mathbf{y}$, is closer to $\bar{z}_1$ if

$$z > \frac{1}{2}(\bar{z}_1 + \bar{z}_2).$$

If the proportions of observations, p_1 and p_2, from the two populations are known and the populations are both multivariate normal, then the prior probabilities p_1 and p_2 can be used to modify the classification rule. The observation, $\mathbf{y}$, is assigned to group G_1 if

$$(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} \mathbf{y} > \frac{1}{2}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} (\overline{\mathbf{y}}_1 + \overline{\mathbf{y}}_2) + \ln \frac{p_2}{p_1}$$

and to G_2 otherwise. Thus, knowledge of the prior probabilities pushes the classification toward the group containing the larger proportion of observations.

Quadratic Discriminant Analysis

Linear discriminant analysis produces the best possible classification rate for populations with equal covariance structures. If the covariance structures of the two populations are unequal ($\Sigma_1 \neq \Sigma_2$), it is no longer possible to pool the covariance matrices ($\mathbf{S}_i$) from each of the samples. Under this condition, the formulation of the discriminant function is no longer linear in **y**. In order to maintain the optimal classification rate, the function becomes

$$D_i^2(\mathbf{y}) = (\mathbf{y} - \overline{\mathbf{y}}_i)' \mathbf{S}_i^{-1} (\mathbf{y} - \overline{\mathbf{y}}_i),$$

where $\mathbf{S}_i$ is the covariance matrix for the ith group, i = 1, 2. An observation, **y**, is assigned to the group for which $D_i^2(\mathbf{y})$ is smallest. This discriminant function is quadratic in **y**.

As with LDA, knowledge of prior probabilities p_1 and p_2, coupled with multivariate normality of the populations, leads to a discriminant function that requires stronger evidence of membership for classification into the lower probability group.

The resulting quadratic classification function becomes

$$Q_i = \ln p_i - \frac{1}{2}\ln|\mathbf{S}_i| - \frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}}_i)'\mathbf{S}_i^{-1}(\mathbf{y} - \bar{\mathbf{y}}_i).$$

An observation, **y**, is classified into the group, G_i, for which Q_i is largest.

Logit

Logit derives its name from the fact that the dependent variable in the model is the natural logarithm of the odds, where odds is the ratio of the probabilities of falling into one of two categories on a variable of interest (e.g., mode of transportation selected) (Demaris 1992). The 'log odds' is a function of a set of explanatory variables and is a special case of a loglinear model (In the current study, LOG models will be assumed to be linear models in a manner similar to LDA models.). In its most basic form, LOG modeling uses aggregated categorical explanatory variables (such as those found in a contingency table) to predict the log odds. If one or more of the explanatory variables are continuous, aggregated data is no longer appropriate unless each of the conditions is replicated. In this case, it is necessary to use disaggregated data and the resulting method is known as logistic regression. Since the current study contains continuous independent variables, logistic regression is used. The model is still referred to as logit as is common in the literature (see, for example, Salchenberger, Cinar, and Lash 1992).

The basic form of the binary LOG model is

$$\ln \frac{\pi_i}{1-\pi_i} = \beta' \mathbf{y}_i ,$$

where π_i is the conditional probability that observation *i* belongs to group 1 given $\mathbf{y}_i$, $1-\pi_i$ is the conditional probability that observation *i* belongs to group 2 given $\mathbf{y}_i$, and β is a vector of parameters to be estimated. Solving for π_i, the LOG model becomes

$$\pi_i = \frac{e^{\beta' y_i}}{1 + e^{\beta' y_i}} .$$

The parameters for the LOG model are estimated using the method of maximum likelihood. The estimated model is

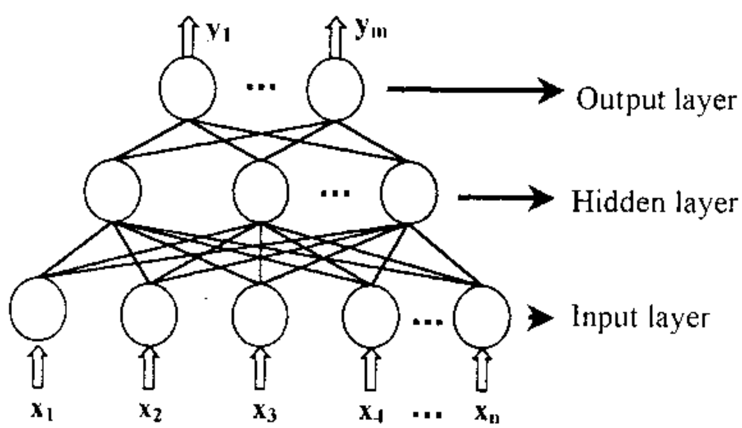
$$p_i = \frac{e^{b' y_i}}{1 + e^{b' y_i}} .$$

The LOG model can be extended to classification involving more than two groups. The resulting multinomial LOG model gives the conditional probabilities of membership in each of the groups.

Backpropagation Neural Network

Neural networks are characterized by three principle elements: 1) topology, 2) learning rule, and 3) recall (Simpson 1996). Topology refers to the number of layers in the network, the number of neurons in each layer, and the interconnections among the neurons. The learning rule is the means by which the parameters governing the transformation of inputs to outputs are updated as a result of experience. Recall is the method by which information stored in the network is retrieved.

Figure 1. General backpropagation neural network.



A BPNN's topology consists of an input layer, one or more hidden layers, and an output layer of neurons connected in a feedforward fashion. Information input to the network flows in one direction only. Figure 1 shows a generalized BPNN with one hidden layer and full connections only between successive layers. Connections can also be made directly from the neurons in the input layer to neurons in the output layer. Each connection has an associated weight that represents the strength of the relationship between the pair of neurons being connected. The vector of all weights represents the current state of network memory. It has been demonstrated that networks of this topology are capable of learning any continuous input-output mapping to any degree of accuracy (Widrow and Lehr 1990). However, although single hidden-layer topologies are capable of performing complex data transformations, such as those requiring nonconvex and/or disconnected decision regions, they cannot form all such regions (Gibson and Cowan 1990).

Although there are a number of learning rules available for training a BPNN, the most common is the generalized-delta rule which is a nontrivial extension of the learning rule for a simple Adaline network (Widrow and Lehr 1990). This learning rule is generally attributed to Rumelhart, Hinton, and Williams (1986) but a number of other researchers discovered it independently (Werbos 1974; Parker 1982; LeCun 1985). The generalized-delta rule uses the differences between the actual and desired outputs of all output neurons to calculate adjustments for the connection weights. Using gradient descent in the weight space, the algorithm propagates the error backward through the network, adjusting the connection weights so as to minimize the value of the error function. The most common of which is the quadratic error function.

The backpropagation algorithm as proposed by Rumelhart, Hinton, and Williams (1986) is described below (The notation used and description of the backpropagation algorithm given below follow that of Golden 1996). The notation used for the algorithm is as follows:

t = the number of the iteration;

c = the number of input neurons;

d = the number of hidden layer neurons;

p = the number of output neurons;

V = d x c-dimensional input neuron to hidden-layer neuron weight matrix;

W = p x d-dimensional hidden-layer neuron to output neuron weight matrix;

b = p-dimensional output neuron weight bias vector;

q = d-dimensional hidden-layer neuron weight bias vector.

The algorithm consists of the following steps:

1. Initialize all weights in the network to small values using either a Gaussian or uniform distribution with a mean of zero and variance σ^2.

2. Pick a training exemplar, **x**(t), at random. Let $\mathbf{x}(t) = [\mathbf{s}(t), \mathbf{o}(t)] = [\mathbf{s}^k, \mathbf{o}^k]$, where $\mathbf{s}(t) \in \Re^c$ is an input vector and $\mathbf{o}(t) \in \Re^p$ is the desired output vector, k = random integer between 1 and the number of training exemplars, and $(\mathbf{s}^k, \mathbf{o}^k)$ is an element of the training set $\{(\mathbf{s}^1, \mathbf{o}^1;), \ldots, (\mathbf{s}^n, \mathbf{o}^n)\}$.

3. Compute the hidden-layer output vector as $\mathbf{h}(t) = \underline{\Im}(\mathbf{V}\mathbf{s}(t) + \mathbf{q})$ where $\underline{\Im}_i : \Re^d \rightarrow (0,1)^d$ is defined such that the *i*th element of $\underline{\Im}$, $\Im_i : \Re \rightarrow (0,1)$ is $\Im_i(x) = 1/[1+\exp(-x)]$ for all $x \in \Re$ (Other transfer functions have also

been used such as the hyperbolic

tangent, $\Im_i(x)=[e^x - e^{-x}]/[e^x + e^{-x}]$ for all $x \in \Re$).

4. Compute the output-layer output vector as $\mathbf{r}(t)=\Im(\mathbf{W}\mathbf{h}(t)+\mathbf{b})$.

5. Compute the output error for the network as $\delta_r(t)=\mathbf{o}(t)-\mathbf{r}(t)$, where $\delta_r(t)$ is a p-dimensional vector.

6. Propagate the error backward from the output neurons to the hidden-layer neurons. The error at the hidden-layer neurons is given by $\delta_h(t)=\mathbf{W}^T\mathbf{D}_r(\mathbf{r}(t))\delta_r(t)$, where $\mathbf{D}_r(\mathbf{r}(t))$ is a p-dimensional matrix whose *i*th diagonal element is $r_i(t)(1-r_i(t))$ where $r_i(t)$ is the *i*th element of $\mathbf{r}(t)$ and all off-diagonal elements are zero.

7. Adjust the weights, representing the distributed memory of the mapping, using the following equations:

$$\mathbf{W}(t+1)=\mathbf{W}(t)+\gamma\mathbf{D}_r(\mathbf{r}(t))\delta_r(t)\mathbf{h}(t)^T;$$

$$\mathbf{V}(t+1)=\mathbf{V}(t)+\gamma\mathbf{D}_h(\mathbf{h}(t))\delta_h(t)\mathbf{s}(t)^T;$$

$$\mathbf{b}(t+1)=\mathbf{b}(t)+\gamma\mathbf{D}_r(\mathbf{r}(t))\delta_r(t);$$

$$\mathbf{q}(t+1)=\mathbf{q}(t)+\gamma\,\mathbf{D}_h(\mathbf{h}(t))\delta_h(t).$$

The d-dimensional matrix $\mathbf{D}_h(\mathbf{h}(t))$ has *i*th diagonal elements of $h_i(t)(1-h_i(t))$ where $h_i(t)$ is the *i*th element of $\mathbf{h}(t)$ and all off-diagonal elements are zero.

8. Increment t = t+1 and return to step 2 if the maximum desired number of iterations, t_{max}, has not been reached and the algorithm has not converged, $(1/t)\sum_{k=1}^{t}|\mathbf{o}(k)-\mathbf{r}(k)|<\varepsilon$, where ε is the maximum acceptable error. Otherwise, stop.

Modular Neural Network

The MNN architecture for the proposed study is shown in figure 2. It consists of a collection of feedforward networks that act as local experts and a single gating network. The local experts compete to learn mappings of subportions of the input space (Jordan and Xu 1995). The notation used to describe the MNN is as follows:

t = the number of the input exemplar, t = 1,... N;

n = the number of input neurons;

m = the number of ouput neurons;

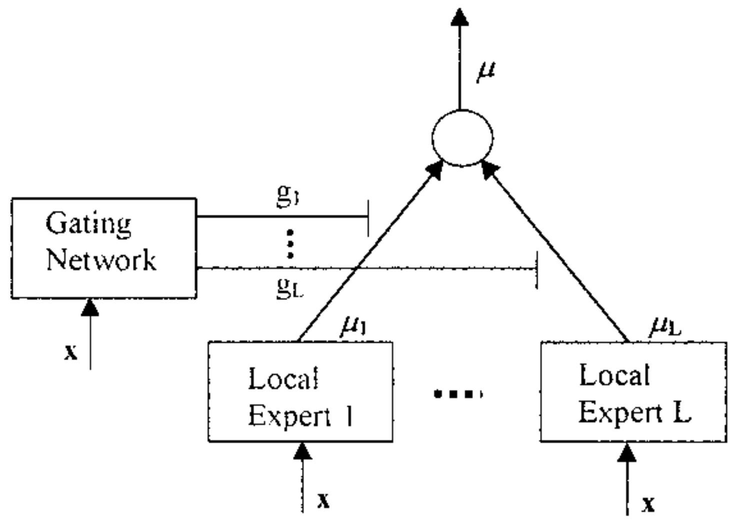
χ = $\{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)}), t=1,K, N\}$ is the set of input/output training data vectors;

s = $\{s_1, K, s_L\}$ is the set of pre-activation values of the gating network output layer;

L = the number of local experts;

W = $[\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_L]$ the matrix of weights from the gating network ($\mathbf{W}_0$) and all local experts ($\mathbf{W}_1, \ldots, \mathbf{W}_L$).

Figure 2. General modular neural network architecture for current study.



The input vectors, $\mathbf{x}^{(t)} \in \Re^n$, and corresponding desired output vectors, $\mathbf{y}^{(t)} \in \Re^m$, are the same for each of the networks. The gating network outputs are constrained to be nonnegative and to sum to one. To accomplish this, the gating network calculates its outputs using the 'softmax' function

$$g_i(\mathbf{x}^{(t)}, \mathbf{W}_0) = \frac{e^{s_i}}{\sum_{l=1}^{L} e^{l}}, i = 1, \ldots, L \text{ (Bridle 1989).}$$

The g_i assign a probability to each local expert's mapping of a given input vector, $\mathbf{x}^{(t)} \in \Re^n$, to the corresponding output vector, $\mathbf{y}^{(t)} \in \Re^m$. This assignment of probabilities partitions the input space so as to maximize the log likelihood function,

$$\ln L(\mathbf{W}, \chi) = \sum_{t=1}^{N} \ln \sum_{l=1}^{L} g_l(\mathbf{x}^{(t)}, \mathbf{W}_0) P(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, \mathbf{W}_l),$$

where $\mathbf{W}_l$ is the vector of weights from the lth local expert. The overall MNN output vector μ is a linear combination of the actual output vectors μ_1 to μ_L from the L local experts. That is,

$$\mu = \sum_{i=1}^{L} g_i \mu_i .$$

In order to train the MNN, all of the weights for the local expert and gating networks are adjusted simultaneously using the backpropagation algorithm with an appropriate learning rule.

Radial Basis Function Network

Radial basis function networks approach classification problems in a slightly different fashion than do BPNNs and MNNs. Rather than using supervised training to estimate all parameters in the network, RBF networks have a hidden layer of radial basis function kernel neurons that are trained using an unsupervised clustering algorithm such as K-means (Ceccarelli and Hounsou 1996). These neurons are known as pattern recognition (or prototype) neurons because they respond most strongly to vector patterns that come closest to matching their own vector pattern. A radial basis function is any symmetric function whose response to a given input is stronger the closer the input is to the kernel of the function (Chen and Lin 1996). That is, the radial basis functions form spherical receptive fields in $\Re^n$ for n-dimensional inputs. The radial basis function chosen for the proposed study is Gaussian, which is the most common function used for this purpose (Neruda 1995).

Figure 3 shows the architecture for the general RBF network to be used in the proposed study. It is based on the RBFNN proposed by Moody and Darken (1989). The notation used to describe the RBFNN is as follows:

n = the number of input neurons;

k = the number of pattern recognition neurons;

l_k = the Euclidean distance of an input vector from the kth pattern recognition neuron;

α = the learning rate used for updating pattern recognition centroids.

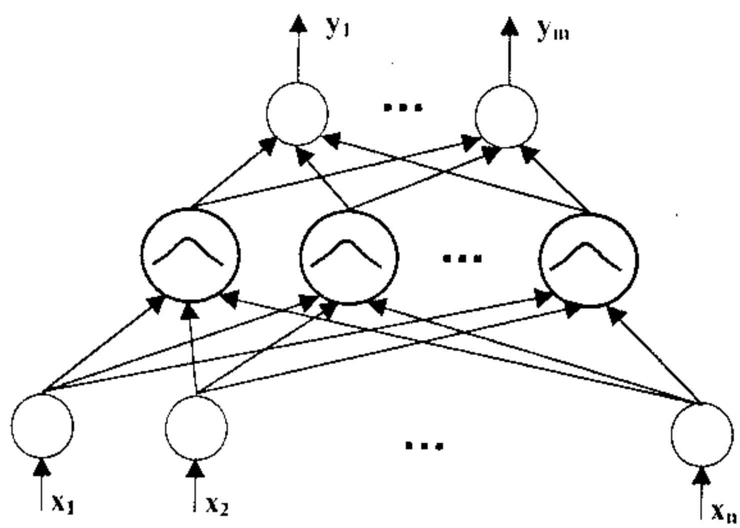
The network is fully connected and feedforward. The weights, c_{ij}, j = 1, . . ., n, from each input layer neuron to a given pattern recognition neuron form a vector, $\mathbf{c}_i$, i = 1, . . , k, representing the centroid of a cluster in the input space $\Re^n$. These centroids are set dynamically as follows:

1. Disable all cluster centroids $\mathbf{c}_1$, ..., $\mathbf{c}_k$;
2. Read an input observation $\mathbf{x}^{(t)}$ and calculate its Euclidean distance,

$$l_i^{(t)} = \| \mathbf{x}^{(t)} - \mathbf{c}_i \| = \sqrt{\sum_{j=1}^{n} (x_j - c_{ij})^2} ,$$

i = 1, . . ., number of enabled centroids, to each enabled cluster centroid. If it is not within a distance r of an existing cluster centroid, enable a pattern recognition neuron and assign it a cluster centroid, $\mathbf{c}_k = \mathbf{x}^{(t)}$, otherwise assign the input vector to the closest cluster. This process continues until either all pattern recognition neurons have been enabled or the maximum number of specified iterations has been reached.

Figure 3. General radial basis function network for current study.



After an input observation has been assigned to a cluster, its Euclidean distance is then processed by the Gaussian radial basis function,

$$v_i^{(t)} = \exp(\frac{I_i^{(t)^2}}{\sigma_i^2}), \ i = 1, \ldots, k$$

to produce the corresponding pattern recognition neuron output, where

$$\sigma_i^2 = \frac{1}{2} \sum_{p=1}^{2} \| \mathbf{c}_i - \mathbf{c}_{ip} \|$$

is calculated using the two nearest cluster centroids to the *i*th pattern recognition neuron. Next, the corresponding cluster centroid, $\mathbf{c}_k$, is updated as

$$\mathbf{c}_i^{(new)} = \mathbf{c}_i^{(old)} + \alpha(\mathbf{x}^{(t)} - \mathbf{c}_i^{(old)}),$$

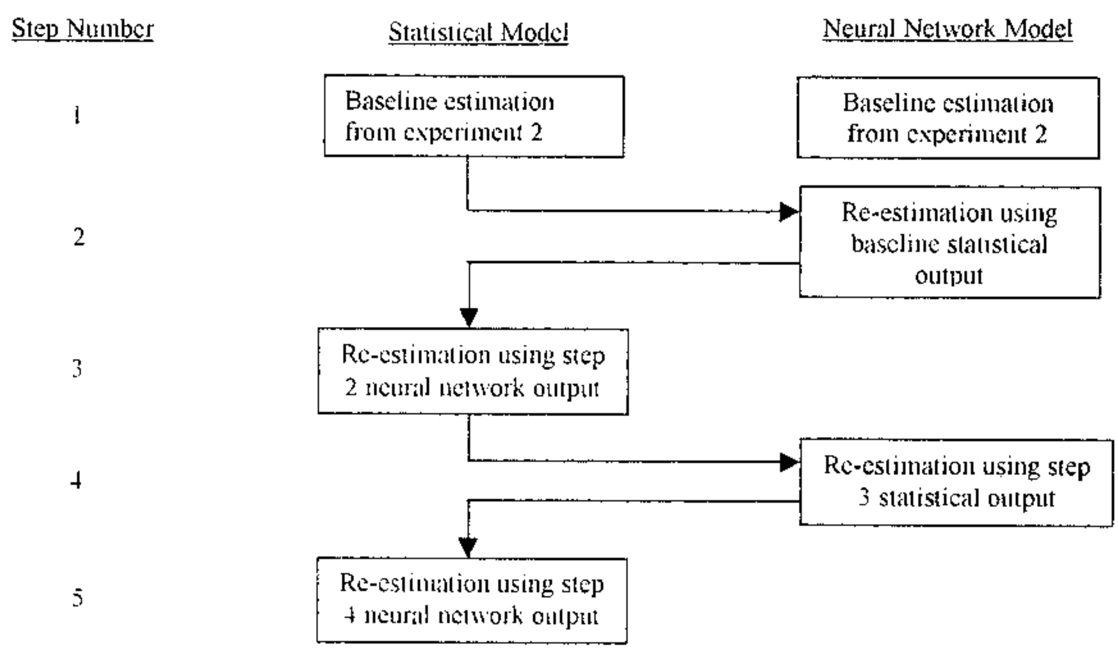
where α is reduced as the number of iterations increases. The training of the pattern recognition neurons is terminated when either the maximum number of specified iterations has been reached or the learning rate has decayed to 0.

After training of the pattern recognition neurons has been completed, the supervised training of the BPNN portion of the network begins. The training procedure is the same as that specified above for BPNNs, except the input to the neurons in the output layer of the BPNN is the weighted sum of the RBFNN outputs. During the supervised training, no adjustments are made to the pattern recognition neuron centroids.

Iterative Procedure

The iterative procedure used in this study is detailed in figure 4. It begins with a re-estimation of the relevant neural network model using as additional input, the output of a baseline (baseline means that no model outputs are used as inputs to estimate the model) statistical model. In the second step, the output from the newly estimated neural network model is used as input to the corresponding statistical model, which is in turn re-estimated. The procedure continues by re-estimating the neural network model using the output from the newly re-estimated statistical model as additional input. The new output replaces the original set of statistical output. In the last step of the iterative procedure, the most recent neural network output is used as input for statistical model re-estimation.

Figure 4. Iterative procedure for combining statistical and neural network models.



The Classification Problems

The twelve classification problems investigated include two (DFWProp and DFWEqual) involving transportation mode choice data collected by the North Central Texas Council of Governments (NCTCOG) (NCTCOG 1984). In addition to these two problems, ten simulated problems are also investigated. These simulated problems are designed to provide five levels of difficulty for the six different classification models.

Transportation Mode Choice Problems

The complete set of NCTCOG data files contains demographic information for 2,471 households from the Dallas/Fort Worth (D/FW) metroplex. Person-trip information covering 6,403 persons and 18,840 person-trips is also contained in the NCTCOG files. The data in these files are used by NCTCOG to develop the models, including mode choice, necessary for transportation planning purposes. A subset of 10,891 person-trip observations representing all home-based work trips, are used in this study. Home-based work trips are work-related trips that have a person's domicile as either the origin or destination of the trip. This category of trips was selected because home-based work trips represent the major source of peak-hour congestion on the major arterials during morning and afternoon rush hours (Rothenberg 1992). These peak-volume hours are the two most congested traffic periods on any given work day and are, therefore, of primary concern to transportation planners. The 10,891 person-trip records are contained in two files consisting of data collected from the 1984 Home Interview and 1984 On-Board Transit surveys conducted by NCTCOG in the D/FW metroplex. The person-trip data represents all of the complete home-based work trip records from the two files (4,816 from the Home Interview Survey and 6,075 from the On-Board Transit Survey).

Table 1. Number of exemplars by mode of transportation

Mode of transportation	Number of available records
Mode 1	3,591
Mode 2	892
Mode 3	4,164
Mode 4	2,244
Total	10,891

Based on a number of classification variables, the task is to assign each observation to one of four transportation modes: drive alone (Mode 1), ride share with one or more passengers (Mode 2), transit with walk access (Mode 3), or transit with drive access (Mode 4). The breakdown of the total number of person-trip records is shown in Table 1. These modes of transportation are the possible values of the dependent variable and represent four of the five modes used by NCTCOG to formulate their mode choice multinomial LOG models. The drive-alone mode includes all trips by personal vehicle (except motorcycle) made by one person alone (NCTCOG 1990). The ride share with one or more passengers mode includes all personal vehicle trips with two or more people in the vehicle. The ride share with two or more passengers mode is a combination of two modes, ride share with two passengers and ride share with three or more passengers, used by NCTCOG. These two modes were combined for the current study because there were too few trips with three or more riders to form a separate category (As it is, the combined mode has the fewest exemplars.). Both categories of transit trips refer to trips made by bus. The walk-access mode refers to trips that began with the rider walking to the bus stop while the ride access mode refers to bus trips begun by driving to the bus stop.

For the first of these mode-choice problems, DFWProp, the training, testing, combined, and validation data sets required for training and validating all neural network models, are constituted in approximately the same proportions of the four transportation mode classes as existed in the original survey sample. The creation of these data sets was accomplished using the uniform random number generator in SAS. The validation data set was created first. This was accomplished by generating a uniform random number between 0 and 1 and selecting the record for inclusion if the value of the random number was between 0 and 0.6. The data not selected for the validation data set was placed in the combined data set. The creation of the training and testing data sets followed a similar pattern. In this case, the combined data set was used to select the records for the two files and a 70 percent/30 percent split was used for the training and testing data sets, respectively. The file creation procedure was repeated twenty times in order to satisfy the data needs for the desired number of replications.

For the second of the mode-choice problems, DFWEqual, the proportions of the four classes were made approximately equal in the training, testing, combined, and validation data sets. To accomplish this, the original population of person-trip records was reduced in size so that there was an equal number of records in each of the four classes. The reduction was effected by using Bernoulli random variables in Excel with appropriate probabilities to select 892 records from each of modes 1 ($p = 0.248399$), 3 ($p = 0.214217$), and 4 ($p = 0.397504$) and including all 892 available mode 2 records.

This provided the largest possible set of observations from which to construct the required data sets, while still maintaining the same number of records for each mode of transportation. From this point, the process used to build the training, testing, combined, and validation data sets was identical to the one used to construct the proportional data sets.

Eleven independent variables are included in this study. They are:

X_OR_MI - x coordinate of travel zone in which the trip originated

Y_OR_MI - y coordinate of travel zone in which the trip originated

X_DS_MI - x coordinate of travel zone in which the trip terminated

Y_DS_MI - y coordinate of travel zone in which the trip terminated

MAN_DIS - Manhattan distance between origin and destination zones

STARTMIL - start time of the trip in military time

ARRIVEMIL - arrival time at the destination in military time

TIME - duration of the trip in minutes

PEOPLE - the number of people residing in the household

CARS - the number of available vehicles at the household

INCOME - the combined income of all persons in the household (Income is an ordinal variable with ten categories and is coded using nine variables in a thermometer format.).

These variables were selected because they were common to both surveys (or could be created from information given), were easy to generate, and because they have been shown to be related to transportation mode choice decisions (Ortuzar and Willumsen 1994). They do not match precisely the variables used by NCTCOG in their mode choice models, but some of the NCTCOG mode choice variables (e.g., access/egress time) were not part of the survey data and were not available for use in this study (NCTCOG 1990).

Simulation Problems

The remaining ten classification problems use data generated via Monte Carlo simulation. All needed random variables were generated using the normal random number generator in SAS. The data for the ten simulation based classification problems are in one of five different configurations. Table 2 presents the parametric characteristics of each of the five simulation data configurations. The first two simulation configurations match exactly those used by Patuwo, Hu, and Hung (1993). Simulations 3 and 4 are logical extensions of the first two simulations designed to make the classification task more difficult. Simulation 5 is unique in that it makes use of a mixture distribution that is mostly contained within the vector space of another population. All of the main groups and contamination portions of the data generated are bivariate normal. All five configurations represent two class problems. Figures 5a-5e display graphically the five simulated data configurations. Each of the five configurations represents a slightly different challenge for the six classifiers used in this study.

Table 2. Description of the five simulated data configurations.

	Population 1		Population 2[a]	
Configuration	Main Group	Contamination	Main Group	Contamination
SIM1	N((5.5), 5.5, .3)[b]	none	N((15.5), 5, 5, .3)	none
SIM2	N((5.5), 5.5, .3)	none	N((15,5), 15, 5, .3)	none
SIM3	N((5,5), 5.5, .3)	none	N((15,5), 5, 5, .3)	N((-2.5.5),5,5,.3)[c]
SIM4	N((5.5), 5,5, .3)	N((22.5.5), 5, 5,.3)[c]	N((15,5), 5, 5, .3)	N((-2.5,5),5,5,.3)[c]
SIM5	N((2.5,2.5), 2,2, 0)	none	N((0,0), $(1.5)^{0.5}$, $(1.5)^{0.5}$,0)[d]	N((5,5), $(1.5)^{0.5}$, $(1.5)^{0.5}$, 0)[d]

[a] For the five proportional simulation problems, population 2 contains 70 percent of the total number of exemplars.

[b] The notation used is as follows: N((μ_1,μ_2), σ_1, σ_2, ρ).

[c] The contamination portions contain 30 percent of the exemplars generated from the given population.

[d] Each portion contains 50 percent of the exemplars generated from population 2.

Five of the problems, one from each of the five data configurations, use data containing equal numbers of observations from each of two classes. This is called the "equal' condition. Two hundred observations (100 from population 1 and 100 from population 2) were generated for the combined data set used for estimating the parametric model parameters. These 200 exemplars were then randomly divided using SAS generated uniform random variables so that 70% of the exemplars were placed in the corresponding training file and the remaining 30%of the 200 exemplars were placed in the corresponding testing file. The validation data set is comprised of 2000 exemplars, 1000 each from populations 1 and 2. Thirty sets of training, testing, combined, and validation files were created to satisfy the data requirements for the desired number of replications.

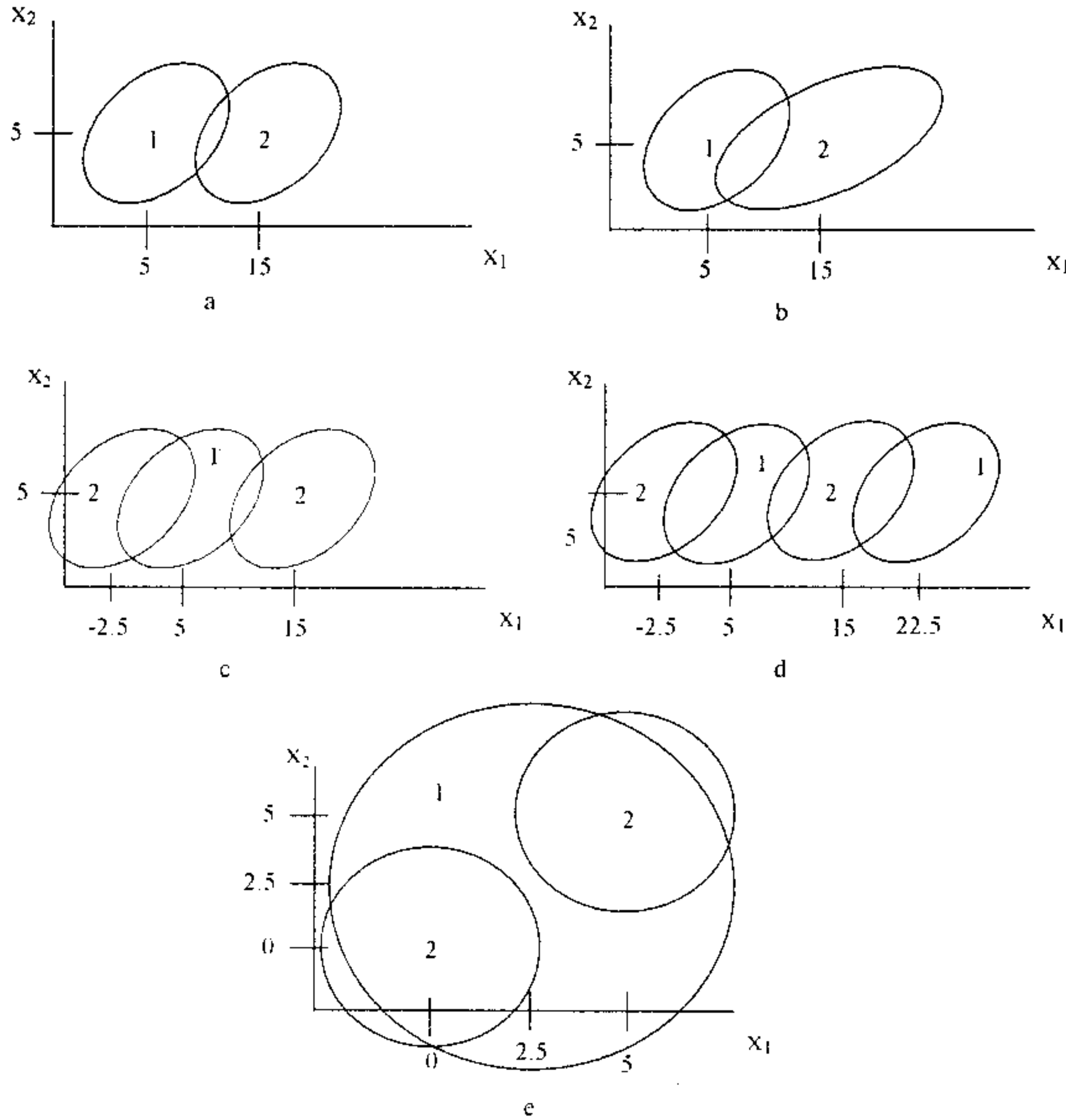
The other five problems use data that are split so that 30% of the exemplars come from population 1 and 70% come from population 2. This is called the 'proportional' condition. The total numbers of exemplars used to create the training, testing, combined, and validation data sets were the same as in the equal condition. The differences in the file structures consist only in the proportions of exemplars generated from the two populations.

In the first two problems, SIM1a and SIM1b (figure 5a), correlated bivariate normal random variables with equal covariance structures was generated for each of the two groups. The next two problems, SIM2a and SIM2b (figure 5b) use correlated bivariate normal random variables with unequal covariance structures for the two groups. For classification problems SIM3a and SIM3b (figure 5c), correlated bivariate normal random variables were generated in the same manner as for problems SIM1a and SIM1b. In addition, contamination was added to the second group by generating correlated bivariate normal random variables with a different mean from the main group from population 2. The contamination changes the covariance structure of group 2 from that in simulation 1. In problems SIM4a and SIM4b, both groups are contaminated as shown in figure 5d. The main groups and the contamination for population 2 for these two problems were generated from the same populations as were used for problems SIM3a and SIM3b. The contamination for population 1 is also correlated bivariate normal and has the same covariance structure as its corresponding main group. This means that the population 1 data and the population 2 data have the same covariance structures. The last two problems, SIM5a and SIM5b, use uncorrelated bivariate normal data for population 1 and a mixture distribution comprised of two uncorrelated bivariate normal distributions, both having the same covariance structure but different means. The resulting data sets have the form shown in figure 5e. In this case, unlike the other four simulation conditions, the data from population 2 are almost entirely contained in the population 1 vector space (see Figure 5e on page 55).

Experimental Designs

In order to investigate the twelve classification problems and adequately address the stated research questions, 47 experiments were performed. The 47 designs for these experiments can be seen in tables 3-10. Each experiment was structured using a randomized block design (Winer, Brown, and Michels 1991) with 20 blocks for all experiments using the D/FW transportation mode choice data and 30 blocks for all simulation experiments. The blocks represent particular realizations of the data configurations. The first six experiments (table 3) were designed to test whether the six classification methods differ in their baseline classification performance when the population proportions and therefore the sample proportions of all of the classes are equal. These comparisons were made for each of the six data configurations. The dependent variable is the proportion of observations from the validation sample classified correctly. The design for the second six experiments (table 4) was identical to that of the first six except that, the population proportions and therefore the sample class proportions are not equal. These six experiments are intended to test the baseline performance of the six classification methods under conditions of unequal population proportions.

Figure 5. Five simulation study population distributions: a, equal covariance structures (SIM1a and SIM1b); b, unequal covariance structures (SIM2a and SIM2b); c, group two contaminated, unequal covariance structures (SIM3a and SIM3b); d, both groups contaminated, equal covariance structures (SIM4a and SIM4b); e, both groups uncorrelated, group two bimodal (SIM5a and SIM5b).



The remaining 35 experiments (tables 5-10) also used randomized block designs with 20 blocks for all D/FW transportation mode choice experiments and 30 blocks for all simulation experiments. In these experiments, the iterative procedure for combining the statistical and neural network models was tested. In each of these experiments, the baseline performance of a given statistical or neural network model was compared to the first and second iteration results of that model coupled with each of the appropriate complementary models. For example, table 5 shows the designs used to examine the performance of the LOG models. For the D/FW transportation mode choice data, LOG models were paired only with BPNN and MNN models for the iterative procedure. For the five simulation conditions, LOG models were paired with all three types of neural network models for the iterative procedure. Similar comparisons were made for the LDA, QDA, BPNN, MNN, and RBFNN models. In each case, the dependent measure is classification rate. Tables 3 through 10 are shown on the following pages.

Table 3. Experimental designs for the equal class proportions investigation.

Data		Classification Method					
Configuration	Replication	LOG	LDA	QDA	BPNN	MNN	RBFNN
DFWEqual	1 20		. . . . Classification rate				
SIM1a	1 30		. . . . Classification rate				
SIM2a	1 30		. . . . Classification rate				
SIM3a	1 30		. . . . Classification rate				
SIM4a	1 30		. . . . Classification rate				
SIM5a	1 30		. . . . Classification rate				

Table 4. Experimental designs for the unequal class proportions investigation.

Data Configuration	Replication	Classification Method					
		LOG	LDA	QDA	BPNN	MNN	RBFNN
DFWProp	1 … 20			.... Classification rate			
SIM1b	1 … 30			.... Classification rate			
SIM2b	1 … 30			.... Classification rate			
SIM3b	1 … 30			.... Classification rate			
SIM4b	1 … 30			.... Classification rate			
SIM5b	1 … 30			.... Classification rate			

Table 5. Experimental design for the iterative combined models investigation using LOG paired with BPNN, MNN, and RBFNN.

Data Configuration	Replication	Classification Method: LOG						
		Baseline	BPNN1	BPNN2	MNN1	MNN2	RBFNN1	RBFNN2
DFWProp	1 … 20				.... Classification rate			
SIM1b	1 … 30				.... Classification rate			
SIM2b	1 … 30				.... Classification rate			
SIM3b	1 … 30				.... Classification rate			
SIM4b	1 … 30				.... Classification rate			
SIM5b	1 … 30				.... Classification rate			

Table 6. Experimental design for the iterative combined models investigation using LDA paired with BPNN, MNN, and RBFNN.

Data Configuration	Replication	Classification Method: LDA						
		Baseline	BPNN1	BPNN2	MNN1	MNN2	RBFNN1	RBFNN2
DFWProp	1 20	. . . . Classification rate						
SIM1b	1 30	. . . . Classification rate						
SIM2b	1 30	. . . . Classification rate						
SIM3b	1 30	. . . . Classification rate						
SIM4b	1 30	. . . . Classification rate						
SIM5b	1 30	. . . . Classification rate						

Table 7. Experimental design for the iterative combined models investigation using QDA paired with BPNN, MNN, and RBFNN.

Data Configuration	Replication	Classification Method: QDA						
		Baseline	BPNN1	BPNN2	MNN1	MNN2	RBFNN1	RBFNN2
DFWProp	1 20	. . . . Classification rate						
SIM1b	1 30	. . . . Classification rate						
SIM2b	1 30	. . . . Classification rate						
SIM3b	1 30	. . . . Classification rate						
SIM4b	1 30	. . . . Classification rate						
SIM5b	1 30	. . . . Classification rate						

Table 8. Experimental design for the iterative combined models investigation using BPNN paired with LOG, LDA, and QDA.

Data Configuration	Replication	Classification Method: BPNN						
		Baseline	LOG1	LOG2	LDA1	LDA2	QDA1	QDA2
DFWProp	1							
		. . . . Classification rate						
	20							
SIM1b	1							
		. . . . Classification rate						
	30							
SIM2b	1							
		. . . . Classification rate						
	30							
SIM3b	1							
		. . . . Classification rate						
	30							
SIM4b	1							
		. . . . Classification rate						
	30							
SIM5b	1							
		. . . . Classification rate						
	30							

Table 9. Experimental design for the iterative combined models investigation using MNN paired with LOG, LDA, and QDA.

Data Configuration	Replication	Classification Method: MNN						
		Baseline	LOG1	LOG2	LDA1	LDA2	QDA1	QDA2
DFWProp	1							
		. . . . . . Classification rate						
	20							
SIM1b	1							
		. . . . Classification rate						
	30							
SIM2b	1							
		. . . . Classification rate						
	30							
SIM3b	1							
		. . . . Classification rate						
	30							
SIM4b	1							
		. . . . Classification rate						
	30							
SIM5b	1							
		. . . . Classification rate						
	30							

Table 10. Experimental design for the iterative combined models investigation using RBFNN paired with LOG, LDA, and QDA.

Data Configuration	Replication	Classification Method: RBFNN Baseline	LOG1	LOG2	LDA1	LDA2	QDA1	QDA2
SIM1b	1 … 30	.... Classification rate						
SIM2b	1 … 30	.... Classification rate						
SIM3b	1 … 30	.... Classification rate						
SIM4b	1 … 30	.... Classification rate						
SIM5b	1 … 30	.... Classification rate						

Experimental Procedure

Each of the twelve experiments (six equal and six proportional) investigating the baseline performance of the six classification methods used the same basic procedure. To begin, 20 sets of randomly generated parameter estimation and validation files for the DF/W transportation mode choice configuration and 30 sets for each of the five simulation configurations were generated as described above. An assumption made was that the number of total observations in each data set is the number that would be available in a corresponding real-world application (This is obviously true for the DFWEqual and DFWProp conditions since the observations available for use came from actual survey results.).

The next step was to estimate the parameters for each of the six models for each of the replicated data sets. The three statistical models were estimated using SAS for

Windows v6.12. To estimate the LOG model, procedure LOGISTIC was used. For the LDA and QDA models procedure DISCRIM was used. For both the LDA and QDA models, prior probabilities were used to adjust the constant of the discriminant function for the six proportional classes experiments. For all statistical models, the combined data sets were used for parameter estimation and the corresponding validation data sets were used to test the classificatory performance of the models.

For estimating the neural network models, NeuralWare's NeuralWorks Professional II/PLUS Win32 was used. The settings for the various training parameters for each of the neural network models were determined as a result of past experience or experimentation. For example, in order to determine a viable set of BPNN training parameter values for the simulation 1 proportional condition, 160 models using different combinations of parameter settings were trained and tested. The following training parameters were varied systematically: the number of hidden layer neurons (from 1 to 10), the training rule (generalized delta rule versus cumulative delta rule), the transfer function (sigmoid versus hyperbolic tangent), and the input neuron connection scheme (whether or not the input neurons, in addition to being connected to the hidden layer, were connected directly to the output layer). The final training parameter values were those that allowed the specific network to generalize well. Since there are literally tens of thousands of possible parameter value combinations to be tested, it is impossible to optimize the training parameter settings for even a single validation sample. The best that can be expected is that the final training parameter settings chosen will result in acceptable network performance. Once the network training parameters were set, those settings were maintained for each of the replications for the given classification problem and model. Some of the training parameter settings did differ for one or more of the twelve baseline classification problems.

Both the training and corresponding testing files were used for estimating the weights, i.e. the model parameters, for all neural networks. Exemplars from the training file were randomly presented to the network and the error generated by each exemplar was propagated backward through the network so that the weights could be appropriately adjusted. After 100 exemplars had been presented, the testing file was presented to the current network and the network's classificatory performance was noted. If the network's classificatory performance was better than the network's baseline performance against the testing file, the current network parameters were saved. This procedure was repeated until the total number of desired training exemplars (between 10,000 and 32,000 depending on the type of network, the data configuration, and the iteration) had been presented. NeuralWorks Professional II uses a function called SaveBest to perform this procedure automatically. The final weights for a given neural network resulting from this procedure, provide the best classificatory performance of all the networks tested against the testing data set used by SaveBest.

The remaining 35 experiments involved extensions, using the previously defined iterative procedure, of the six proportional classes baseline experiments. The iterative procedure is an extension of Markham and Ragsdale (1995) and is intended to enhance the performance of both the statistical and neural network models over their baseline performance as determined in the six proportional classes experiments. The data sets used in the six baseline experiments were also used for the iterative procedure experiments. The same was true of the parameter settings for each of the neural network models except, in some cases, the number of hidden layer neurons. The classification rate against the validation data sets was used as the performance measure for each of the re-estimated models.

CHAPTER 4

DATA ANALYSIS AND RESULTS

This chapter presents the data analysis for and results of the 47 experiments described in chapter 3. All data analyses were performed using Microsoft Excel 97 SR-1. Macros, using the Excel two-factor analysis of variance (ANOVA) without replication (i.e., randomized block design) procedure and separately programmed Tukey post-hoc comparison procedure, were used to analyze each of the experimental results. An alpha of 0.05 is used for all ANOVAs, multiple comparisons, and t-tests performed. The *p*-values are reported for the 47 ANOVAs. For comparison purposes, the Tukey post-hoc tests were also performed at alpha = 0.01. Both sets of Tukey results for each experiment are presented below. The following sections detail the analyses and results of baseline equal condition, baseline proportional condition, and iterative procedure experiments. The results of comparisons between the baseline equal condition performance and baseline proportional condition performance of a given model under a specific data configuration follows.

Equal Condition Experiments

Table 11 shows the results of the ANOVA and post-hoc comparisons for the simulation 1 equal condition. Overall, the six classification models perform reasonably well on what is a linearly separable problem. The difference in classificatory

Table 11. ANOVA and Tukey results for simulation 1 equal condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.011045717	29	0.000380887	7.385952556	0.000000000
Models	0.007822067	5	0.001564413	30.336265828	0.000000000
Error	0.007477517	145	0.000051569		
Total	0.026345300	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.006385031	0.01

Means joined by a line are not significantly different.

RBFNN1c	BPNN1c	MNN1c	QDA1c	LOG1c	LDA1c
0.8343	0.8397	0.8408	0.8507	0.8508	0.8509

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.005375488	0.05

Means joined by a line are not significantly different.

RBFNN1c	BPNN1c	MNN1c	QDA1c	LOG1c	LDA1e
0.8343	0.8397	0.8408	0.8507	0.8508	0.8509

performance from the best to worst models is only 1.66%. The reason that the performance of the models is not better is because of a combination of sampling error and some degree of overlap of the two populations. The ANOVA results indicate that there are significant differences in classificatory performance for both the classification models and replications. The post-hoc comparisons indicate that the three parametric models are statistically better than the three neural network models but not statistically different from each other. This is true at both alpha = 0.05 and alpha = 0.01. Although LDA is Bayes optimal under the simulation 1 configuration and the LOG and QDA models are not, their performances are indistinguishable. This is likely because of the large sample sizes used. As for the neural network models, the MNN and BPNN do not differ statistically, but both are significantly better than the RBFNN.

The results for the simulation 2 equal condition can be found in table 12. As for simulation 1, the overall ANOVA results indicate that there are significant differences among both the classification models and across replications. The overall performance of the six models is poorer than under simulation 1. The increase in the variance of population 2 results in the amount of group overlap being much greater in the simulation 2 configuration. The simulation 2 configuration is Bayes optimal for QDA and , as expected, QDA performs significantly better (as much as 5.98% better) than all other models at both alpha levels. The MNN and RBFNN models do not differ and neither do the RBFNN and BPNN models. The MNN, however, is significantly better than the BPNN, even at alpha = 0.01. This is a first indication that the predominantly used BPNN may not be the best neural network model under all classification conditions. When the covariance structures of the classes differ, the LDA and LOG models, which were optimal for the simulation 1 configuration, drop off in their performance relative to all other models.

Table 12. ANOVA and Tukey results for simulation 2 equal condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.015991974	29	0.000551447	2.383271313	0.000397739
Models	0.081982740	5	0.016396548	70.863377050	0.000000000
Error	0.033550468	145	0.000231383		
Total	0.131525182	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.013524886	0.01

Means joined by a line are not significantly different.

LOG2e	LDA2e	BPNN2e	RBFNN2e	MNN2e	QDA2e
0.7267	0.7322	0.7581	0.7655	0.7729	0.7865

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.011386455	0.05

Means joined by a line are not significantly different.

LOG2e	LDA2e	BPNN2e	RBFNN2e	MNN2e	QDA2e
0.7267	0.7322	0.7581	0.7655	0.7729	0.7865

The simulation 3 configuration introduces contamination, albeit rather systematic contamination, to the simulation 1 classification problem. The data are no longer linearly separable but require a quadratic form to best separate them. Under this configuration, the performance of the LDA and LOG models, the optimal linear classifiers, again fall significantly relative to all of the other models (see table 13). The spread between the best and worst models jumps to 8.89%. The QDA and MNN models do not differ statistically, but they are both significantly better than the other four models. Once again, the MNN was better than the BPNN, even at alpha = 0.01.

Table 13. ANOVA and Tukey results for simulation 3 equal condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.016194290	29	0.000558424	2.020029003	0.0036152 49
Models	0.248316490	5	0.049663298	179.650835079	0.0000000 00
Error	0.040084301	145	0.000276443		
Total	0.304595082	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.014783303	0.01

Means joined by a line are not significantly different.

LOG3e	LDA3e	BPNN3e	RBFNN3e	MNN3e	QDA3e
0.6712	0.6716	0.7384	0.7405	0.7552	0.7601

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.012445902	0.05

Means joined by a line are not significantly different.

LOG3e	LDA3e	BPNN3e	RBFNN3e	MNN3e	QDA3e
0.6712	0.6716	0.7384	0.7405	0.7552	0.7601

Table 14. ANOVA and Tukey results for simulation 4 equal condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.132357840	29	0.004564063	2.759236422	0.000037327
Models	1.185023824	5	0.237004765	143.282884870	0.000000000
Error	0.239845051	145	0.001654104		
Total	1.557226715	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.036161773	0.01

Means joined by a line are not significantly different.

LDA4e	LOG4e	QDA4e	MNN4e	RBFNN4e	BPNN4e
0.4895	0.4896	0.4990	0.6375	0.6551	0.6689

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.030444203	0.05

Means joined by a line are not significantly different.

LDA4e	LOG4e	QDA4e	MNN4e	RBFNN4e	BPNN4e
0.4895	0.4896	0.4990	0.6375	0.6551	0.6689

The results, shown in table 14, for the simulation 4 equal condition indicate an overwhelming superiority, by as much as 17.94%, of the three neural network models over their parametric counterparts and overall significance for both models and replications. This configuration is the same as simulation 3 except for an additional layer of contamination added to group 1. This additional contamination results in classes that are neither linearly nor quadratically separable. It appears as though the parametric models are separating the exemplar space in the same manner as in simulation 1, thereby misclassifying all of the contamination exemplars. In addition, because the sample variances for the two groups are larger than for simulation 1, there is more variation in the placement of the discriminant function and therefore poorer performance against the main groups from the two populations. Although the performance of the neural networks drops off relative to simulation 3, they are all better able to identify the non-overlapping pieces of the two classes than are the parametric models. Their poorer performance in simulation 4 relative to simulation 3 is likely due to the increased overlap of the two classes. Under simulation 4, the BPNN model performance, although not different from that of the RBFNN, is significantly better than the performance of the MNN model.

Table 15 shows the results for the simulation 5 equal condition. The table indicates, as is the case for each of the other four simulation configurations, that both models and replications are significant. Under this configuration, the results of the post-hoc tests are the same for both alpha = 0.05 and alpha = 0.01. The three neural network models and the QDA model do not differ significantly but as with simulation 4, the LOG and LDA models place essentially all of the observations into only one of the groups. The difference between the best and worst performers is 27.9%, which is nearly 10% greater than for simulation 4.

Table 15. ANOVA and Tukey results for simulation 5 equal condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.015195994	29	0.000524000	1.569767469	0.044257062
Models	3.018461528	5	0.603692306	1808.505512849	0.000000000
Error	0.048402056	145	0.000333807		
Total	3.082059578	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.01624487	0.01

Means joined by a line are not significantly different.

LDA5e	LOG5e	QDA5e	RBFNN5e	MNN5e	BPNN5e
0.5028	0.5029	0.7731	0.7757	0.7793	0.7818

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.013676379	0.05

Means joined by a line are not significantly different.

LDA5e	LOG5e	QDA5e	RBFNN5e	MNN5e	BPNN5e
0.5028	0.5029	0.7731	0.7757	0.7793	0.7818

The last of the equal condition experiments to be discussed is the one using D/FW transportation mode choice data. Unlike the simulated data, the multivariate distribution of the mode choice data is unknown. Although each of the variables alone is easy to understand, jointly, they represent a less well defined classification problem than do the simulations. In addition, classification complexity is increased because there are four

Table 16. ANOVA and Tukey results for D/FW transportation mode choice equal condition

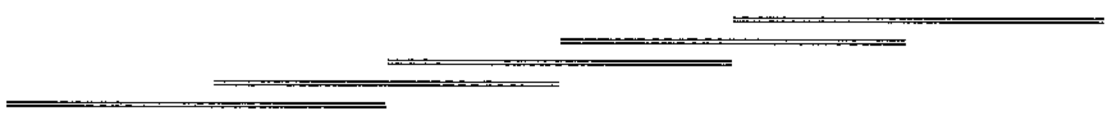
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.010353366	19	0.000544914	3.571407625	0.000019459
Models	0.032960897	5	0.006592179	43.205645420	0.000000000
Error	0.014494797	95	0.000152577		
Total	0.057809060	119			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.013451113	0.01

Means joined by a line are not significantly different.

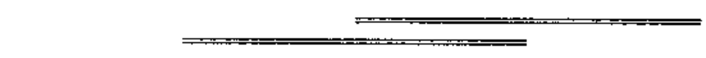


BPNNdc	RBFNNdc	LDAdc	LOGdc	QDAdc	MNNdc
0.5380	0.5504	0.5595	0.5676	0.5757	0.5889

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.011324346	0.05

Means joined by a line are not significantly different.



BPNNdc	RBFNNdc	LDAdc	LOGdc	QDAdc	MNNdc
0.5380	0.5504	0.5595	0.5676	0.5757	0.5889

groups instead of two The realism of this configuration is diminished because the equal condition was artificially created. The natural groups have proportion differences as great as 467%. The results, shown in table 16, indicate that both the models and replications are significant and that all of the models are able to perform above the 50% level, which would be the level of performance expected if all observations were placed in just one of the two classes. The best performing model is the MNN which is significantly better than all of the other models and is 5.09% better than the worst model (BPNN). Moreover, the BPNN model's performance is statistically lower than all of the other models. Under this configuration, the QDA, which is second only to the MNN, is again significantly better than LDA and, as is the case in each of the five simulation experiments, there is no statistical difference between the LOG and LDA models.

Baseline Proportional Condition Experiments

Although the overall results for simulation 1 are significant for both models and replications (see table 17), the total spread from the best technique, LOG, to the worst, RBFNN, is only 1.59%. As is the case under the equal condition, the three parametric models performed significantly better than all three neural network models, but do not perform significantly differently from each other. At alpha = 0.01 the best neural network model, BPNN, does not differ significantly from the parametric models. Among the neural networks, BPNN and MNN do not differ. Both, however, differ from the RBFNN at alpha = 0.05 and alpha = 0.01.

Although the level of performance declines for each of the six classification models in the simulation 2 baseline proportional condition when compared to the simulation 2 equal configuration, the overall results are still significant for models. The results, shown in table 18, are also significant for replications. The ordering of the six models' performances under both configurations remains the same except for LDA and

Table 17. ANOVA and Tukey results for simulation 1 baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.012553540	29	0.000432881	8.529493476	0.000000000
Models	0.005937140	5	0.001187428	23.397115812	0.000000000
Error	0.007358901	145	0.000050751		
Total	0.025849582	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.006334186	0.01

Means joined by a line are not significantly different.

RBFNN1u	MNN1u	BPNN1u	QDA1u	LDA1u	LOG1u
0.8500	0.8585	0.8597	0.8656	0.8658	0.8659

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.005332682	0.05

Means joined by a line are not significantly different.

RBFNN1u	MNN1u	BPNN1u	QDA1u	LDA1u	LOG1u
0.8500	0.8585	0.8597	0.8656	0.8658	0.8659

LOG, which do not differ statistically. The QDA model is once again significantly better than the other five models and the MNN is significantly better than the BPNN. The RBFNN model does not differ from the MNN, but is significantly better than the BPNN model. The absolute difference from the best (QDA) to worst (LOG) model increases to 11.68%, nearly double the 5.98% spread obtained under the simulation 2 equal condition.

Table 18. ANOVA and Tukey results for simulation 2 baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Rows	0.018447196	29	0.000636110	1.789965870	0.013581947
Columns	0.383241079	5	0.076648216	215.682267127	0.000000000
Error	0.051529462	145	0.000355376		
Total	0.453217737	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.016761471	0.01

Means joined by a line are not significantly different.

LOG2u	LDA2u	BPNN2u	RBFNN2u	MNN2u	QDA2u
0.6447	0.6462	0.7215	0.7369	0.7407	0.7615

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.0141113	0.05

Means joined by a line are not significantly different.

LOG2u	LDA2u	BPNN2u	RBFNN2u	MNN2u	QDA2u
0.6447	0.6462	0.7215	0.7369	0.7407	0.7615

As is the case for all of the previously described experiments, both models and replications are significant under the simulation 3 baseline proportional condition (see table 19). The three neural network models, although not statistically different from each other, are significantly superior than the parametric models at both alpha = 0.05 and alpha

Table 19. ANOVA and Tukey results for simulation 3 baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.033603140	29	0.001158729	2.554135875	0.000136607
Models	0.410487224	5	0.082097445	180.963826124	0.000000000
Error	0.065781818	145	0.000453668		
Total	0.509872182	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.018938138	0.01

Means joined by a line are not significantly different.

LDA3u	LOG3u	QDA3u	RBFNN3u	BPNN3u	MNN3u
0.6469	0.6489	0.7239	0.7464	0.7539	0.7610

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.015943812	0.05

Means joined by a line are not significantly different.

LDA3u	LOG3u	QDA3u	RBFNN3u	BPNN3u	MNN3u
0.6469	0.6489	0.7239	0.7464	0.7539	0.7610

Table 20. ANOVA and Tukey results for simulation 4 baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.012127179	29	0.000418179	2.240503550	0.000959189
Models	0.010816946	5	0.002163389	11.590935551	0.000000002
Error	0.027063513	145	0.000186645		
Total	0.050007637	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.012147206	0.01

Means joined by a line are not significantly different.

QDA4u	LOG4u	LDA4u	MNN4u	RBFNN4u	BPNN4u
0.6977	0.6996	0.6997	0.7075	0.7113	0.7194

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.010226601	0.05

Means joined by a line are not significantly different.

QDA4u	LOG4u	LDA4u	MNN4u	RBFNN4u	BPNN4u
0.6977	0.6996	0.6997	0.7075	0.7113	0.7194

= 0.01. Each of the three neural network models improves their classificatory performance over that in the simulation 3 equal condition. The QDA model, as was the case under the corresponding equal condition, is statistically superior to the LOG and LDA models. It does, however, drop from being the best performer to being the fourth best. The difference, in classificatory performance, between the best (MNN) and worst (LDA) models is 11.41%.

As is the case for the simulation 4 equal condition (see table 14), the three neural network models under the simulation 4 proportional condition have higher classification rates than the three parametric models (see table 20). Only the BPNN and RBFNN models are significantly better than the best statistical model, LDA. Unlike most of the previous results, the BPNN model is significantly better at alpha = 0.05 than the MNN model. The absolute difference between them, however, is only 1.17%. The gap between the best model (BPNN) and the worst model (QDA) has decreased from 17.94% to only 2.17%. The large improvement in the three parametric models' performance is likely due to the reduction in the number of population 1 exemplars. The resulting population 1 and population 2 samples begin to resemble those of the simulation 3 configuration.

The simulation 5 baseline proportional condition results are shown in table 21. The three neural network models perform significantly better than the three parametric models. The MNN model is best and is statistically superior to the BPNN model at both alpha = 0.05 and alpha = 0.01. The RBFNN model's performance falls between the other two neural network types and is not significantly different from either of them. The QDA model, as in the simulation 5 equal condition (see table 15), is once again statistically superior to both the LOG and LDA models. Furthermore, the QDA model's classificatory performance drops by 1.03%, while the LOG and LDA models improve their performances by 19.79% and 19.8%, respectively. This large increase in the LOG and LDA performance has reduced the difference between the best model (MNN) and the worst model (LDA), to only 12.57% from what had been a 27.9% difference under the corresponding equal condition.

Table 21. ANOVA and Tukey results for simulation 5 baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.007331778	29	0.000252820	2.181346719	0.001375386
Models	0.524893711	5	0.104978742	905.763405892	0.000000000
Error	0.016805622	145	0.000115901		
Total	0.549031111	179			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.009572197	0.01

Means joined by a line are not significantly different.

LDA5u	LOG5u	QDA5u	BPNN5u	RBFNN5u	MNN5u
0.7008	0.7008	0.7628	0.8163	0.8205	0.8265

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.008058729	0.05

Means joined by a line are not significantly different.

LDA5u	LOG5u	QDA5u	BPNN5u	RBFNN5u	MNN5u
0.7008	0.7008	0.7628	0.8163	0.8205	0.8265

The D/FW transportation mode choice baseline proportional condition matches the proportions of exemplars from each of the four transportation mode choices to those that occur in the original sample. Under this more realistic condition, all six models show greatly improved performance when compared to their performance under the corresponding equal condition (see tables 16 and 22). As in the parallel equal condition,

Table 22. ANOVA and Tukey results for D/FW transportation mode choice baseline proportional condition

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.001829189	19	0.000096273	1.782364340	0.036044303
Models	0.053564421	5	0.010712884	198.334310717	0.000000000
Error	0.005131356	95	0.000054014		
Total	0.060524966	119			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.87	0.008003284	0.01

Means joined by a line are not significantly different.

LDAdu	RBFNNdu	LOGdu	QDAdu	BPNNdu	MNNdu
0.6872	0.6993	0.7036	0.7055	0.7346	0.7479

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.1	0.006737878	0.05

Means joined by a line are not significantly different.

LDAdu	RBFNNdu	LOGdu	QDAdu	BPNNdu	MNNdu
0.6872	0.6993	0.7036	0.7055	0.7346	0.7479

the MNN model is still significantly better than all other models (this is true even at alpha = 0.01). Under this configuration, the BPNN model, which increases its performance by 19.66%, becomes the second best performer, and is significantly better than the remaining four models at both alpha = 0.05 and alpha = 0.01. The spread in performance from the best (MNN) to the worst (LDA) models is 6.07%.

Iterative Procedure Experiments

The iterative procedure was performed only under the proportional condition. This was done primarily because of an assumption that in most real-world circumstances, the number of items to be classified is not the same for each of the possible classes. This contention is supported by the D/FW transportation mode choice data set used in this study. Although the 467% proportion difference between the largest and smallest classes in this data set may be large compared to some real-world conditions, there are few naturally occurring phenomena in which equal numbers derive for each of the classes.

The results of the comparisons are presented in tables 23 - 57. The discussion of the results is organized by data configuration beginning with simulation 1 and ending with the D/FW transportation mode choice configuration. Within each data configuration, the models are presented in the same order: LOG, LDA, QDA, BPNN, MNN, and RBFNN. The results presented in all of the tables are presented smallest to largest, left to right, to only four significant digits. The following coding scheme should be used to interpret the column headings for the Tukey post-hoc comparison results presented in tables 23 - 57:

Baseline results:	LOG	= logit;
	LDA	= linear discriminant analysis;
	QDA	= quadratic discriminant analysis;
	BPNN	= backpropagation neural network;
	MNN	= modular neural network;
	RBFNN	= radial basis function neural network.
Column 1:	Data configuration	= 1 for simulation 1;
		= 2 for simulation 2;
		= 3 for simulation 3;
		= 4 for simulation 4;
		= 5 for simulation 5;
		= d for transportation mode choice.
Column 2:	Proportional	= u.
Column 3:	Iteration	= 1 for iteration 1;
		= 2 for iteration 2.
Column 4	Output model	= o for LOG;
		= f for LDA;
		= q for QDA;
		= b for BPNN;
		= m for MNN;
		= r for RBFNN.
Column 5:	Paired model	= o for LOG;
		= f for LDA;
		= q for QDA;
		= b for BPNN;
		= m for MNN;
		= r for RBFNN.

Simulation 1 Iterative Models

The LOG model showed the highest level of classificatory performance under the simulation 1 baseline proportional condition (see table 17). Although the LOG model is not significantly better than the LDA and QDA models, it is significantly better than all three of the neural network models. Table 23 shows the results of iterating the LOG

Table 23. Iterative procedure results for LOG model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.019127243	29	0.000659560	4.118972431	0.000000002
Models	0.010889629	6	0.001814938	11.334342382	0.000000000
Error	0.027862157	174	0.000160127		
Total	0.057879029	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.011574703	0.01

Means joined by a line are not significantly different.

lu2or	lu1or	lu1ob	lu1om	lu2ob	LOG	lu2om
0.8619	0.8625	0.8637	0.8647	0.8652	0.8659	0.8842

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.009795756	0.05

Means joined by a line are not significantly different.

lu2or	lu1or	lu1ob	lu1om	lu2ob	LOG	lu2om
0.8619	0.8625	0.8637	0.8647	0.8652	0.8659	0.8842

model with each of the neural network models. None of the iterations lead to performance levels that are significantly lower than the LOG baseline performance. On the other hand, LOG model performance does drop off for the first iteration with each of the three models. The drop in performance gets slightly worse for the second iteration with the RBFNN. When iterated a second time with the BPNN and MNN models, the performance of the LOG model improves. In fact, the second pairing with the MNN leads to a performance level that is significantly better than the baseline LOG performance as well as being significantly better than the first iteration performance result. The total increase over the baseline LOG result is 1.83%.

The classificatory performance of the baseline LDA model does not improve statistically (see table 24) by performing two iterations with any of the three neural network models. When paired with the RBFNN, the performance of the LDA model drops off for both iterations, with the second iteration performing significantly worse than the baseline model. The first and second iteration performance results for the pairings with the MNN and BPNN models, leave the baseline performance essentially unchanged.

The results presented in table 25, show that the baseline QDA classificatory performance is not improved by providing it with the output of any of the three neural network models. Not only does its performance not improve, but with every first iteration pairing its performance declines. This trend continues with each second iteration pairing. So, the overall significance of the models factor is due to a significant

Table 24. Iterative procedure results for LDA model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.013956335	29	0.000481253	20.019890696	0.000000000
Models	0.001966617	6	0.000327769	13.635051865	0.000000000
Error	0.004182740	174	0.000024039		
Total	0.020105692	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.004484695	0.01

Means joined by a line are not significantly different.

lu2fr	lu1fr	lu2fb	LDA	lu1fm	lu2fm	lu1fb
0.8573	0.8618	0.8657	0.8658	0.8658	0.8658	0.8658

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.003795431	0.05

Means joined by a line are not significantly different.

lu2fr	lu1fr	lu2fb	LDA	lu1fm	lu2fm	lu1fb
0.8573	0.8618	0.8657	0.8658	0.8658	0.8658	0.8658

decrease in the performance of the QDA model as opposed to an increase. The decrease is most acute for the pairings with the RBFNN models that led to a drop in performance of 1.77% from the baseline QDA classification rate. Although each of the second iteration pairings lead to lower classificatory performance than the corresponding first iteration, in no case was the second iteration performance significantly lower than that obtained with the first iteration.

Table 25. Iterative procedure results for QDA model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.023825911	29	0.000821583	8.437984192	0.000000000
Models	0.005271174	6	0.000878529	9.022840529	0.000000014
Error	0.016941898	174	0.000097367		
Total	0.046038982	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.009025754	0.01

Means joined by a line are not significantly different.

1u2qr	1u1qr	1u2qb	1u2qm	1u1qm	1u1qb	QDA
0.8479	0.8539	0.8549	0.8565	0.8584	0.8587	0.8656

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.007638562	0.05

Means joined by a line are not significantly different.

1u2qr	1u1qr	1u2qb	1u2qm	1u1qm	1u1qb	QDA
0.8479	0.8539	0.8549	0.8565	0.8584	0.8587	0.8656

Table 26. Iterative procedure results for BPNN model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.020419535	29	0.000704122	14.613697914	0.000000000
Models	0.000333562	6	0.000055594	1.153818454	0.333396605
Error	0.008383724	174	0.000048182		
Total	0.029136820	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.006349228	0.01

Means joined by a line are not significantly different.

lu2bq	lu1bq	lu1bf	BPNN	lu2bo	lu1bo	lu2bf
0.8572	0.8587	0.8592	0.8597	0.8600	0.8608	0.8612

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.005373399	0.05

Means joined by a line are not significantly different.

lu2bq	lu1bq	lu1bf	BPNN	lu2bo	lu1bo	lu2bf
0.8572	0.8587	0.8592	0.8597	0.8600	0.8608	0.8612

The results shown in table 26 provide the first occasion for which the overall results for models is not significant. The BPNN baseline performance falls in the middle of the results from the iteration models, but none of the performance results differ from the baseline statistically. The gap between the best and worst performing models was only 0.4%. The highest performance is found on the second pairing with the output from the LDA model.

Table 27. Iterative procedure results for MNN model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.016628406	29	0.000573393	17.225353305	0.000000000
Models	0.000748574	6	0.000124762	3.747994075	0.001561674
Error	0.005792069	174	0.000033288		
Total	0.023169049	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.005277392	0.01

Means joined by a line are not significantly different.

MNN	1u2mq	1u2mo	1u1mo	1u1mq	1u1mf	1u2mf
0.8585	0.8591	0.8607	0.8616	0.8622	0.8623	0.8645

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.004466296	0.05

Means joined by a line are not significantly different.

MNN	1u2mq	1u2mo	1u1mo	1u1mq	1u1mf	1u2mf
0.8585	0.8591	0.8607	0.8616	0.8622	0.8623	0.8645

As with the BPNN model above, the best classificatory performance obtained from a MNN model comes as a result of the second pairing with the output from the LDA model (see table 27). However, in this case, the second iteration LDA/MNN model coupled performance is significantly better than the MNN model baseline performance at both alpha = 0.05 and alpha = 0.01. The increase, although significant, is only 0.6%. In addition, all of the other iteration pairings lead to higher performance levels, though none of them are significantly better than the MNN model baseline classification rate.

The overall results, presented in table 28, for models is once again significant for the RBFNN model. None of the iteration model performance results, however, are significantly different than the RBFNN model baseline performance. As is the case for all of the other simulation 1 iteration model results, the difference between the best and worst performing models is quite small (0.42%) for the RBFNN models.

In summary, of the three parametric models, only the best baseline performing LOG model improves significantly as a result of the iteration process, and this by only 1.83%. This improvement, however, obtained as a result of the second iteration pairing with MNN output, yields the overall best performing simulation 1 model. There is also significant improvement in both the RBFNN and MNN models resulting from pairings with LDA output. The MNN model improves to within 1.97% of the best performing LOG model while the RBFNN improves to within 2.57% of the same LOG model. This 2.57% gap is the largest between any pair of the best of each of the six model types.

Table 28. Iterative procedure results for RBFNN model, simulation 1

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.029971263	29	0.001033492	8.924654388	0.000000000
Models	0.003680400	6	0.000613400	5.296977526	0.000048488
Error	0.020149529	174	0.000115802		
Total	0.053801192	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.009843168	0.01

Means joined by a line are not significantly different.

lu2rq	RBFNN	lu2rf	lu1rq	lu2ro	lu1ro	lu1rf
0.8453	0.8500	0.8523	0.8539	0.8551	0.8572	0.8585

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.008330346	0.05

Means joined by a line are not significantly different.

lu2rq	RBFNN	lu2rf	lu1rq	lu2ro	lu1ro	lu1rf
0.8453	0.8500	0.8523	0.8539	0.8551	0.8572	0.8585

Simulation 2 Iterative Models

The LOG model is the poorest absolute performer and is statistically poorer than all but the LDA model under the simulation 2 baseline proportional condition (see table 18). Iterating with the output of each of the three neural network models leads to a

Table 29. Iterative procedure results for LOG model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.052846024	29	0.001822277	3.674187184	0.000000048
Models	0.199320048	6	0.033220008	66.980238812	0.000000000
Error	0.086298310	174	0.000495967		
Total	0.338464381	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.020370591	0.01

Means joined by a line are not significantly different.

LOG	2u2ob	2u1ob	2u1om	2u2or	2u1or	2u2om
0.6447	0.7121	0.7130	0.7202	0.7336	0.7388	0.7425

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.017239781	0.05

Means joined by a line are not significantly different.

LOG	2u2ob	2u1ob	2u1om	2u2or	2u1or	2u2om
0.6447	0.7121	0.7130	0.7202	0.7336	0.7388	0.7425

significant increase in the classificatory performance of the LOG model at both alpha = 0.05 and alpha = 0.01 (see table 29). The increase obtained by pairing the LOG model with the MNN model (9.78%) results in a performance level that is only 1.9% poorer than the best baseline model (QDA). The baseline performance of the LOG model is 11.68% poorer than the baseline QDA model. The baseline QDA model performance, as

Table 30. Iterative procedure results for LDA model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.028506739	29	0.000982991	2.817702007	0.000016330
Models	0.255355040	6	0.042559173	121.994064181	0.000000000
Error	0.060702102	174	0.000348863		
Total	0.344563882	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.017084581	0.01

Means joined by a line are not significantly different.

LDA	2u1fb	2u2fb	2u2fr	2u1fr	2u2fm	2u1fm
0.6462	0.7280	0.7339	0.7398	0.7408	0.7551	0.7553

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.014458807	0.05

Means joined by a line are not significantly different.

LDA	2u1fb	2u2fb	2u2fr	2u1fr	2u2fm	2u1fm
0.6462	0.7280	0.7339	0.7398	0.7408	0.7551	0.7553

will be discussed below (see table 31). does not improve by iterating with any of the neural network models This model remains the best performing model when compared to all of the simulation 2 proportional models. Another point of interest is that the second iteration results from the pairing with the MNN is significantly better than the first iteration results at both alpha = 0.05 and alpha = 0.01.

Similar to the LOG model above, the baseline LDA model performance improves significantly by pairing with the output from any of the three neural network models (see table 30). The 10.91% increase in absolute classificatory performance, obtained on the first iteration pairing with the MNN model output, is even larger than that realized with the LOG model. The best LDA model is only 0.62% below the classification rate of the best performing QDA baseline model. None of the first and second iteration models for a given neural network differ from each other statistically.

The baseline QDA model, with its 0.7615 classification rate, is the overall best performing model. All iterative pairings with neural network models lead to a significant decrease in QDA model classificatory performance at both alpha = 0.05 and alpha = 0.01 (see table 31). None of the iterative procedure QDA results, for simulation 2, differ statistically. The worst performing QDA model, the second iteration pairing with the output from the corresponding RBFNN model, is only 1.3% below the baseline model performance level.

Although the QDA model baseline classificatory performance does not improve with the iterative procedure, its output significantly improves the performance of the worst performing baseline neural network model, the BPNN (see table 32). This performance improvement is true for both the first and second iterations at both alpha = 0.05 and alpha = 0.01. The pairings with the LOG and LDA models do not fare as well. There is a significant drop with the first iteration LOG pairing followed by a significant increase in performance from the first to the second iteration (though not statistically different from the BPNN baseline performance). The first and second iteration pairings

Table 31. Iterative procedure results for QDA model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.017888835	29	0.000616856	9.838303241	0.000000000
Models	0.002928079	6	0.000488013	7.783369201	0.000000198
Error	0.010909707	174	0.000062699		
Total	0.031726620	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.007242842	0.01

Means joined by a line are not significantly different.

2u2qr	2u1qr	2u1qm	2u2qm	2u2qb	2u1qb	QDA
0.7485	0.7515	0.7519	0.7523	0.7539	0.7540	0.7615

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.006129671	0.05

Means joined by a line are not significantly different.

2u2qr	2u1qr	2u1qm	2u2qm	2u2qb	2u1qb	QDA
0.7485	0.7515	0.7519	0.7523	0.7539	0.7540	0.7615

with LDA output sandwich the baseline result and do not differ statistically from each other or the BPNN baseline performance. The best performing BPNN model is only 1.21% poorer than the baseline QDA model.

Table 32. Iterative procedure results for BPNN model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.038390049	29	0.001323795	2.614620525	0.000064020
Models	0.099246045	6	0.016541008	32.670062048	0.000000000
Error	0.088097026	174	0.000506305		
Total	0.225733120	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.020581788	0.01

Means joined by a line are not significantly different.

2u1bo	2u2bo	2u1bf	BPNN	2u2bf	2u2bq	2u1bq
0.6798	0.7127	0.7163	0.7215	0.7278	0.7469	0.7494

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.017418519	0.05

Means joined by a line are not significantly different.

2u1bo	2u2bo	2u1bf	BPNN	2u2bf	2u2bq	2u1bq
0.6798	0.7127	0.7163	0.7215	0.7278	0.7469	0.7494

The MNN model, the best absolute performing neural network under the simulation 2 configuration, does not improve statistically by iterating with any of the parametric models (see table 33). The best performance is obtained using the output from the QDA models, with the first iteration result being slightly better than the second iteration result. The closest the MNN model comes to the baseline QDA classificatory

Table 33. Iterative procedure results for MNN model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.027668638	29	0.000954091	4.746113478	0.000000000
Models	0.067617950	6	0.011269658	56.060773084	0.000000000
Error	0.034978479	174	0.000201026		
Total	0.130265067	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.012968894	0.01

Means joined by a line are not significantly different.

2u1mo	2u2mo	MNN	2u1mf	2u2mf	2u2mq	2u1mq
0.6965	0.7329	0.7407	0.7467	0.7495	0.7499	0.7503

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.010975671	0.05

Means joined by a line are not significantly different.

2u1mo	2u2mo	MNN	2u1mf	2u2mf	2u2mq	2u1mq
0.6965	0.7329	0.7407	0.7467	0.7495	0.7499	0.7503

performance is within 1.12% at a 0.7503 classification rate. The only model that is significantly worse than the baseline BPNN model is the first pairing with the LOG output. This LOG output, which comes from the baseline LOG model, represents a model performing at only a 0.6447 classification rate, which is the poorest of all six of the baseline models. The second iteration MNN performance with LOG output yields a result that, while not quite as good as the baseline result, is no longer statistically different from the MNN baseline performance.

Table 34. Iterative procedure results for RBFNN model, simulation 2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.042383352	29	0.001461495	10.921395356	0.000000000
Models	0.004260990	6	0.000710165	5.306890601	0.000047423
Error	0.023284581	174	0.000133819		
Total	0.069928924	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.010581243	0.01

Means joined by a line are not significantly different.

2u2ro	RBFNN	2u1ro	2u1rf	2u2rf	2u2rq	2u1rq
0.7337	0.7369	0.7376	0.7396	0.7400	0.7421	0.7491

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.008954984	0.05

Means joined by a line are not significantly different.

2u2ro	RBFNN	2u1ro	2u1rf	2u2rf	2u2rq	2u1rq
0.7337	0.7369	0.7376	0.7396	0.7400	0.7421	0.7491

Table 34 presents the iterative results for the RBFNN models. Coupled with QDA model baseline output, the RBFNN performance improves significantly at alpha = 0.05 and alpha = 0.01. This brings the RBFNN model performance to within 1.24% of the baseline QDA model. No other iterative RBFNN models differ statistically from the baseline model and no second iteration models differ statistically from their first iteration counterparts.

In summary, although the iterative process leads to significantly improved performance for the LOG, LDA, BPNN, and RBFNN models and absolute improved performance for the MNN model, it is not effective in improving any model to a performance level higher than that of the baseline QDA model. It brings the next best performing model, the first iteration LDA model paired with MNN output, to within 0.62% of the baseline QDA model. The gap in performance of 11.68% that had occurrs between the best and worst baseline models was reduced to only 1.9% between the baseline QDA model and the second iteration LOG model paired with MNN output after all of the iterations are complete.

Simulation 3 Iterative Models

The absolute baseline performance of the LOG model under simulation 3 configurations, is second poorest of the six classification models and is not statistically different from the LDA model, the worst performer (see table19). Its performance is improved significantly (at both alpha = 0.05 and alpha = 0.01) with every iterative pairing, showing its best performance when paired a second time with MNN model output (see table 35). The MNN is the best performing simulation 3 baseline model. The best performing LOG model has a classification rate, 0.7617, just slightly better than the baseline MNN model (0.7610). While not statistically different, the second iteration LOG performance using MNN model output is higher than the corresponding first iteration performance.

Table 35. Iterative procedure results for LOG model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.055809529	29	0.001924467	3.519362699	0.000000137
Models	0.319741967	6	0.053290328	97.454536915	0.000000000
Error	0.095147105	174	0.000546822		
Total	0.470698600	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.021389482	0.01

Means joined by a line are not significantly different.

LOG	3u1ob	3u2ob	3u2or	3u1or	3u1om	3u2om
0.6489	0.6990	0.7150	0.7521	0.7553	0.7585	0.7617

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.018102076	0.05

Means joined by a line are not significantly different.

LOG	3u1ob	3u2ob	3u2or	3u1or	3u1om	3u2om
0.6489	0.6990	0.7150	0.7521	0.7553	0.7585	0.7617

As was true for the baseline LOG model discussed above, all iterative pairings with the LDA model result in a significant increase in classification rate at both alpha = 0.05 and alpha = 0.01 (see table 36). From being the poorest absolute baseline performing model, the first iteration LDA model paired with MNN model output yields a

Table 36. Iterative procedure results for LDA model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.036841763	29	0.001270406	3.238577139	0.000000933
Models	0.325779681	6	0.054296613	138.415454001	0.000000000
Error	0.068255462	174	0.000392273		
Total	0.430876906	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.01811637	0.01

Means joined by a line are not significantly different.

LDA	3u2fb	3u1fb	3u2fr	3u1fr	3u2fm	3u1fm
0.6469	0.7429	0.7458	0.7515	0.7563	0.7679	0.7713

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.015332018	0.05

Means joined by a line are not significantly different.

LDA	3u2fb	3u1fb	3u2fr	3u1fr	3u2fm	3u1fm
0.6469	0.7429	0.7458	0.7515	0.7563	0.7679	0.7713

performance level that is 12.44% above its baseline performance, 0.98% better than the best performing MNN model baseline performance, and the best overall performance of any simulation 3 configuration model. In no case under this data configuration, does the second iteration results differ statistically from the first iteration results.

Table 37. Iterative procedure results for QDA model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.039498311	29	0.001362011	9.537903865	0.000000000
Models	0.023007907	6	0.003834651	26.853338251	0.000000000
Error	0.024847164	174	0.000142800		
Total	0.087353382	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.010930522	0.01

Means joined by a line are not significantly different.

QDA	3u2qr	3u1qr	3u1qb	3u2qb	3u1qm	3u2qm
0.7239	0.7476	0.7498	0.7520	0.7531	0.7553	0.7574

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.009250581	0.05

Means joined by a line are not significantly different.

QDA	3u2qr	3u1qr	3u1qb	3u2qb	3u1qm	3u2qm
0.7239	0.7476	0.7498	0.7520	0.7531	0.7553	0.7574

The QDA model generally follows the pattern exhibited by the LOG and LDA models under simulation 3. That is, the baseline QDA classification rate is statistically the poorest model at both alpha = 0.05 and alpha = 0.01 (see table 20). Another similarity to the LOG and LDA models results is that the best absolute performance of any QDA model comes as a result of pairing with MNN model output (see table 37). In fact, for each of the three parametric models, the two best absolute performances come as a result of pairings with MNN model output. The maximum increase in the performance (3.35%) over the baseline QDA model is less than that for the LOG and LDA models, but is already significantly better than either of those two models. This improvement leaves the best QDA model performance just 0.36% below the best baseline performing MNN model.

The results shown in table 38 indicate that the iterative procedure is not effective in improving the performance of the BPNN baseline model. In fact, the only statistical differences are seen with pairings resulting in lower classification rates. For example, the first pairing with LOG output leads to an 8.59% drop in performance, which is the largest absolute decrease found thus far due to an iterative pairing. The second pairing with LOG output, however, results in a significant increase in performance of 3.35%, which, while still statistically poorer than the baseline BPNN model (at both alpha = 0.05 and alpha = 0.01), is moving in the desired direction. Although the movements are not significant, the change in performance from the first to the second pairings with LDA and QDA output also show movement in the desired direction.

The results for the MNN models, presented in table 39, are similar to those seen for the BPNN models (see table 38). There are no models that perform significantly better than the baseline MNN model, but the first pairing with LOG output leads to a significant decrease in classificatory performance at both alpha = 0.05 and alpha = 0.01. The second pairing with LOG output yields a significant increase in the MNN model's performance. In fact, the absolute performance of the iteration 2, LOG paired, MNN model is slightly better than the baseline performance. Once again, the iteration process

Table 38. Iterative procedure results for BPNN model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.038854220	29	0.001339801	1.940986961	0.004902774
Models	0.187750412	6	0.031291735	45.332750115	0.000000000
Error	0.120106588	174	0.000690268		
Total	0.346711220	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.024031777	0.01

Means joined by a line are not significantly different.

3u1bo	3u2bo	3u1bf	3u2bf	3u1bq	BPNN	3u2bq
0.6680	0.7035	0.7247	0.7384	0.7525	0.7539	0.7539

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.020338271	0.05

Means joined by a line are not significantly different.

3u1bo	3u2bo	3u1bf	3u2bf	3u1bq	BPNN	3u2bq
0.6680	0.7035	0.7247	0.7384	0.7525	0.7539	0.7539

yields movement in the desired direction. The first pairing with LDA output yields the best absolute performance of any MNN model under the simulation 3 configuration but leaves the best MNN model 0.81% below the best performing LDA model.

Table 39. Iterative procedure results for MNN model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.030358338	29	0.001046839	2.732673505	0.000028986
Models	0.044333012	6	0.007388835	19.287846351	0.000000000
Error	0.066656345	174	0.000383082		
Total	0.141347695	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.017902893	0.01

Means joined by a line are not significantly different.

3u1mo	3u1mq	3u2mq	MNN	3u2mo	3u2mf	3u1mf
0.7193	0.7564	0.7575	0.7610	0.7618	0.7620	0.7632

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.015151351	0.05

Means joined by a line are not significantly different.

3u1mo	3u1mq	3u2mq	MNN	3u2mo	3u2mf	3u1mf
0.7193	0.7564	0.7575	0.7610	0.7618	0.7620	0.7632

Although the overall RBFNN model's results are significant for replications, they are not significant for models (see table 40). From an absolute performance perspective, however, the baseline RBFNN model performs poorest while the first pairing with LDA output yields the best classificatory performance. The performance of this first iteration RBFNN model is only 1.04% below that of the MNN baseline model. Further, all second iteration models result in lower absolute performance levels than do their first iteration counterparts.

Table 40. Iterative procedure results for RBFNN model, simulation 3

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.041438624	29	0.001428918	11.026143677	0.000000000
Models	0.000562564	6	0.000093761	0.723497823	0.631209788
Error	0.022549293	174	0.000129594		
Total	0.064550481	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.010412834	0.01

Means joined by a line are not significantly different.

RBFNN	3u2ro	3u2rq	3u1rq	3u1ro	3u2rf	3u1rf
0.7464	0.7472	0.7475	0.7488	0.7505	0.7505	0.7506

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.008812458	0.05

Means joined by a line are not significantly different.

RBFNN	3u2ro	3u2rq	3u1rq	3u1ro	3u2rf	3u1rf
0.7464	0.7472	0.7475	0.7488	0.7505	0.7505	0.7506

In summary, all three parametric models improve significantly, at both alpha = 0.05 and alpha = 0.01, over their baseline models. In fact, all iterative pairings with each of the three neural network models yield significantly better performance results. This is especially true of the worst performing LOG and LDA models. The iteration process results in an LDA model that yields the best absolute classification rate (0.7713) and outperforming all neural network models. Unlike the parametric cases, there is no significant improvement in any of the neural network models as a result of the iterative process. There is some absolute improvement from the baseline MNN and RBFNN models. The best MNN model's performance was just 0.81% below the best performing LDA model. The largest difference between the best performing of each of the six models is only 2.07% between best LDA and best RBFNN models.

Simulation 4 Iterative Models

Simulation 4 presents an essentially insurmountable problem for all three parametric models. The alternating group structure of the data leads the LOG, LDA, and QDA baseline models to place practically all of the validation exemplars into only the higher proportion group. The fact that each of their baseline performances is slightly below the expected 70% correct is because of a small number of outliers in some of the replication samples. Thus, under the simulation 4 configuration, the three parametric models, without neural network output to help them, cease being discriminant models entirely. This is not true of the BPNN, MNN, or RBFNN models. Each of these models and their iterative counterparts exhibit true discriminant behavior under the simulation 4 configuration.

Table 41. Iterative procedure results for LOG model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.014170186	29	0.000488627	2.419302950	0.000234004
Models	0.008578757	6	0.001429793	7.079226926	0.000000925
Error	0.035142814	174	0.000201970		
Total	0.057891757	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.012999323	0.01

Means joined by a line are not significantly different.

LOG	4u2ob	4u1ob	4u2or	4u1or	4u1om	4u2om
0.6996	0.7063	0.7073	0.7099	0.7127	0.7151	0.7212

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.011001423	0.05

Means joined by a line are not significantly different.

LOG	4u2ob	4u1ob	4u2or	4u1or	4u1om	4u2om
0.6996	0.7063	0.7073	0.7099	0.7127	0.7151	0.7212

The baseline classificatory performance of the simulation 4 configuration LOG model is significantly improved, at both alpha = 0.05 and alpha = 0.01, by pairing it with first iteration RBFNN output and first and second iteration MNN outputs (see table 41). The absolute increase in classification rate is 2.16%. In addition to the improvement in overall classificatory performance, the actual validation file classification results for one replication indicate that the paired LOG models truly discriminate between the two groups. The classification rates for the two groups' change from baseline rates of 0% correct for group 1 and 100% correct for group 2 to 49% correct for group 1 and 84% correct for group 2 for the LOG model paired with first iteration MNN output. A perusal of several other validation results files revealed similar changes in the nature of LOG model performance.

The performance results for LDA, shown in table 42, are similar to those found for the LOG models. The baseline LOG performance is just below the 70% expected by placing all observations into the dominant class. The statistically best performing models are those paired with first and second iteration MNN model outputs. As is the case with the iterative LOG models, the nature of the LDA model output changed as a result of being paired with neural network model output. For example, one replication result shows a change from 0% to 40% correct for group 1 while the classification rate for group 2 dropped from 100% to 87%. The maximum increase in absolute performance over the baseline result is 1.53%.

The classification rate of the baseline QDA model does not change significantly by iterating with the any of the three neural networks' outputs (see table 43). The nature of the QDA results, however, do change in the same manner as for the LOG and LDA models. Thus, while the absolute QDA performance improves by only 0.87%, the performance within each group changes dramatically. One example shows a 30% improvement in group 1 performance and a 12% decrease in group 2 performance. More

Table 42. Iterative procedure results for LDA model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.016619338	29	0.000573081	2.385319130	0.000292505
Models	0.006701224	6	0.001116871	4.648722675	0.000207624
Error	0.041804062	174	0.000240253		
Total	0.065124624	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.014177894	0.01

Means joined by a line are not significantly different.

LDA	4u2fb	4u1fb	4u2fr	4u1fr	4u2fm	4u1fm
0.6997	0.7011	0.7045	0.7101	0.7110	0.7136	0.7150

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.011998856	0.05

Means joined by a line are not significantly different.

LDA	4u2fb	4u1fb	4u2fr	4u1fr	4u2fm	4u1fm
0.6997	0.7011	0.7045	0.7101	0.7110	0.7136	0.7150

importantly, the QDA model is able to discriminate between the two groups when using output from a neural network while being totally unable to do so without the neural network output.

Table 43. Iterative procedure results for QDA model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.045332763	29	0.001563199	4.137985252	0.000000002
Models	0.012604212	6	0.002100702	5.560824530	0.000026859
Error	0.065731645	174	0.000377768		
Total	0.123668620	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.017778279	0.01

Means joined by a line are not significantly different.

4u1qb	4u2qb	QDA	4u2qm	4u2qr	4u1qr	4u1qm
0.6865	0.6881	0.6977	0.7004	0.7056	0.7057	0.7064

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.015045889	0.05

Means joined by a line are not significantly different.

4u1qb	4u2qb	QDA	4u2qm	4u2qr	4u1qr	4u1qm
0.6865	0.6881	0.6977	0.7004	0.7056	0.7057	0.7064

The simulation 4 baseline performance of the BPNN model is statistically better than that of any of the corresponding iterative models (see table 44). There are no significant differences between any pair of iterative models, but there is an absolute drop from the baseline performance of as much as 3.79%. Of all the iterative pairings, only those with the LOG output show movement in the desired direction when going from the first to the second iteration.

Table 44. Iterative procedure results for BPNN model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.0682665052	29	0.002353967	4.291953980	0.000000001
Models	0.026960512	6	0.004493419	8.192784101	0.000000082
Error	0.095432131	174	0.000548461		
Total	0.190657695	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.021421495	0.01

Means joined by a line are not significantly different.

4u1bo	4u2bf	4u1bq	4u2bq	4u1bf	4u2bo	BPNN
0.6815	0.6864	0.6885	0.6911	0.6930	0.6939	0.7194

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.01812917	0.05

Means joined by a line are not significantly different.

4u1bo	4u2bf	4u1bq	4u2bq	4u1bf	4u2bo	BPNN
0.6815	0.6864	0.6885	0.6911	0.6930	0.6939	0.7194

While the overall results for replications are significant for both the MNN and RBFNN model experiments, the results for models are not (see tables 45 and 46). While the first iteration with LOG output improves the absolute performance of the MNN model, the total improvement is only 0.88%. For the RBFNN model, all of the iteration classification rates are less than the baseline rate. Each of the second iteration results show movement do show movement in the desired direction, however.

Table 45. Iterative procedure results for MNN model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.051552077	29	0.001777658	5.976683152	0.000000000
Models	0.002257731	6	0.000376288	1.265123715	0.275862918
Error	0.051753198	174	0.000297432		
Total	0.105563006	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.015775062	0.01

Means joined by a line are not significantly different.

4u1mo	MNN	4u1mf	4u2mf	4u2mq	4u1mq	4u2mo
0.7052	0.7075	0.7088	0.7089	0.7096	0.7120	0.7163

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.013350551	0.05

Means joined by a line are not significantly different.

4u1mo	MNN	4u1mf	4u2mf	4u2mq	4u1mq	4u2mo
0.7052	0.7075	0.7088	0.7089	0.7096	0.7120	0.7163

Table 46. Iterative procedure results for RBFNN model, simulation 4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.041762167	29	0.001440075	6.731454269	0.000000000
Models	0.000735729	6	0.000122621	0.573178969	0.751343698
Error	0.037224200	174	0.000213932		
Total	0.079722095	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.013378738	0.01

Means joined by a line are not significantly different.

4u1rq	4u1rf	4u2rq	4u2rf	4u1ro	4u2ro	RBFNN
0.7049	0.7062	0.7071	0.7077	0.7084	0.7087	0.7113

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.011322525	0.05

Means joined by a line are not significantly different.

4u1rq	4u1rf	4u2rq	4u2rf	4u1ro	4u2ro	RBFNN
0.7049	0.7062	0.7071	0.7077	0.7084	0.7087	0.7113

In summary, while the three neural network models demonstrate superior absolute baseline performance when compared to the three parametric models (two were statistically superior), none of them improve significantly via the iterative process. Neither is the nature of their output changed as a result of iterative pairings with parametric model output. They are each able, within their limits, to place observations into one of the two desired groups by discriminating each groups relevant spatial characteristics under both baseline and iterative conditions. This ability to discriminate is not true of the three parametric models. The parametric models' baseline performance consists of placing essentially all observations into the higher proportion group without recourse to the spatial characteristics of the observations. Through pairing the parametric models with output from any of the three neural networks, they become able to identify spatial regions belonging to each of the two groups and assign observations accordingly. In addition, both the LOG and LDA models improve significantly by iterating with the neural network models and the second iteration LOG model paired with MNN output became the best performing model under the simulation 4 configuration.

Simulation 5 Iterative Models

As is the case for the simulation 4 configuration, the baseline performance of the LOG model represents the placement of all observations into the higher proportion group. The fact that the average baseline classificatory performance is slightly better than the 70% expected (see table 21) is likely due to extreme outliers from the smaller group being classified correctly. The addition of output from one of the neural network models as input to the LOG model dramatically changes its ability to discriminate between the groups (see table 47). One typical example, a first iteration pairing with MNN output, yieldes a group 1 classification rate of 70% and a group 2 classification rate of 90.6%. All iterative pairings lead to significant improvements in LOG model performance at both alpha = 0.05 and alpha = 0.01. There is a significant improvement, again at both alpha levels, between the first and second iteration pairings with BPNN output. These models, however, perform significantly poorer than the models paired with RBFNN and

MNN outputs. The best absolute performance, like for many of the previous configurations, is with MNN output. This second iteration MNN paired LOG model performs at a level 13.4% above the baseline LOG model.

Table 47. Iterative procedure results for LOG model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.035739077	29	0.001232382	2.858024989	0.000012431
Models	0.501924317	6	0.083654053	194.002652914	0.000000000
Error	0.075028898	174	0.000431201		
Total	0.612692292	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.018994014	0.01

Means joined by a line are not significantly different.

LOG	5u1ob	5u2ob	5u2or	5u1or	5u1om	5u2om
0.7008	0.7374	0.7927	0.8211	0.8245	0.8333	0.8348

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.016074774	0.05

Means joined by a line are not significantly different.

LOG	5u1ob	5u2ob	5u2or	5u1or	5u1om	5u2om
0.7008	0.7374	0.7927	0.8211	0.8245	0.8333	0.8348

Similar to almost every data configuration, the LDA model under the simulation 5 configuration closely mimics the behavior of the LOG model (see table 48). The LDA model baseline performance results from placing all observations into the higher proportion group. A LDA model's ability to classify observations into the correct group as opposed to the larger group occurs, once again, only when paired with the output from

Table 48. Iterative procedure results for LDA model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.046592596	29	0.001606641	2.508090674	0.000130164
Models	0.400528774	6	0.066754796	104.209374536	0.000000000
Error	0.111461512	174	0.000640583		
Total	0.558582882	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.023150743	0.01

Means joined by a line are not significantly different.

LDA	5u1fb	5u2fb	5u2fr	5u2fm	5u1fr	5u1fm
0.7008	0.7673	0.7677	0.8177	0.8221	0.8234	0.8283

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.019592645	0.05

Means joined by a line are not significantly different.

LDA	5u1fb	5u2fb	5u2fr	5u2fm	5u1fr	5u1fm
0.7008	0.7673	0.7677	0.8177	0.8221	0.8234	0.8283

one of the three neural network models. Shifts in classification rate from 0% to 71.8% and 100% to 89.2% for groups 1 and 2 respectively, are typical of iterative results. An overall increase of 12.75% is obtained by pairing the LDA model with the output from the first iteration MNN model. The performance increases over the baseline LDA model engendered by all of the possible pairings are significant at alpha = 0.05 and alpha = 0.01.

Table 49. Iterative procedure results for QDA model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.031499900	29	0.001086203	6.571972015	0.000000000
Models	0.093082600	6	0.015513767	93.864589128	0.000000000
Error	0.028758400	174	0.000165278		
Total	0.153340900	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.011759391	0.01

Means joined by a line are not significantly different.

QDA	5u1qb	5u2qb	5u2qr	5u1qr	5u1qm	5u2qm
0.7628	0.8027	0.8063	0.8183	0.8196	0.8269	0.8291

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.00995206	0.05

Means joined by a line are not significantly different.

QDA	5u1qb	5u2qb	5u2qr	5u1qr	5u1qm	5u2qm
0.7628	0.8027	0.8063	0.8183	0.8196	0.8269	0.8291

Table 50. Iterative procedure results for BPNN model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.121577043	29	0.004192312	4.113549784	0.000000002
Models	0.274307283	6	0.045717881	44.858966995	0.000000000
Error	0.177331574	174	0.001019147		
Total	0.573215900	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.029200851	0.01

Means joined by a line are not significantly different.

5u1bo	5u1bf	5u2bf	5u2bo	5u2bq	5u1bq	BPNN
0.6982	0.7487	0.7673	0.7711	0.7963	0.7964	0.8163

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.024712896	0.05

Means joined by a line are not significantly different.

5u1bo	5u1bf	5u2bf	5u2bo	5u2bq	5u1bq	BPNN
0.6982	0.7487	0.7673	0.7711	0.7963	0.7964	0.8163

The QDA model under simulation 5 configurations does not follow the pattern of the LOG and LDA models. Its baseline performance represents true discrimination between the two groups. Its baseline performance, which is statistically better than LOG and LDA and statistically worse than all three neural network models, is significantly improved upon (at both alpha = 0.05 and alpha = 0.01) by all of the iterative pairings (see table 49). The best absolute performance results derive once again from a pairing with MNN second iteration output. The total increase in classification rate of 6.63% brings the best performing QDA model in line with the best performances of the other five models.

The BPNN baseline performance is not outperformed by any of the iteratively paired models (see table 50). In fact, four of the six pairings result in significant decreases in classificatory performance. The most dramatic of these decreases came as a result of the first iteration pairing with LOG output. The resulting 11.81% decrease in performance is partially erased with the second LOG output pairing which leads to a significant 7.29% increase over the first iteration results. Although this still leaves the classification rate significantly lower than the baseline rate, its movement is in the desired direction.

No significant improvement is obtained by iterating the MNN model with the output from any of the three parametric models (see table 51). In fact, no significant differences are found between any of the iterative models and the baseline MNN model One interesting point is that the first iteration pairings with each of the parametric models leads to absolute decreases in MNN performance. The second iteration models each showed a reverse of this trend, i.e., showed improved performance, and two of the three second iteration models performed at slightly higher than baseline levels.

Table 51. Iterative procedure results for MNN model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.025042435	29	0.000863532	6.352000560	0.000000000
Models	0.003718095	6	0.000619683	4.558282511	0.000254361
Error	0.023654690	174	0.000135946		
Total	0.052415220	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.010665006	0.01

Means joined by a line are not significantly different.

5u1mf	5u2mf	5u1mo	5u1mq	MNN	5u2mq	5u2mo
0.8190	0.8228	0.8229	0.8248	0.8265	0.8310	0.8316

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.009025873	0.05

Means joined by a line are not significantly different.

5u1mf	5u2mf	5u1mo	5u1mq	MNN	5u2mq	5u2mo
0.8190	0.8228	0.8229	0.8248	0.8265	0.8310	0.8316

The overall test of the RBFNN models is not significant (see table 52). The initial simulation 5 baseline performance of the RBFNN model is the best absolute performance of any RBFNN model. The results across the baseline and all iterative models are very consistent with the largest gap being only 0.84%.

Table 52. Iterative procedure results for RBFNN model, simulation 5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.032129249	29	0.001107905	7.182116354	0.000000000
Models	0.001160100	6	0.000193350	1.253412551	0.281519686
Error	0.026841043	174	0.000154259		
Total	0.060130392	209			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.011360623	0.01

Means joined by a line are not significantly different.

5u2rf	5u2ro	5u1rq	5u2rq	5u1ro	5u1rf	RBFNN
0.8121	0.8155	0.8169	0.8170	0.8171	0.8179	0.8205

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.00961458	0.05

Means joined by a line are not significantly different.

5u2rf	5u2ro	5u1rq	5u2rq	5u1ro	5u1rf	RBFNN
0.8121	0.8155	0.8169	0.8170	0.8171	0.8179	0.8205

In summary, while the LOG and LDA baseline models are statistically the poorest of the six classification models and the QDA baseline model is statistically poorer than all three neural network models, the iterative process significantly improves all three parametric models. Not only are the parametric models improved, but their best absolute performances placed them first (LOG), third (QDA), and fourth (LDA) among the best results of the six models. The iterative process, on the other hand, fails to improve any of the neural network models significantly over their baseline performances. Only the MNN model show even a small amount of absolute improvement (0.51%) over its baseline performance as a result of iterating with the parametric outputs.

D/FW Mode Choice Iterative Models

Iterating with both BPNN and MNN output resulted in significant improvement of the LOG model at both alpha = 0.05 and alpha = 0.01 (see table 53). The second iteration pairing with MNN output yields a 3.67% increase in classificatory performance over the LOG baseline model. This performance is statistically better than either of the pairings with BPNN output. While not significantly different, the second iteration pairings each yield performance results in the desired direction.

The baseline performance of the LDA model is both absolutely and statistically the worst of the six baseline-classification models. Similar to the LOG baseline model, iteration with both BPNN and MNN outputs yields significantly better classificatory performance from the LDA models at alpha = 0.05 and alpha = 0.01(see table 54). The maximum improvement of 4.46% over the baseline model is the largest for any of the six models. Also in line with the LOG model results, is the absolute improvement from first to second iteration for pairings with both BPNN and MNN outputs, and that the pairings with MNN output result in significantly better performance than pairings with BPNN output.

Table 53. Iterative procedure results for LOG model, D/FW mode choice

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.002605511	19	0.000137132	3.708005993	0.000023163
Models	0.017595953	4	0.004398988	118.947120723	0.000000000
Error	0.002810687	76	0.000036983		
Total	0.023012151	99			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.82	0.006554378	0.01

Means joined by a line are not significantly different.

LOG	du1ob	du2ob	du1om	du2om
0.7036	0.7296	0.7337	0.7383	0.7403

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
3.98	0.005412121	0.05

Means joined by a line are not significantly different.

LOG	du1ob	du2ob	du1om	du2om
0.7036	0.7296	0.7337	0.7383	0.7403

The QDA baseline model is, as with the other two parametric models, improves significantly, at both alpha = 0.05 and alpha = 0.01, by the iterative process (see table 55). In contrast to the LOG and LDA model cases, there is no significant difference between the pairings with BPNN and MNN outputs. However, there is movement in the desired direction from iteration 1 to iteration 2 for both the BPNN and MNN pairings. Also, the best absolute performance comes with the second iteration pairing with MNN output, which yields a 3.76% improvement over the QDA baseline model.

Table 54. Iterative procedure results for LDA model, D/FW mode choice

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.002805647	19	0.000147666	4.643727479	0.000000721
Models	0.026814436	4	0.006703609	210.812316265	0.000000000
Error	0.002416720	76	0.000031799		
Total	0.032036802	99			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.82	0.006077688	0.01

Means joined by a line are not significantly different.

LDA	du1fb	du2fb	du1fm	du2fm
0.6872	0.7217	0.7227	0.7312	0.7318

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
3.98	0.005018506	0.05

Means joined by a line are not significantly different.

LDA	du1fb	du2fb	du1fm	du2fm
0.6872	0.7217	0.7227	0.7312	0.7318

Table 55. Iterative procedure results for QDA model, D/FW mode choice

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.006528660	19	0.000343614	11.378610970	0.000000000
Models	0.021601230	4	0.005400307	178.828727783	0.000000000
Error	0.002295064	76	0.000030198		
Total	0.030424954	99			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.82	0.00592274	0.01

Means joined by a line are not significantly different.

QDA	du1qb	du1qm	du2qb	du2qm
0.7055	0.7412	0.7419	0.7426	0.7431

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
3.98	0.004890561	0.05

Means joined by a line are not significantly different.

QDA	du1qb	du1qm	du2qb	du2qm
0.7055	0.7412	0.7419	0.7426	0.7431

For two of the three first iteration pairings, those with LDA and LOG outputs, there is a significant decrease in the classificatory performance of the BPNN models at both alpha = 0.05 and alpha = 0.01 (see table 56). The opposite is true of the first pairing with QDA output. Not only was there a significant increase in performance, but this first iteration model yielded the best performance of any BPNN model. The first iteration

Table 56. Iterative procedure results for BPNN model, D/FW mode choice

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.002936701	19	0.000154563	2.707434394	0.000587316
Models	0.041141005	6	0.006856834	120.108960807	0.000000000
Error	0.006508083	114	0.000057088		
Total	0.050585790	139			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.008464413	0.01

Means joined by a line are not significantly different.

du1bf	du2bf	du1bo	du2bo	BPNN	du2bq	du1bq
0.7217	0.7217	0.7243	0.7314	0.7346	0.7620	0.7647

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.007163495	0.05

Means joined by a line are not significantly different.

du1bf	du2bf	du1bo	du2bo	BPNN	du2bq	du1bq
0.7217	0.7217	0.7243	0.7314	0.7346	0.7620	0.7647

performance drops for the LDA and LOG pairings are improved upon, albeit non-significantly, by the second iteration pairings. The BPNN model performance resulting from the second iteration pairing with LDA output remains statistically below that of the baseline model. The second iteration results from the LOG pairing are statistically indistinguishable from the baseline performance. The slight absolute performance decrease resulting from the second iteration pairing with QDA output leaves the BPNN model's performance statistically unchanged from the first iteration result.

Table 57. Iterative procedure results for MNN model, D/FW mode choice

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*
Replications	0.002069181	19	0.000108904	3.742709488	0.000005254
Models	0.023669444	6	0.003944907	135.5744849	0.000000000
Error	0.003317139	114	0.000029098		
Total	0.029055764	139			

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
5.01	0.006042994	0.01

Means joined by a line are not significantly different.

du1mf	du1mo	du2mf	du2mo	MNN	du2mq	du1mq
0.7321	0.7330	0.7360	0.7394	0.7479	0.7632	0.7655

TUKEY MULTIPLE COMPARISON TEST

Critical Q	Distance	Alpha
4.24	0.00511423	0.05

Means joined by a line are not significantly different.

du1mf	du1mo	du2mf	du2mo	MNN	du2mq	du1mq
0.7321	0.7330	0.7360	0.7394	0.7479	0.7632	0.7655

In general, the iterative process results for the MNN model parallel those found for the BPNN model discussed above. The baseline MNN model improves significantly, at both alpha = 0.05 and alpha = 0.01, by iterating with QDA output (see table 57). As with the BPNN model, the first iteration with QDA output yields the best performing MNN model. The corresponding second iteration performance results are slightly worse but not significantly so. All of the pairings with LOG and LDA model outputs result in classificatory performance that is statistically poorer than the baseline model. However, similar to the BPNN model case, the absolute change in performance from each of these first iteration pairings to the corresponding second iteration pairings is in the desired direction. In addition, the second pairing with LOG output leads to a significant improvement in performance over the first iteration result.

In summary, all five of the models to which the iterative process is applied improve their classification rates significantly (alpha = 0.05 and alpha = 0.01) over their baseline performances. What began as a 6.07% difference between the best and worst of the six baseline models (MNNdu and LDAdu) is reduced to a 3.37% difference between the best and worst performances of the six best iterative models (du1mq and du2fm). Although only one of the second iteration results is significantly better than the corresponding first iteration result, none of the second iteration pairings lead to statistically poorer model performance. All but two lead to performances in the desired direction. All parametric iterative models demonstrated higher classification rates than their baseline counterparts

Baseline Proportional Condition Versus Equal Condition Analysis

In order to better understand the effect of differences in the relative proportions of exemplars, the results from the equal and baseline proportional conditions are compared using two-sample t-tests in Excel with alpha = 0.05. The results for each of the six models under each of the six data configurations are presented in tables 58 to 63. The direction of the differences between the baseline proportional and equal condition models tends to be consistent across all six models within each of the data configurations, with one minor exception. While the direction of the differences is generally consistent, their magnitude can vary greatly.

Table 58. Simulation 1 proportional condition versus equal condition t-test results

Model	Mean Values		Difference	*p*-value
	Proportional	Equal		
LOG	0.8659	0.8508	0.0151	1.29E-09
LDA	0.8658	0.8509	0.0149	9.34E-10
QDA	0.8656	0.8507	0.0149	2.40E-09
BPNN	0.8597	0.8397	0.0200	2.27E-08
MNN	0.8585	0.8408	0.0177	8.64E-07
RBFNN	0.8500	0.8343	0.0157	2.44E-05

The classificatory performance difference for each of the six models under the simulation 1 configuration is positive and significant (see table 58). The differences for all six of the models are relatively small, averaging approximately 1.31%. Although all three of the parametric models show slightly smaller differences than do the neural network models, there is no reason to believe that different mechanisms are responsible for the changes. As noted earlier, the simulation 1 populations are linearly separable except for the overlapping portions. This improvement in performance under the proportional condition may be due to the fact that more of the overlapping portion of the two populations is apportioned to the larger group.

The results for the simulation 2 configuration are opposite those found under simulation 1. The classificatory performance of each of the six models is statistically better under the equal condition than under the proportional condition (see table 59). This universal decrease in classification rates is likely because a larger number of observations are located within the overlapping portion of the two populations. In addition, the differences for the LOG and LDA models are much greater than the differences for the remaining models. One explanation for this result is that these models are unable to exploit the differences in the covariance structures of the two populations. Therefore, although the discriminant line shifts toward the group 1 centroid as a result of the proportionality shift, the discriminant function is unable to use the information from each of the separate covariance structures to determine where to place the discriminant line. Its final placement results in a higher misclassification of group 2 observations.

Table 59. Simulation 2 proportional condition versus equal condition t-test results

	Mean Values			
Model	Proportional	Equal	Difference	p-value
LOG	0.6447	0.7267	-0.0820	1.83E-30
LDA	0.6462	0.7322	-0.0860	9.11E-31
QDA	0.7615	0.7865	-0.0250	7.46E-17
BPNN	0.7215	0.7581	-0.0366	3.18E-05
MNN	0.7407	0.7730	-0.0323	6.74E-10
RBFNN	0.7370	0.7655	-0.0285	1.99E-07

The three parametric models perform significantly poorer under the proportional condition than under the equal condition for simulation 3 data (see table 60). Under the equal condition, the relative proportions of exemplars in group 1, the main part of group 2, and the contaminated part of group 2 (see figure 5c) are 0.50, 0.35, and 0.15, respectively. Under the proportional condition the relative proportions of exemplars change to 0.30, 0.49, and 0.21, respectively, for these same three groups. For the LOG and LDA models, this de-emphasis of group 1 in favor of group 2 leads to an increase in the covariance structure of the two groups and a shift of the discriminant line toward the group-1 centroid. This, in turn, leads to more of the main-part-group-2 observations being classified correctly, but higher misclassification rates for both group-1 observations and the contamination-portion-of-group-2 observations. The smaller decrease in the QDA model's performance appears to result from its focus on the main-part-of-group-2 under the proportional condition compared to an emphasis on group 1 under the equal condition.

Table 60 Simulation 3 proportional condition versus equal condition t-test results

Model	Mean Values		Difference	p-value
	Proportional	Equal		
LOG	0.6489	0.6712	-0.0223	0.0008
LDA	0.6469	0.6716	-0.0247	0.0003
QDA	0.7239	0.7601	-0.0362	4.03E-10
BPNN	0.7539	0.7384	0.0155	0.0124
MNN	0.7610	0.7552	0.0058	0.1885
RBFNN	0.7464	0.7405	0.0059	0.2485

The three neural network models, on the other hand, all improve their absolute classificatory performance under the proportional condition, although only the MNN model does so significantly. Although the neural networks are less than perfect, they appear to identify and isolate the three regions of the data under both the equal and proportional conditions. Their improved performance under the proportional condition can be explained by a decrease in the number of group 1 observations overlapping with group 2 data regions.

All six of the classification models perform significantly better under the proportional condition for simulation 4 (see table 61). The average difference between the equal and proportional classificatory performance for the parametric models (20.63%) is slightly more than 3.5 times as large as that for the three neural network models (5.89%). Under the equal condition, the linear models (LOG and LDA) place the discriminant line halfway between the group-1 and group-2 centroids. Because the group-1 centroid is to the right of the group-2 centroid, all observations to the right of the discriminant line are assigned to group 1 and those to the left of the discriminant line are assigned to group 2. (The theoretical group-1 centroid is at 10.25,5 and the theoretical group-2 centroid is at 9.75,5.) This assignment strategy results in fewer than 50% of the observations being correctly classified. Under the equal condition, the QDA model can do no better than to isolate the major portion of either group 1 or group 2, depending on the particular sample characteristics. This also leads to a classification rate of less than 50%. The large jumps in classification rate for the parametric models occur because of the shift in the relative proportions of observations in favor of group 2 as opposed to an improvement in the discriminatory power of the models. Under the proportional condition, all three parametric models cannot but assign all observations to the more populous group, thereby, improving their classificatory performance to nearly 70% each.

The neural network models are able to isolate the four data regions under both the equal and proportional conditions. The nearly uniform improvement in their classificatory performance under the proportional condition appears to be caused by the redistribution of the observations, which results in better definition of the more populous group and less data overlap.

Table 61. Simulation 4 proportional condition versus equal condition t-test results

Model	Mean Values		Difference	*p*-value
	Proportional	Equal		
LOG	0.6996	0.4896	0.2100	1.92E-20
LDA	0.6997	0.4896	0.2101	1.92E-20
QDA	0.6977	0.4990	0.1987	1.04E-24
BPNN	0.7194	0.6689	0.0505	1.09E-07
MNN	0.7075	0.6376	0.0699	7.26E-06
RBFNN	0.7113	0.6551	0.0562	1.91E-14

For simulation 5, the coincident group-1 and group-2 theoretical centroids, coupled with no correlation between the independent variables, leave the linear models (LOG and LDA) with only the ability to split the data regions symmetrically. This results in a classification rate of approximately 50% under the equal condition (see table 62). Under the proportional condition, the LOG and LDA models assign essentially all observations to group 2, thereby, significantly improving their performance by nearly 20%. In both the equal and proportional conditions, the QDA model is able to place an ellipse around the group 2 data regions. The slight but significant decrease in performance under the proportional condition occurs because of a decrease in the group 1 classification rate. This decrease more than offsets the concomitant increase in the group 2 classification rate, caused by the proportional dominance of group 2 in the assignment function.

The three neural network models each improve their performance under the proportional conditions. An increase in the relative density of group 2 observations and a reduction in the number of group 1 observations occupying group 2 space results in an improvement of the neural network classificatory performance.

Table 62. Simulation 5 proportional condition versus equal condition t-test results

	Mean Values			
Model	Proportional	Equal	Difference	p-value
LOG	0.7008	0.5029	0.1979	2.47E-43
LDA	0.7008	0.5028	0.1980	2.15E-43
QDA	0.7628	0.7731	-0.0103	0.0489
BPNN	0.8163	0.7818	0.0345	1.34E-08
MNN	0.8265	0.7793	0.0472	9.45E-13
RBFNN	0.8205	0.7757	0.0448	5.10E-15

The performance results shown in table 63 indicate that there was significant improvement in the performance of all six classification models when going from the equal to the proportional condition for the D/FW mode choice data. The primary reason for the better performance is the improved domain coverage afforded by the additional exemplars for groups 1, 3 and 4 under the proportional condition. Unlike the five simulation conditions described above, the total numbers of exemplars available for training and validation of the models was not the same under the equal and proportional conditions for the D/FW mode choice problem. Under the proportional condition, there were an additional 7,323 exemplars available to assist in the models' training and validation.

Table 63. D/FW mode choice proportional condition versus equal condition t-test results

Model	Mean Values		Difference	p-value
	Proportional	Equal		
LOG	0.7036	0.5676	0.1360	2.39E-25
LDA	0.6872	0.5595	0.1277	1.83E-29
QDA	0.7055	0.5757	0.1298	1.95E-28
BPNN	0.7346	0.5380	0.1966	3.59E-31
MNN	0.7479	0.5889	0.1590	3.83E-22
RBFNN	0.6993	0.5504	0.1489	1.27E-20

In summary, there are significant differences between the equal and proportional condition models in 34 out of 36 (94.4%) cases. In terms of the magnitude and direction of the performance changes, the results for the LOG and LDA models are generally the same for all six data configurations. All of the performance differences are significant for these models Increases occur for the simulation 1, 4, and 5 configurations and the D/FW mode choice configuration and decreases occur for the simulation 2 and 3 configurations.

Under simulation 1, 3, and 4 configurations, the QDA model behaves very much like its linear parametric counterparts in terms of the direction and amount of classificatory performance change. The same can be said concerning the D/FW mode choice data but it must be noted that the performance differences for the RBFNN and

MNN models are not very different from those for the three parametric models. Under the simulation 2 configuration, the QDA model, while changing in the same direction as the two linear models, changes in the same direction and order of magnitude as the three neural network models. The behavior of the QDA model under simulation 5 appears to resemble that of the three neural network models except that its performance declines slightly when moving from the equal to the proportional condition while the neural networks improve their performance. This difference may be accounted for by the neural networks' abilities to isolate the disjoint regions of group 2 while the QDA model must treat them as one unit

The differences in performance levels between the equal and proportional conditions are remarkably consistent. This is true for all six data configurations and holds for both the direction and magnitude of the changes. Not once does one of the neural networks change its performance in a different direction from the other two neural networks under a given data configuration. For the five simulation configurations, the largest spread in performance change is 1.94% for simulation 4. The largest difference for the D/FW mode choice configuration is 4.77%. This percentage is still small when compared to the 20.83% difference found between the LDA and QDA models under simulation 4. The similarity of reaction to the proportional shift in observations implies that three neural network models included in this study use proportionality information in similar ways.

Summary

This chapter presents the results of the 47 experiments performed during this study. These experiments include those comparing the performance of the six classification models under the equal and proportional data conditions. The comparisons are reported for the five simulations and the D/FW data configurations. Statistically, the QDA and MNN models outperform their parametric and neural network counterparts respectively. The QDA models are statistically the best or equivalent to the best parametric models in each of the six equal and six proportional conditions. The MNN models are statistically the best or equivalent to the best neural network models under all but the simulation 4 equal and proportional conditions. Under simulation 4 the BPNN models were statistically best.

The iterative process results are a little less consistent than those for the baseline experiments. In only 2 of the 35 iteration experiments is the baseline model statistically the best. In four more of the experiments, the baseline model has the highest absolute classification rate but is not statistically different from one or more of the corresponding iterative models. Nine of the baseline models improved as a result of the iterative process, but not statistically so. The remaining 20 improved both absolutely and statistically as a result of the iterative process. Parametric models account for 14 of the 20 significantly improved models. In each of these 14 models, the best performance comes as a result of pairing with either first iteration (4 models) or second iteration (10 models) MNN output. In three of the four cases in which first iteration MNN output yields the best classificatory performance, the results with second iteration MNN output are not statistically inferior. Under the simulation 1 configuration, improvement to the three neural network models results from pairings with LDA output. Under the simulation 2 configuration, neural network improvements result from pairings with QDA output. That is, the pairings with the Bayes optimal models yield the most improvement in the neural network models under the more ideal data conditions. Under simulation 3, 4, and 5 configurations, there are no significant improvements in any of the neural network models. Finally, all five of the D/FW mode choice configuration models are improved significantly over their baseline performances by incorporating the results of a complementary model.

Lastly, the results for the equal versus baseline proportional conditions indicate that, depending on the population data conditions, it is sometimes better to have equal numbers of observations from each group and sometimes better to have a proportional representation of population groups. Whether equal or proportional representation is best seems to depend on the degree of group overlap, the covariance structures of the groups, and the domain coverage of the observation space. If the D/FW mode choice results are indicative of those expected from real-world data sets, the domain coverage seems to be the dominant factor affecting classificatory performance.

CHAPTER 5

DISCUSSION OF RESULTS

This study involves a broad-based investigation of six classification models that have been shown to have potential in solving statistical classification problems. This study is unique in that thoroughly investigates a number of research questions that have not been adequately addressed to date. For example, a thorough search of the literature uncovered no studies that compared the classificatory performance of as many parametric and neural network models. Especially unusual is the inclusion of the logit (LOG), modular neural network (MNN), and radial basis function neural network (RBFNN) models with the more common linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and backpropagation neural network (BPNN) models.

The classificatory performance of these models was compared under a number of data configurations and observation proportionality conditions as detailed previously. The application of the three neural network models to one of these data configurations, D/FW transportation mode choice decisions, is a new application. More generally, according to current literature, at least two of the neural network models, MNN and RBFNN, have not been applied to any transportation mode choice classification problems. Typically, only the LOG model is used for mode choice classification problems.

Another unique contribution of this study is the introduction of an iterative process for improving the classificatory performance of parametric and/or neural network models which have been combined into model pairs. Markham and Ragsdale (1995) included Mahalanobis distances as inputs to a BPNN model, for the purpose of classifying observations from two real-world data sets into one of two groups. The current study is different and new in that it passes both parametric and neural network outputs between model pairs in an iterative fashion. This allows a more complete assessment of the effect of the classification results of one type of model on the classificatory performance of a complementary type of model.

This chapter continues with a discussion of the results of the various classification model comparisons and their relationship to the three research questions posed in this study. A discussion of the study's research contributions is next, followed by an explication of the limitations of the study. The chapter ends with an examination of some future research opportunities.

Conclusions

Research Question #1

> How effective is an integrative approach that uses classification results from both neural network and statistical models for data classification, compared to the corresponding stand-alone procedures? For which data configuration(s) is this approach effective and likely to show the greatest potential for improved classificatory performance?

This study used an iterative procedure (see figure 4) to integrate the results of one classification model into the classification efforts of a complementary model. The overall results indicate that baseline classification models can, in general, be improved upon absolutely, and often significantly, by incorporating the results of a complementary model in the model estimation process. This having been said, it is also clear that this process can, and sometimes does, lead to significantly poorer classificatory performance. Even though improved classificatory performance is generally possible using the iterative process, the specific pair of models yielding the best absolute classification results varies, depending on the data conditions.

Markham and Ragsdale (1995) used the classification results of a linear discriminant analysis model (Mahalanobis distances) as inputs to BPNN models. For their two small real-world data sets, they found significantly better results from the integrated BPNN model than for either a LDA or BPNN model alone. In fact, not only were their overall results significant, but for every replication performed for each of their data sets, the classificatory performance of the baseline BPNN was better than that of the Mahalonobis distances and the integrated BPNN was better than the baseline BPNN model. While there are several data configurations in the current study for which the baseline BPNN model performs better than the corresponding LDA model (the equivalent of Mahalanobis distances) for all replications (e.g., the simulation 2 and D/FW mode-choice proportional configurations), the corresponding first iteration BPNN model results are not better than the baseline results across all replications for any data configuration. In fact, under the simulation 3, 4, and 5, and D/FW mode choice configurations, the first iteration BPNN model results are significantly below the corresponding baseline results. These results do not lend support to Markham and

Ragsdale's hope that integrating LDA output into a BPNN model will always result in superior classificatory performance.

The iterative process, an operationalization of classification model integration, led to mixed results. In 15 of the 35 iterative procedure experiments, a second iteration model resulted in the best absolute performance. In 15 of the remaining 20 experiments, the corresponding second iteration model did not result in statistically lower performance than the first iteration model. For these pairings, the second iteration is unnecessary unless even the smallest absolute performance improvement is of value. In 14 of the 35 experiments, a first iteration model performed best, implying that the second iteration, for the corresponding model pairing, is unnecessary. In addition, of the 102 first and second iteration performance means estimated, 33 of the corresponding first and second iteration models exhibited lower classification rates than the baseline model and for 14 of these pairs the performance decrease was significant. So, while the integration of models often leads to significantly improved performance, these results imply that model integration can just as easily lead to significant performance deterioration, depending on the selection of complementary model types for a given data condition.

Table 64 presents the results of the best performing baseline and overall model for each of the six data configurations. Because of the large number of models (6 baseline and 36 iterative models for each simulation, 6 baseline and 24 iterative models for D/FW mode choice), it would not have been practical to determine statistically the best overall model for each configuration. However, for all but one of the data configurations, simulation 2, the baseline performance was improved upon absolutely as a result of the iterative process. For three of the data configurations, simulation 1, simulation 2, and D/FW mode choice, the baseline and best overall model types were the same. For two of these data configurations, simulation 1 and D/FW mode choice, the baseline models were improved significantly (LOG by 1.83% and MNN by 1.76%, respectively). Generally speaking, the improvements across the six data configurations were modest at best, averaging less than 1% (~0.94%). Even when simulation 2 is not factored in, the average change is only approximately 1.13%. Therefore, if a practitioner or researcher is willing to test the entire array of classification models used in this study to determine which is best, the incremental improvement resulting from the iterative procedure would only be of benefit when the cost of misclassification is very high.

Except for the data configurations that are ideally suited to one or more of the parametric models, simulations 1 and 2, the best performing baseline models are neural networks. Three of four of these best performing neural networks are MNN models. Under simulation 2 conditions, the MNN baseline performance results are second only to those of the QDA model and after one iteration, the QDA/MNN paired model produces results that are not statistically distinguishable from the baseline QDA model. Each of simulations 3, 4, and 5 were designed to not be linearly separable and in each of these cases, the linear baseline models' classification rates were statistically below each of the neural network models. Yet, after iterating with MNN output, both of these linear models improved their performance significantly and one of them had the highest classification rate of any of the six models. The best performing linear models, having been paired with MNN output, improved on their baseline performance by an average of

9.24%. The implication of these results is that regardless of the data configuration (at least across the range represented in the current study) the neural network model of choice should be a MNN instead of the more commonly used BPNN. A second implication is that even when the populations are known not to be linearly separable, linear parametric models (such as LOG and LDA) should not be discounted.

Table 64. Best performing baseline and overall models for all data configurations

	Baseline		Overall			
Data Configuration	Model	Performance	Model	Performance	Iteration	Paired Model
Simulation 1	LOG	0.8659	LOG	0.8842	2	MNN
Simulation 2	QDA	0.7615	QDA	0.7615	baseline	none
Simulation 3	MNN	0.7610	LDA	0.7713	1	MNN
Simulation 4	BPNN	0.7194	LOG	0.7212	2	MNN
Simulation 5	MNN	0.8265	LOG	0.8348	2	MNN
D/FW mode choice	MNN	0.7479	MNN	0.7655	1	QDA

Research Question #2

> What is the relationship between the relative proportion of training, testing, and validation exemplars in each of the output classes and the performance of each of the classification models? At what proportional mix does a given classification method perform best?

In real-world applications, the relative proportions of exemplars in the various groups can vary greatly. For example, in the 'new firm' failure data set used by Jain and Nag (1997), the proportional split was 71% failed and 29% successful businesses. In the classic Iris data set used by Fisher (1936), there were equal numbers of observations in each of three groups. How various classification models react to different proportional mixes of exemplars is an important question. The current study provides some insights concerning the performance characteristics of the six classification models included for two proportional mixes under each of the six data configurations.

To begin with, it appears that the three neural network models are less sensitive than the parametric models, especially the LOG and LDA models, to changes in the proportional mix of the classification groups. For every data configuration except simulation 2, all three neural network models exhibit higher absolute classification rates under the proportional condition. Although most of the differences are significant, they are generally smaller than those of the parametric models. This appears to be because of two factors. First, the neural network models are able to isolate disjoint regions representing the same group regardless of the proportional mix or the distorted covariance structure. Thus, for data configurations such as simulations 3, 4, and 5, they are not confused by the contamination or bimodal regions and are able to enclose each data region with hyperplanes used for determining group membership. Second, the neural network models, unlike the parametric models, do not use group centroids in their classification algorithms. Therefore, centroids distorted by the addition of contamination, do not directly influence the classification of exemplars by the neural network models

The results indicate that all three of the neural network models perform better under the proportional condition except for the simulation 2 configuration. Their improved performance appears to be a result of two factors. First, the increased number of exemplars in group 2 under the proportional conditions allows for improved definition of the group 2 boundaries and therefore, by default, the corresponding group 1 boundaries. This appears to be especially true for the D/FW mode choice configuration.

Second, a smaller number of exemplars from group 1 exists in the overlapping regions under each of the simulation configurations, thereby allowing the neural networks to assign those spaces to the dominant group with a smaller resulting misclassification penalty.

The results for the three parametric models are not as clear as those for the neural networks. The LOG and LDA models appear to be useful for only the configurations in which there is some reasonable means to linearly separate the groups. They perform better under the proportional condition when the groups have the same covariance structure (simulation 1 and D/FW mode choice configurations) and perform better under the equal condition when the covariance structures differ (simulation 2 and 3 conditions). The QDA model exhibits equal condition/proportional condition performance patterns that mirror those of the LOG and LDA models, except for the simulation 5 case. Under the simulation 5 configuration, the QDA model is able to encircle the bimodally distributed group 2 exemplars under both the equal and proportional conditions. The slight, but significant, decrease in the QDA model's classificatory performance under the proportional condition is likely a result of an increase in the area allocated to group 2 which in turn increases the misclassification rate for group 1 exemplars.

Research Question #3

> How will the six classification models perform relative to each other under each of the six data configurations (D/FW transportation mode choice and five simulation configurations) within each of the two class proportion conditions, equal and proportional (25/25/25/25 and 33/8/38/21 for the 4-class D/FW transportation mode choice configuration and 50/50 and 70/30 for the simulation configurations)?

Models, including those used for classification applications and research, are often selected because they are the only ones known by the investigator or because they have become a *de facto* standard for the particular task at hand. For business applications and research, the LDA and BPNN models have become the primary parametric and neural network standards, respectively. This appears to be the case irrespective of the group data characteristics. Is their status as *de facto* standards warranted? The results of the equal condition and baseline proportional condition comparisons done in this study would indicate not.

As expected, the LOG, LDA, and QDA models perform well under the simulation 1 configuration whether the groups have equal numbers of exemplars or not. The LOG and LDA models are, after all, optimal classifiers under conditions of normality with equal covariance structures. The essentially equivalent performance of the QDA models under the simulation 1 configuration is likely due to the relatively large number of exemplars (200 total) in the training data sets. As the number of training exemplars increases, the relative advantage of pooling the covariance matrices diminishes, thereby reducing the advantage of techniques such as LDA.

While the three neural network models performed statistically more poorly than their parametric counterparts under the simulation 1 configuration, from a practical perspective, they each appear to be able to isolate the two groups nearly as well as the parametric methods. Thus, all six of the models appear to be viable alternatives under conditions of bivariate normality with equal covariance structures, unless even a very small performance advantage is important.

When the covariance structures of the two bivariate normal groups differ, the QDA model becomes the best choice. This is not unexpected since it was developed specifically for this condition. The more commonly used parametric methods, LDA and LOG, are unable to match not only the performance of the QDA model, but also that of the three neural network models. This is true regardless of the proportional mix of exemplars. The LDA and LOG models are unable to exploit the differences in covariance structures that allows the QDA model to offset the inclination to assign exemplars to the group with the larger values on the diagonal of its covariance matrix. Of note regarding the neural network models is that the BPNN performs poorest of the three under both the equal and proportional conditions, and in both cases its performance is significantly below that of the corresponding MNN models. The BPNN model, therefore, should not generally be the neural network model of choice when simulation 2 conditions prevail.

The ability to linearly separate groups 1 and 2 is greatly diminished under simulation 3 relative to simulations 1 and 2. The classificatory performance of the LOG and LDA models is expectedly reduced under this condition because they are unable to recognize that group 2 is not contiguous and have no mechanism for isolating the group 2 contamination region. This problem with recognizing disjoint regions extends to the simulation 4 condition and becomes so acute that the LOG and LDA models lose all ability to discriminate between the two groups regardless of the proportional mix of exemplars. It appears, in fact, that if the intervening space between disjoint regions of a group is populated with exemplars from another group, the LOG and LDA models lose their discriminatory power almost completely.

The QDA model was not affected in the same way as the LOG and LDA models by the addition of the single disjoint region of contamination under the simulation 3 configuration. In fact, under the equal condition, the QDA model demonstrated the best absolute performance of any of the six models although it was not statistically superior to the MNN model. The QDA model maintains its discriminant ability because the contiguous space occupied by group 1 can be isolated quadratically (except for the regions overlapping group 2). With the addition of the second contamination region under the simulation 4 configuration, the QDA model loses its ability to discriminate between the two groups. This is true under both the equal and baseline proportional conditions. Its performance characteristics mirror those of the LOG and LDA models and therefore, it should be eliminated from contention if simulation 4 conditions prevail.

The first four simulation configurations are formulated as variations of the same basic scheme. Each of these simulations is based primarily on two bivariate normally distributed groups, correlated with $r = 0.3$ and non-coincident means. The variations involve changes in covariance structure and/or the addition of bivariate normally distributed contamination. The simulation 5 configuration, on the other hand, presents an entirely new data configuration. Only group 1 under simulation 5 is bivariate normally distributed and unlike the other simulation configurations, the independent variables are uncorrelated. Group 2, on the other hand, is bimodal. It is a composite of two bivariate normally distributed distributions and has a mean that is coincident with that of group 1. The coincident means coupled with exact symmetry and almost complete overlap of the two data groups renders the LOG and LDA models useless as discriminant tools. As was the case under the simulation 4 configuration, they are only capable of assigning essentially all exemplars to one group.

The QDA model and the three neural network models, unlike the LOG and LDA models, are each able to discriminate between the two simulation 5 configuration groups. In both the equal and baseline proportional conditions, the three neural network models appear to better isolate the bimodal group 2 exemplars than does the QDA model. This is especially true under the baseline proportional condition where the larger number of group 2 exemplars provides better definition of the group 2 boundary. In both the equal and baseline proportional conditions there are only small differences among the performance levels of the three neural network models and therefore, any of these models might be used under similar data conditions. While the BPNN model shows slightly better absolute performance than the MNN model under the equal condition, under the more realistic baseline proportional condition the MNN model performs significantly better than the BPNN model. This, again, brings into question the widespread use of the BPNN model in most business applications and research. The results under the simulation 5 configuration also, again, lend support for the use of neural network models over their parametric counterparts.

The D/FW mode choice data presents the most realistic challenge to the six classification models. The equal condition was fabricated to demonstrate classification model performance characteristics that might be expected when real-world data presents itself in generally equal proportions and in amounts too small to completely define the group structure. Under these conditions, the MNN model performance is significantly superior to each of the other five models. The MNN model appears to require the least amount of information to isolate the four groups. This might be because of its ability to break the classification problem into smaller pieces, each of which requires fewer exemplars. The BPNN and RBFNN models, on the other hand, demonstrate the poorest classificatory performance with the BPNN model performing statistically below all of the other models.

When the D/FW mode choice data are used in their natural proportions (i.e., the baseline proportional condition), the MNN model still provides statistically the highest classificatory performance of any of the six models. Under this condition, the BPNN model moves from last to second best. In fact, it shows the largest increase in performance of any of the models, which implies that it is more sensitive to domain coverage than the other models under these data conditions. Again, the RBFNN model performs relatively poorly compared to the best models. It is impossible to determine precisely why this is the case, but one explanation is that border exemplars were placed in multi-group clusters rather than same group clusters during the unsupervised portion of the RBFNN training. The parametric models maintain the same relative order among themselves (from best to worst: QDA, LOG, LDA) as under the corresponding equal condition, but the performance gaps among them is larger than for the best neural network models. This makes them a less desirable alternative under the more realistic conditions.

To summarize the results concerning research question 3, when the data conditions are nearly ideal for a given parametric model (simulation 1 configuration for LDA and LOG and simulation 2 configuration for QDA) those models are viable classification alternatives. In fact, if the objective is to select the best baseline model, the parametric alternatives would be preferred under these ideal conditions. As the data conditions deteriorate, however, the BPNN and MNN models become more appealing alternatives. Given that the real world rarely provides us with ideal distributional structures, the current results would indicate that for most real-world classification tasks, neural networks should be accorded significant consideration.

Research Contributions

Much research has been done in an attempt to better understand the nature of classification. Efforts have been made to place bounds on the types of data environments amenable to classification and to characterize the efficacy of the many available classification models and algorithms. In addition, researchers have investigated the utility of various new classification models and extensions of existing models in an attempt to improve their speed and/or accuracy. The current study contributes to these efforts in three unique ways.

The first contribution concerns the data used to evaluate the classificatory performance of the various models and the iterative procedure. While two of the simulated data configurations (simulations 1 and 2) are commonly found in the classification research literature, the current study extends the simulated data configurations in particular directions not hitherto investigated. The systematic violation of the ideal conditions of simulations 1 and 2 by the introduction of contamination of a ghost-like or shadow-like nature, as in the simulation 3 and 4 configurations, provides a direct means of testing the robustness of each of the six classification models. The use of live transportation mode choice data is also unique to this study. This not only provided an opportunity to evaluate the relative performance of the six classification models on a realistic classification task, but also contributes to a better understanding of their potential efficacy in a transportation-planning context. This evaluation also represents a first step in an attempt to determine whether or not the LOG model deserves its preeminent place in transportation mode choice modeling.

The second contribution is the inclusion of a unique mix of classification models. Not only were the more commonly evaluated models such as LDA, QDA, and BPNN included, but also rarely used models such as the MNN and RBFNN models were evaluated in this study. While some studies have focused attention on the MNN or RBFNN models either alone or in comparison with a BPNN model, few if any have compared the three neural network models in a single study or evaluated their performance against the three parametric models included in this study. Thus, the current study allowed for an assessment of the heretofore primacy of the LDA and BPNN models.

The third contribution is the introduction of the iterative procedure for developing classification models. While Markham and Ragsdale (1995) found that results from a linear classification model (LDA) could be of value as inputs to a BPNN model, the current study tested the utility of not only LDA outputs but also LOG and QDA outputs as inputs to each of the three neural network models. In addition, the iterative procedure allowed for a determination of the efficacy of neural network outputs as inputs to each of the three parametric models. The evaluation of the iterative procedure also provided information concerning the relative usefulness of specific pairings of parametric and neural network models.

Study Limitations

While the merits of the current study are many, there are a number of potential and actual limitations that need to be addressed. Every effort has been made to ensure that there are no internal validity problems. No human subjects were used in this study (although the cross sectional D/FW mode choice data was captured from human subjects) so issues of internal validity such as history, maturation, testing, and experimental mortality (Campbell and Stanley 1963) are not an issue. The treatments applied to the simulated and live data sets were computer based classification algorithms, which for a given set of data are deterministic in nature. That is, given the same random number generator seed, the results of each of the six classification models are infinitely repeatable for a given data set whether simulated or live.

Issues concerning construct validity with respect to the models selected has been addressed in supporting studies demonstrating that these models can reasonably estimate the posterior probabilities of group membership (see chapter 2 for details). In addition, the average classification rate used in this study as the measure of classificatory performance, the number of exemplars correctly classified divided by the total number of attempted classifications, is the measure most commonly found in this type of research.

As is the case for any simulation study, there is concern as to the generalizability or external validity of the current results. As indicated earlier, it is highly unlikely that any of the data conditions simulated in this study would be realized in more natural circumstances. However, it is of value to understand how the various classification models and their iterative hybrids react to known data configurations and adulterations. This understanding allows future researchers and practitioners to more intelligently select classification models based upon the similarity of their data to the idealized data used for testing the models in this study.

The D/FW mode choice data presents a slightly different generalization problem. There is no question as to the authenticity of the data, and to the extent that it was collected with appropriate care, of its validity. In any case, this study does not purport to guarantee the reasonableness of the data but to test the efficacy of the various classification models and the iterative procedure. This said, however, it is important to realize that these data are unique to a given time (1984) and location (D/FW metroplex) and that the results might differ for other times and/or locations even though the classification domain may be the same. It is also possible that the results would differ if different independent mode choice variables were to be included in the classification task. While the results would not likely remain constant over the complete range of times, locations, and variable choices, it is true that individuals attempt to maximize the utility of their mode choices regardless of the particular time and location conditions.

Therefore, the current results provide a useful starting point for the mode choice model selection process. Further tests with other mode choice data sets will need to be performed before more definitive statements can be made regarding the generalization of the results obtained in this study.

Another limitation concerning the generalizability of the results is the number of training, testing, and validation exemplars used for model development and validation. There are two aspects to this issue. First, is the absolute number of each type of exemplar generated for the simulated data sets. The results are tied to having a total of 200 exemplars to train the parametric models and 140 training and 60 testing exemplars for developing each of the neural network models. Larger or smaller numbers of exemplars or the choice of a different split for the training and testing data sets might yield different results. The second aspect of this issue is the choice of the proportions of exemplars used for groups 1 and 2 under the simulation configurations and the 4 groups under the D/FW mode choice configuration. While the 50/50 and 30/70 splits were deemed reasonable for the simulated data (Jain and Nag 1997), other splits might result in different performance levels of the various classification models. The same could be said of the 25/25/25/25 and natural splits used for the D/FW mode choice configuration, although in this case the natural split does reflect the actual survey results.

The particular parameter values selected for the various neural network models provide another limitation of the current study. The selection of parameter settings for each of the types of neural networks can have a dramatic affect on the classification results. Currently, no agreed upon procedure exists for finding optimal parameter settings for any of the neural network models used in this study. While much care was taken in the parameter setting selection process, there was no way to guarantee that the final parameter settings were in fact optimal. Hundreds of models were developed to test various combinations of such values as the number of hidden layer neurons, the learning rule, the connection scheme, the transfer function, error offsets, the number of local experts, and the number of prototype neurons, as appropriate for the given type of neural network. However, many parameters were not varied during the testing process because of time limitations or past experience. It is possible that other parameter settings would have yielded different results. For example, the relatively poor showing of the RBFNN models *vis-à-vis* the BPNN and MNN models might be because of a poor choice for the number of prototype neurons or the combination of some number of other parameter values.

Future Research

A number of potentially viable research streams present themselves as a result of the current research findings. Two of these that appear to be especially promising are detailed below. The first of these involves the iterative procedure used to improve the performance of both parametric and neural network models. In the current study, the iterative procedure always started with the baseline output of a parametric model being added as additional input to a neural network model.

Future research might examine the impact of reversing the direction of information flow, i.e., passing baseline neural network output first to a parametric model. The current study used only two iterations of each model, which while generally yielding improved classification results, did not do so in all cases. In addition, a wide variety of performance patterns were observed for different model pairs as a result of iteration. In general, no systematic patterns could be discerned from the iteration results. Future research could use a delta value rather than an absolute number of iterations as a termination criterion so that any systematic patterns might be manifested more completely.

The current study used only the overall classification rate as a performance measure. While this is the most widely used measure of classificatory performance, the results of this study imply that the iterative procedure can differentially affect not only the overall classification rates of the models, but also the group specific classification rates. Given that many applications (e.g., medical diagnosis) require a focus on the classification rates within one or more of the groups involved as opposed to only the overall rate across groups, a future study might investigate the impact of the iterative procedure on these classification rates. In the current study, the average absolute classification performance across a number of trials was investigated. Another measure of performance that might be investigated is the variation in classification rates across replications. Even if there is little increase in the absolute performance of a classification model resulting from the iterative procedure, a decrease in the variation of the results across replications would indicate a more stable and reliable and therefore desirable model.

The results generated using the D/FW mode choice data point to a second potentially valuable research stream. Transportation planning is a costly but critical endeavor practiced worldwide in a nearly uniform manner. The success with which transportation decisions, including mode choice, can be modeled, directly affects the evolution of transportation networks and options and therefore the quality of life in our urban centers. The current results imply that models other than the traditional LOG model might be viable alternatives for mode choice modeling. To test this, mode choice data from other urban areas should be included in a study to help determine the efficacy of the various classification models and the iterative procedure used in the current research.

REFERENCES

Ahmad, S., and V. Tresp. 1996. Classification with missing and uncertain inputs. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 63-68. New York: IEEE.

Altman, E. I., G. Marco, and F. Varetto. 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. *Journal of Banking & Finance*. 18: 505-529.

Anand, R., K. G. Mehrotra, C. K. Mohan, and S. Ranka. 1993. An improved algorithm for neural network classification of imbalanced training sets. *IEEE Transactions on Neural Networks*. 4: 962-969.

Anderson, J. A. 1972. A simple neural network generating an interactive memory. *Mathematical Biosciences*. 14: 197-220.

Anderson, J. A., J. W. Silverstein, S. A. Ritz, and R. S. Jones. 1977. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*. 84: 413-451.

Asmis, E. 1984. *Epicurus' Scientific Method*. Ithaca, NY: Cornell University Press.

Birley, S., and Westhead, P. 1994. New producer service businesses: Are they any different from new manufacturing ventures? *Service Industry Journal*. 14: 455-481.

Bonelli, P., and A. Parodi. 1991. An efficient classifier system and its experimental comparison with two representative learning methods on three medical domains. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, University of California, San Diego, July 13-16, 1991, ed. R. K. Belew and L. B. Booker, 288-295. San Mateo, CA: Morgan Kaufman.

Bounds, D. G., and P. J. Lloyd. 1990. A comparison of neural network and other pattern recognition approaches to the diagnosis of low back disorders. *Neural Networks*. 3: 583-591.

Bridle, J. S. 1989. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neuro-computing: Algorithms, architectures, and applications*, ed. F. Fogelman-Soulie and J. Herault, 227-236. New York: Springer.

Burns, J. O., and S. M. Groomer. 1983. An analysis of tax court decisions that assess the profit motive of farming-oriented operations. *The Journal of the American Taxation Association*. 4: 23-39.

Campbell, C. M. III. 1997. The determinants of dismissals, quits, and layoffs: A multinomial logit approach. *Southern Economic Journal*. 63: 1066-1073.

Campbell, D. T., and J. C. Stanley 1963. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Publishing.

Carpenter, G. A., S. Grossberg, and D. B. Rosen. 1991. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*. 4: 759-771.

Carse, B., A. G. Pipe, T. C. Fogarty, and T. Hill. 1995. Evolving radial basis function neural networks using genetic algorithms. In *1995 IEEE International Conference on Evolutionary Computation, ICEC '95, Perth, Western Australia, 29th November-1st December 1995*. Piscataway, NJ: IEEE.

Cawley, G. C., and S. R. Dorling. 1996. Reproducing a subjective classification scheme for atmospheric circulation patterns over the United Kingdom using a neural network. In *Artificial Neural Networks - ICANN 96, 1996 International Conference, Bochum, Germany, July 16-19, 1996 Proceedings*, ed. C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, 281-286. Berlin Springer.

Ceccarelli, M., and J. T. Hounsou. 1996. RBF networks vs. multilayer perceptrons for sequence recognition. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 391-396. New York: IEEE.

Chen, F., and M. Lin. 1996. On the learning and convergence of the radial basis networks. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 281-286. New York: IEEE.

Clark, W. R., P. A. Lachenbruch, and B. Broffitt. 1979. How non-normality affects the quadratic discriminant function. *Communications in Statistics*. A8: 1285-1301.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*. 5: 559-584.

Curram, S. P., and J. Mingers. 1994. Neural networks, decision tree induction and discriminant analysis: an empirical comparison. *Journal of the Operational Research Society*. 45: 440-450.

Damodaran, P. S., S. S. Kolli, and S. M. Alexander. 1993. The investigation of new approaches to business data analysis. *Computers and Industrial Engineering*. 25: 545-548.

Das, S., A. B. Chattopadhyay, and A. S. R. Murthy. 1996. Force parameters for on-line tool wear estimation: A neural network approach. *Neural Networks. 9: 1639-1645.*

Deal, K., and S. J. Edgett. 1997. Determining success criteria for financial products: A comparative analysis of CART, logit and factor/discriminant analysis. *Service Industries Journal*. 17: 489-506.

Demaris, A. 1992. *Logit Modeling: Practical applications*. Sage University Paper series Quantitative Applications in the Social Sciences, series no. 07-086. Newbury Park, CA: Sage Publications.

Desai, V. S., J. N. Crook, and G. A. Overstreet Jr. 1996. *European Journal of Operational Research*. 95: 24-37.

de Villiers, J., and E. Barnard. 1992. Backpropagation neural nets with one and two hidden layers. *IEEE Transactions on Neural Networks*. 4: 136-141.

Dougherty, M. S., H. R. Kirby, and R. D. Boyle. 1994. Using neural networks to recognise, predict and model traffic. In *Artificial Intelligence Applications to Traffic Engineering*, ed. M. Bielli, G. Ambrosino, and M. Boero, 233-250. Utrecht, The Netherlands: VSP.

Doumas, A., K. Mavroudakis, D. Gritzalis, and S. Katsikas. 1995. Design of a neural network for recognition and classification of computer viruses. *Computer & Security*. 14: 435-448.

Dror, I. E., M. Zagaeski, and C. F. Moss. 1995. Three-dimensional target recognition via sonar: A neural network model. *Neural Networks*. 8: 149-160.

Dutta, S., and S. Shekhar. 1988. Bond rating: A non-conservative application of neural networks. *International Conference on Neural Networks*, 443-450.

Fisher, R. A. 1936. The use of multiple measurement in taxonomic problems. *Annals of Eugenics*. 7: 179-188.

Fletcher, S., and E. Goss. 1993. Forecasting with neural networks: An application using bankruptcy data. *Information & Management*. 24: 159-167.

Fritzke, B. 1994. Fast learning with incremental RBF networks. *Neural Processing Letters*. 1: 2-5.

Fukushima, K. 1975. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*. 20: 121-136.

Ghinganda, E. F., and K. Subrahmaniam. 1979. Robustness of the linear discriminant function to nonnormality: Johnson's system. *Journal of Statistical Planning and Inference*. 3: 69-77.

Ghislandi, A. C., A. R. Fijal, and E. J. Christopher. 1994. CATS 1990 household travel survey: Technical documentation for the household, person and trip files. *Working paper 94-0 4*. Chicago Area Transportation Study. Information Services Division. Chicago, Illinois.

Gibson, G. J., and C. F. N. Cowan. 1990. On the decision regions of multilayer perceptrons. *Proceedings of the IEEE*. 78: 1590-1594.

Glorfeld, L. W., and K. H. Ching. 1992. Issues in data representation and performance validation for neural networks. In *1992 Proceedings: Decision Sciences Institute, 1992 Annual Meeting, November 22-24, San Francisco, CA* , 741-743.

Golden, R. M. 1996. *Mathematical methods for neural networks and design*. Cambridge, MA: MIT Press.

Gorman, R. P., and T. J. Sejnowski. 1988. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*. 1: 75-89.

Grossberg, S. 1976. Adaptive pattern classification and universal recoding: Parallel development and coding of neural feature detectors. *Biological Cybernetics*. 23: 121-134.

Hebb, D. O. 1949. *The organization of behavior*. New York: Wiley.

Hergert, F., H. G. Zimmermann, U. Kramer, and W. Finnoff. 1992. Domain independent testing and performance comparisons for neural networks. In *Artificial Neural Networks, 2*, ed. I. Aleksander and J. Taylor, 1071-1075. Amsterdam: Elsevier.

Hernandez, R., A. Varfis, I. Kanellopoulos, and G. Wilkinson. 1992. Development of MLP/LVQ hybrid networks for classification of remotely-sensed satellite images. In *Artificial Neural Networks, 2*, ed. I. Aleksander and J. Taylor, 1193-1196. Amsterdam: Elsevier.

Hollis, J. E. L., D. J. Brown, I. C. Luckraft, and C. R. Gent. 1996. Feature vectors for road vehicle scene classification. *Neural Networks.* 9: 337-344.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences.* 79: 2554-2558.

Huang, C. S., R. E. Dorsey, and M. A. Boose. 1994. Life insurer financial distress prediction: A neural network model. *Journal of Insurance Regulation.* 13: 131-167.

Huang, W. Y., and R. P. Lippmann. 1987. Comparisons between neural net and conventional classifiers. *IEEE 1st International Conference on Neural Networks, San Diego, CA*, 485-493. New York: IEEE.

Hughes, S. 1993. Bankruptcy prediction models. *Credit Control.* 14: 16-22.

Hyer, N. and U. Wemmerlov. 1989. Group technology in US manufacturing industry: a survey of current practices. *International Journal of Production Research* 27: 1287-1304.

Jacobs, R. A. 1988. Increased rates of convergence through learning rate adaptation. *Neural Networks.* 1: 295-307.

Jacobs, R. A., and M. I. Jordan. 1993. Learning piecewise control strategies in a modular neural network architecture. *IEEE Transactions on Systems, Man, and Cybernetics.* 23: 337-345.

Jacobs, R. A., F. Peng, and M. A. Tanner. 1997. A bayesian approach to model selection in hierarchical mixture-of-experts architectures. *Neural Networks.* 10: 231-241.

Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton. 1991. Adaptive mixtures of local experts. *Neural Computation.* 3: 79-87.

Jain, B. A., and B. N. Nag. 1997. Performance evaluation of neural network decision models. *Journal of Management Information Systems.* 14: 201-216.

Jalel, N. A., A. R. Mirzai, J. R. Leigh, and H. Nicholson. 1992. Application of neural networks in process control. In *Neural Network Applications, Proceedings of the Second British Neural Network Society Meeting (NCM(!), London, October 1991*, ed. J. G. Taylor, 101-113. Berlin: Springer.

Jordan, M. I. and R. A. Jacobs. 1992. Hierarchies of adaptive experts. *Advances in Neural Information Processing Systems.* 4: 985-992.

Jordan, M. I., and L. Xu. 1995. Convergence results for the EM approach to mixtures of experts architectures. *Neural Networks.* 8: 1409-1431.

Kahane, L. H. 1996. Congressional voting patterns on NAFTA: An experimental analysis. *American Journal of Economics & Sociolgy.* 55: 395-409.

Kahn, Z. U., S. K. Chawla, and S. T. A. Cianciolo. 1995. Multiple discriminant analysis: Tool for effective marketing of computer information systems to small business clients. *Journal of Professional Services Marketing.* 12: 153-162.

Kandil, N., K. Khorasani, R. V. Patel, and V. K. Sood. 1996. Optimal learning rate for backpropagation neural networks. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 249-252. New York: IEEE.

Kastens, T. L., and A. M. Featherstone. 1996. Feedforward backpropagation neural networks in prediction of farmer risk preferences. *American Journal of Agricultural Economics.* 78: 400-415.

Kitano, H. 1994. Neurogenetic learning: An integrated method of designing and training neural networks using genetic algorithms. *Physica D.* 75: 225-228.

Klecka, W. R. 1980. *Discriminant Analysis.* Sage University Paper series Quantitative Applications in the Social Sciences, series no. 07-019. Beverly Hills and London: Sage Publications.

Knox, C. 1980. CAD/CAM and group technology: the answer to systems integration. *Industrial Engineering* 12: 66.

Kohonen, T. 1972. Correlation matrix memories. *IEEE Transactions on Computers.* C-21: 353-359.

Kosko, B. 1992. *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence.* Englewood Cliffs, NJ: Prentice-Hall.

Lashley, K. S. 1950. In search of the engram. In *Society of experimental biology symposium, No. 4: Psychological mechanisms in animal behavior*, 478-505. London: Cambridge University Press.

LeCun, Y. 1985. Learning processes in an aysmmetric threshold network. *Disordered Systems and Biological Organization*, ed. E. Bienenstock, F. Fogelman Souli, and G. Weisbuch, Berlin: Springer-Verlag.

Lee, D., J. S. Payne, H. Byun, and K. C. Persaud. 1996. Application of radial basis function neural networks to odour sensing using a broad specificity array of conducting polymers. In *Artificial Neural Networks - ICANN 96, 1996 International Conference, Bochum, Germany, July 16-19, 1996 Proceedings*, ed. C. von der Malsburg, W. von Seelen, J. C. Vorbruggen, and B. Sendhoff, 299-304. Berlin: Springer.

Lee, H. and C. Lai. 1993. Supervised fuzzy ART: Training of a neural network for pattern classification via combining supervised and unsupervised learning. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 143-148. New York: IEEE.

Leonard, K. J. 1993. Who's keeping score? A credit scoring model for commercial loan applications. *Business Credit*. 95: 8-12.

Levine, D. S. 1991. *Introduction to neural & cognitive modeling*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Li, E. Y. 1994. Artificial neural networks and their business applications. *Information & Mangement*. 27: 303-313.

Liao, T. W., and L. J. Gen. 1993. An evaluation of ART1 neural models for GT part family and machine cell forming. *Journal of Manufacturing Systems*. 12: 282-290

Lippmann, R. P. 1987. An introduction to computing with neural nets. *IEEE ASSP magazine*. 4: 4-22.

Longo, J. M., and M. S. Long. 1997. Using neural networks to differentiate between winner and loser stocks. *Journal of Financial Statement Analysis*. 2: 5-15.

Madeo, S. A. 1979. An empirical analysis of tax court decisions in accumulated earnings cases. *The Accounting Review*. 10: 538-553.

Malsburg, C. von der. 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*. 14: 85-100.

Manly, B. F. J. 1994. *Multivariate statistical methods: A primer*, 2nd ed. London: Chapman & Hall.

Markham, I. S., C. T. Ragsdale. 1995. Combining neural networks and statistical predictions to slove the classification problem in discriminant analysis. *Decision Sciences.* 26: 229-242.

McClelland, J. L., D. E. Rumelhart, and G. E. Hinton. 1986. The appeal of parallel distributed processing. In *Parallel Distributed Processing. Vol. 1: Foundations*, ed. D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, 3-44. Cambridge, MA: MIT Press.

McCulloch, W. S. and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics.* 5: 115-133.

McLachlan, G. J. 1974. An asymptotic unbiased technique for estimating the error rates in discriminant analysis. *Biometrics.* 30: 239-249.

Mendenhall, W., and T. Sincich. 1993. *A second course in business statistics: Regression analysis.* 4th ed. New York: MacMillan Publishing Company.

Mitchell, D. L., J. Yi, and S. Govind. 1996. Household trip forecasting using neural networks: An exploratory investigation. In *1996 Proceedings: Decision Sciences Institute, 1996 Annual Meeting, November 24-26, Orlando, FL*, 1117-1119.

Minsky, M. and S. Papert. 1969. *Perceptrons.* Cambridge, MA: MIT Press.

Moody, J. and C. J. Darken. 1989. Fast learning networks of locally-tuned processing units. *Neural Computation.* 1: 281-294.

Morley, C. L. 1994. Discrete choice analysis of the impact of tourism prices. *Journal of Travel Research.* 33: 8-14.

Musavi, M. T., W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels. 1992. On the training of radial basis function classifiers. *Neural Networks.* 5: 595-604.

Neruda, R. 1995. Functional equivalence and genetic learning of RBF networks. In *Artificial Neural Nets and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 53-56. Wien, Austria: Springer.

NeuralWare. 1995. *Neural computing: A technology handbook for Professional II/PLUS and NeuralWorks Explorer.* Pittsburg: NeuralWare.

________. 1995. *Reference guide: Software reference for Professional II/PLUS and NeuralWorks Explorer.* Pittsburg: NeuralWare.

North Central Texas Council of Governments. Transportation Department. 1984. *Findings of the 1984 Regional Travel Survey.* Arlington, Texas.

________. 1990. *Multimodal transportation analysis process (MTAP): A travel demand forecasting model.* Arlington, Texas.

Ortuzar, J. de D., and L. G. Willumsen. 1994. *Modelling Transport.* 2nd ed. New York: Wiley.

Owen, W. 1992. Transportation and society. In *Transportation Planning Handbook* ed. J. D. Edwards, 1-18. Englewood Cliffs, NJ: Prentice-Hall.

Parker, D. B. 1982. Learning logic. *Invention Report, S81-64, File 1.* Office of Technology Licensing, Stanford University.

Patuwo, E., M.Y. Hu, and M. S. Hung. 1993. Two-group classification using neural networks. *Decision Sciences.* 24: 825-845.

Pavur, R., D. Mitchell, J. Yi, and J. Newton. 1997. Examining the classification performance of neural network models and parametric methods for classifying distributions from the Weibull family. In *Decision Sciences Institute Southwest Region: Theory and Applications Proceedings, 28th Annual Conference, New Orleans, LA, March 11-15, 1997,* ed. L. R. Heinrichs, 34-36.

Petridis, V., S. Kazarlis, A. Papaikonomou, and A. Filelis. 1992. A hybrid genetic algorithm for training neural networks. In *Artificial Neural Networks, 2*, ed. I Aleksander and J. Taylor, 953-956. Amsterdam: Elsevier.

Poechmueller, W., M. Glesner, and H. Juergs. 1993. Is LVQ really good for classification? - An interesting alternative. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 402-407 New York: IEEE

Prechelt, L. 1996. A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice. *Neural Networks.* 9: 457-462.

Raghupathi, W., L. L. Schkade, and B. B. Raju. 1991. A neural network application for bankruptcy prediction. *Proceedings of the Twenty-fourth Annual Hawaii International Conference on Systems Science.* 4: 147-155.

Reibnegger, G., G. Weiss, G. Werner-Felmayer, G. Judmaier, and H. Wachter. 1991. Neural networks as a tool for utilizing laboratory information: Comparison with linear discriminant analysis and with classification and regression trees. *Proceedings of the National Academy of Sciences, Medical Sciences.* 88: 11426-11430.

Rencher, A. C. 1995. *Methods of multivariate analysis.* New York: Wiley.

Ribeiro, B. E. Costa, and A. Dourado. 1995. Lime kiln fault detection and diagnosis by neural networks. In *Artificial Neural Networks and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 112-115. Wien, Austria: Springer.

Richard, M. D., S. A. LeMay, and G. S. Taylor. 1995. A factor-analytic logit approach to truck driver turnover. *Journal of Business Logistics.* 16: 281-298.

Rochester, N., J. H. Holland, L. H. Haibt, and , W. L. Duda. 1956. Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory.* IT-2: 80-93.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review.* 65: 386-408.

________. 1962 *Principles of Neurodynamics.* New York: Spartan.

Rostar, R., and J. Olejar. 1995. Keystroke dynamics based user authentication using neural networks. In *Artificial Neural Networks and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 194-197. Wien, Austria: Springer.

Roth, M. W. 1990. Survey of neural network technology for automatic target recognition. *IEEE Transactions on Neural Networks.* 1: 28-43.

Rothenberg, M. J. 1992. Urban travel characteristics. In *Transportation Planning Handbook* ed. J. D. Edwards, 19-67. Englewood Cliffs, NJ: Prentice-Hall.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning representations by back-propagating errors. *Nature.* 323: 533-536.

Salchenberger, L. M., E. M. Cinar, and N. A. Lash. 1992. Neural networks: A new tool for predicting thrift failures. *Decision Sciences.* 23: 899-916.

Saxon, H., A. Bulsari, L. Karilainen, and K. Raipala. 1995. Application of neural networks for classification of temperature distribution patterns. In *Artificial Neural Networks and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 18-21. Wien, Austria. Springer.

Shah, C. 1997. Model selection in univariate time series forecasting using discriminant analysis. *International Journal of Forecasting.* 13: 489-500.

Sharda, R., and R. L. Wilson. 1996. Performance comparison issues in neural network experiments for classification problems. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 382-390. New York: IEEE.

Shepherd, A. J. 1997. *Second-Order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons*, Perspectives in Neural Computing, ed. J. G. Taylor. New York: Springer.

Sieger, D. B., and A. B. Badiru. 1993. An artificial neural network case study: Prediction versus classification in a manufacturing application. *Computers & Industrial Engineering*. 25: 381-384.

Shen, Z., M. Clarke, R. W. Jones, and T. Alberti. 1992. Neural network approach for detecting heart disease. In *Artificial Neural Networks, 2*, ed. I. Aleksander and J. Taylor, 1637-1640. Amsterdam: Elsevier.

Shunk, G. A. 1992. Urban transportation systems. In *Transportation Planning Handbook* ed. J. D. Edwards, 1-18. Englewood Cliffs, NJ: Prentice-Hall.

Simpson, P. K. 1996. Foundations of neural networks. In *Neural Networks Theory, Technology, and Applications*, ed. P. K. Simpson, 1-22. New York: IEEE.

Smith, C. A. B. 1946. Some examples of discrimination. *Annals of Eugenics*. 13: 272-282.

Smith, C. G., H. D. Hindman, and S. J. Havlovic. 1997. A discriminant analysis of employees choice in a multiunion representation election. *Canadian Journal of Administrative Sciences*. 14: 235-245.

Specht, D. F. 1990. Probabilistic neural networks. *Neural Networks*. 3: 109-113.

Stern, S. 1993. A disaggregate discrete choice model of transportation demand by elderly and disabled people in rural Virginia. *Transportation Research-A*. 27A: 315-327.

Subramanian, V., M. S. Hung, and M.Y. Hu. 1993. An experimental evaluation of neural networks for classification. *Computers and Operations Research*. 20: 769-782.

Swarup, K. S., and H. S. Chandrasekharaiah. 1992. Pattern classification using neural networks for diagnostic problem solving. In *Artificial Neural Networks, 2*, ed. I. Aleksander and J. Taylor, 1641-1644. Amsterdam: Elsevier.

Tam, K. Y., and M. Y. Kiang. 1992. Managerial applications of neural networks: The case of bank failure predictions. *Management Science*. 38: 926-947.

Tana, S. S., and H. C. Koh. 1997. A multi-layer perceptron model of credit scoring for assessing default risk in charge card applicants. *International Journal of Management*. 14: 250-255.

Toulson, D. L., J. F. Boyce, and C. Hinton. 1992. Data representation and generalization in an application of a feed-forward neural net. In *Neural Network Applications, Proceedings of the Second British Neural Network Society Meeting (NCM91), London, October 1991*, ed. J. G. Taylor, 27-40. London: Springer-Verlag.

Tubbs, J. D., W. A. Coberly, and D. M. Young. 1982. Linear dimension reduction and Bayes classification with unknown population parameters. *Pattern Recognition*. 15: 167-172.

U. S. Department of Transportation. Federal Transit Administration. 1986. *A self-instructing course in disaggregate mode choice modeling*. Washington, D. C.

Van Rooih, A. J. F., L. C. Jain, and R. P. Johnson. 1996. *Neural network training using genetic algorithms*. Series in Machine Perception & Artificial Intelligence, ed. H. Bunke and P. S. P. Wang, no. 26. Singapore: World Scientific.

Velasco, T., and M. R. Rowe. 1993. Back propagation artificial neural networks for the analysis of quality control charts. *Computers & Industrial Engineering*. 25: 397-400.

Venugopal, V., and W. Baets. 1994. Neural networks and statistical techniques in marketing research: A conceptual comparison. *Marketing Intelligence & Planning*. 12: 30-38.

von Neumann, J. 1958. *The computer and the brain*. New Haven: Yale University Press

Wang, S. 1995 The unpredictability of standard back propagation neural networks in classification applications. *Management Science*. 41: 555-559.

Werbos, P. J. 1974. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph. D. diss., Harvard University.

White, H. 1989 Learning in artificial neural networks: A statistical perspective. *Neural Computation*. 1: 425-464.

Whitley, D. 1995. Genetic algorithms and neural networks. In *Genetic Algorithms in Engineering and Computer Science*, ed. G. Winter, J. Periaux, M. Galan, and P. Cuesta, 203-216. New York: John Wiley.

Widrow, B., and M. E. Hoff. 1960. Adaptive switching circuits. *IRE WESCON, Convention Record.* Part 4: 96-104.

Widrow, B., and M. A. Lehr. 1990. 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proceedings of the IEEE.* 78: 1415-1442.

Widrow, B., D. E. Rumelhart, and M. A. Lehr. 1994. Neural networks: Applications in industry, business, and science.

Williamson, A. G., and P. Munson. 1995. Radial basis function neural networks in credit application vetting systems. In *Artificial Neural Networks and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 301-304. Wien, Austria: Springer.

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. *Statistical principles in experimental design, 3rd ed.* New York: McGraw-Hill.

Yi, J., D. Mitchell, and V. R. Prybutok. 1996. Comparison of neural network model with conventional statistical models for prediction of Houston's maximum ozone levels. In *Decision Sciences Institute Southwest Region: Theory and Applications Proceedings, 27th Annual Conference, San Antonio, TX, March 6-9, 1996*, ed. R. Pavur and S. Yager, 57-59.

Yoon, Y., and G. Swales. 1991. Predicting stock price performance: A neural network approach. *Proceedings of the Twenty-fourth Annual Hawaii International Conference on Systems Science.* 4: 156-162.

Yu, D., D. N. Shields, and S. Daley. 1995. Application of radial basis function networks to fault diagnosis for a hydraulic system. In *Artificial Neural Networks and Genetic Algorithms, Proceedings of the International Conference in Ales, France, 1995*, ed. D. W. Pearson, N. C. Steele, and R. F. Albrecht, 100-103. Wien, Austria: Springer.

Zhao, Q., and Z. Bao. 1996. Radar target recognition using a radial basis function neural network. *Neural Networks.* 9: 709-720.