Multi-layered perceptrons for on-line lot sizing : extended abstracts

MULTI-LAYERED PERCEPTRONS FOR ON-LINE LOT SIZING (EXTENDED ABSTRACT)

H.P. Stehouwer[1], *E.H.L. Aarts*[1,2], *and J. Wessels*[1,3]

[1] *Eindhoven University of Technology, Department of Mathematics and Computing Science*
P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands
E-mail: peterst@win.tue.nl
[2] *Philips Research Laboratories*
P.O. Box 80.000, NL-5600 JA Eindhoven, The Netherlands
E-mail: aarts@prl.philips.nl
[3] *International Institute for Applied Systems Analysis*
A-2361 Laxenburg, Austria
E-mail: wessels@iiasa.at

1 Introduction

Production can be viewed as a transformation process in which *materials* are transformed into end products. These transformations require *resources* such as manpower and machines. Lot sizing is concerned with the determination of production quantities (the lot sizes) in order to satisfy the demand for end product over time such that production resources are used efficiently.

A lot of research on lot sizing has been focussed on *deterministic* analysis. For an overview of this literature the reader is referred to [Bahl, Ritzman & Gupta, 1987; Aggarwal & Park, 1993]. Deterministic models are based on the assumption that all problem data are known in advance. However, this assumption is often not justified in practice, where many forms of uncertainty effect the production process. Especially uncertainty in demand can be of great influence on the efficiency of the lot sizing.

In practice one often faces the situation in which demand information comes in gradually. In such situations it is only for a certain small time horizon into the future that demand can be considered deterministic. The problem is how to determine the lot sizes in such an *on-line* planning situation such that production resources are used efficiently. We distinguish between the following three approaches for this problem.

Using a *myopic approach*, at any decision moment, the world is supposed to stop beyond the encountered deterministic time horizon and an optimization or approximation algorithm for the corresponding deterministic lot sizing problem is applied.

In an *explicit modeling approach*, the demand process is explicitly modeled by assuming that demand is the realization of a random process, possibly with unknown parameter values. These unknown parameters may characterize for example the noise part of the demand process or some systematic trend and can be estimated from historical demand data. With such a modelling of the demand process, lot sizing problems can be formulated as

Markov decision problems [Tijms, 1986] and solved as such. However, analysis in this area is usually already technically complicated for relatively simple models of the underlying random processes. Promising is the application of fuzzy optimization techniques, where a possibilistic in stead of a probabilistic analysis is used [Lee, Kramer & Hwang, 1991]. In practice, usually the problem is decomposed in an estimation part and an optimization part. In the estimation part the thus obtained model of the demand process is used to forecast future demand values. In the optimization part these future demand values are considered as “real” demand and are incorporated in a lot sizing procedure.

Using a *black-box approach*, one accepts that one is not able to model the underlying demand process and one takes a parametrized black-box and tries to fit the parameters in such a way that the black-box shows a sensible input-output behavior on at least a representative set of examples. These examples represent situations from the past in which lot sizing decisions had to be taken. The difference is that afterwards it is often quite easy to determine what would have been the optimal decisions.

Advantages of using a myopic approach are the absence of history requirements and its straightforward implementation. Disadvantages are a low performance and a high system nervousness, especially if variances in demand are large; see [Blackburn & Millen, 1980]. Both stochastic and fuzzy models assume that the nature of the underlying random processes are well understood and are not subject to change. In practice, however, this is often not the case. In such cases the black-box approach might be useful.

Artificial neural networks [Arbib, 1987] and more specifically multi-layered perceptrons (MLPs) are interesting candidates for being used as black-boxes in production planning; for their generalization and interpolation abilities, for their pattern recognition skills [Pao, 1989], for their classification capabilities [Zwietering, 1994] and for their ability to adapt to changing circumstances. In [Zwietering, Van Kraaij, Aarts & Wessels, 1991] it was shown that a properly designed and trained MLP outperforms traditional algorithms for the rolling horizon version of the Wagner-Whitin problem [Wagner & Whitin, 1958]. In this paper we consider an on-line lot sizing problem with overtime. We develop a two-stage decision procedure for this problem. In the first stage an MLP classifies the decision situation. It is in this stage that uncertainties are taken into account. The outcome of the first stage is used as input for the second stage, in which a detailed production plan is calculated. The proposed approach combines the classification and pattern recognition abilities of MLPs with traditional deterministic analysis.

The remainder of this paper is organized as follows. In Section 2 we give a brief introduction in MLPs and supervised learning. The on-line lot sizing problem is formulated in Section 3. Based on results for the deterministic finite horizon problem, which is analyzed in Section 4, in Section 5 we derive a two-stage strategy for the on-line lot sizing problem. Finally in Section 6 we discuss some results.

2 Multi-layered perceptrons

In general, a neural network consists of a network of elementary nodes that are linked through weighted connections. The nodes represent computational units, which are capable of performing a simple computation that consists of a summation of the weighted inputs of the node, followed by the addition of a constant called the threshold or bias, and the application of a non-linear response function. The result of the computation of a unit constitutes the output of the corresponding node. Subsequently, the output of a node is used as an input for the nodes to which it is linked through an outgoing connection.

In an MLP the nodes are arranged in layers, and the connections are not allowed to cross a layer, i.e, there are connections between the inputs of the network and the nodes in the first layer and between subsequent layers only.

When applying MLPs for a certain task, besides choosing the number of layers and the number of units per layer, one has to choose the weights such that the network performs the task accurately. These are the parameters of the black-box to be fitted, which in general cannot be determined beforehand. The common names for fitting parameters in the context of neural networks are *learning* or *training*, and when the learning is done on the basis of direct comparison of the output of the network with known correct answers, one speaks of *supervised learning*. In a supervised learning problem, one is given an MLP and a set $\mathcal{S} = \{(x_1, t_1), (x_2, t_2), \ldots, (x_N, t_N)\}$ of examples. Each example consists of an input vector $\mathbf{x_k}$ and a corresponding desired output vector $\mathbf{t_k}$. The problem is to find weights $\mathbf{w}$ such that the difference between the output vector of the MLP on input of a particular $\mathbf{x_k}$ and the target vector $\mathbf{t_k}$ is minimized for the entire set, measured by $E(\mathbf{w}) = \sum_{k=1}^{N} \parallel \mathbf{g}(\mathbf{w}; \mathbf{x_k}) - \mathbf{t_k} \parallel^2$, where $\mathbf{w}$ denotes the weight vector and $\mathbf{g}(\mathbf{w}; \mathbf{x_k})$ denotes the output vector of the MLP with weights $\mathbf{w}$ after processing input vector $\mathbf{x_k}$. This makes the supervised learning problem equivalent to the task of searching weight space for a minimum of $E(\mathbf{w})$. The best-known method for descending the $E(\mathbf{w})$ surface is the *backpropagation algorithm* [Rumelhart, McClelland & Williams, 1986].

A supervised learning problem is called a *classification problem* if the desired output vector $\mathbf{t_k}$ is one of a finite number of possibilities (classes). For any classification problem MLPs offer a possible solution, as an alternative for conventional classification techniques; cf. [Huang & Lippmann, 1988].

Although the ability of an MLP to memorize and recall data in the abovementioned way is impressive, it becomes truly interesting when the network could extend this behavior to similar data it has never seen. This is called *generalization*. MLPs have shown to be able to generalize in a wide variety of tasks.

3 Problem formulation

Consider the situation in which production has to be planned for a product for which demand occurs during discrete time periods labeled $t = 1, 2, \ldots$. Demand for a certain period becomes known $n \geq 0$ periods ahead, so at any moment in time demand is known for n consecutive periods into the future. Demand occurring during a certain period must be satisfied by production during that period or by production during an earlier period. In a period there is a limited regular time production capacity of C units product. At extra production costs it is possible to produce during overtime. Let d_t, X_t, and I_t denote the demand in period t, the production in period t, and the inventory position at the end of period t, respectively. The cost function related to production is denoted by $P(\cdot)$. The cost function related to carrying inventory is denoted by $H(\cdot)$. Without loss of generality we assume $I_0 = 0$. Then the problem can be formulated as choosing $X_1, X_2, \ldots,$ to minimize the average costs per period

$$\lim_{\tau \to \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} [P(X_t) + H(I_t)], \qquad (1)$$

subject to

$$\begin{aligned} I_t = I_{t-1} + X_t - d_t, \qquad & t = 1, 2, \ldots, \\ X_t \geq 0, I_t \geq 0, \qquad & t = 1, 2, \ldots. \end{aligned}$$

In this paper we assume that the production cost function $P(\cdot)$ is piecewise linear and the inventory cost function $H(\cdot)$ is linear as follows

$$P(X) = \begin{cases} 0 & \text{if } X = 0, \\ S + pX & \text{if } 0 < X \leq C, \\ S + pX + r(X - C) & \text{if } X > C, \end{cases}$$

$$H(I) = hI, \text{ for all } I \geq 0,$$

where $p \geq 0$ denotes the regular time production cost per unit product, $h \geq 0$ denotes the holding cost per unit product per period, $S \geq 0$ denotes the setup cost, and $r \geq 0$ denotes the difference in cost per unit product between regular time and overtime production.

Since the planning horizon is infinite, the size of an instance of this problem is unbounded. In general the limit (1) does not exist, and an optimal production policy is not obtainable. We will refer to this problem as the *on-line infinite horizon problem*.

In the introduction we mentioned a number of approaches for dealing with on-line lot sizing situations. One of the approaches was to combine an optimization algorithm for the finite horizon problem in combination with a *rolling horizon procedure*. This means that a so called *rolling plan* is formed by solving the deterministic n period problem and implementing only the first period's decision. One period later, the horizon is updated and the process repeated. This strategy lies at the basis of the approach we propose, and it turns out that the analysis of the deterministic finite horizon problem plays an important role. Therefore, in the next section, we analyze the deterministic finite horizon problem. Without loss of generality we assume that the planning horizon covers periods $1, 2, \ldots, n$ and that $I_0 = I_n = 0$.

4 The deterministic finite horizon problem

The deterministic n period problem can be formulated as to choose $\{X_t\}_{t=1}^n$, to minimize

$$\sum_{t=1}^{n} [\ P(X_t) + H(I_t)\ ], \tag{2}$$

subject to

$$I_0 = I_n = 0, \tag{3}$$

$$I_t = I_{t-1} + X_t - d_t, \qquad t = 1, 2, \ldots, n, \tag{4}$$

$$X_t \geq 0,\ I_t \geq 0, \qquad t = 1, 2, \ldots, n. \tag{5}$$

Special cases of this model include the case $r = 0$ in which the model becomes equivalent to the Wagner-Within model [Wagner & Whitin, 1958] and the case $r = \infty$ in which there is no overtime production possible [Florian & Klein, 1971]. A similar model was analyzed in [Dixon, 1980], in which the amount of overtime production and inventory per period was bounded.

To facilitate the exposition of properties of optimal solutions, let us define the following notions. A solution is called a *production plan* and consists of a sequence $X = (X_1, \ldots, X_n)$. In a production plan X, period t is called a *production period* if $X_t > 0$, and whenever $I_t = 0$, we say X has an inventory regeneration point at the end of period t. Given a production plan X, a subsequence $(X_{u+1}, X_{u+2}, \ldots, X_v)$ of X, with $0 \leq u < v \leq n$, is called a *subplan* of X, if $I_u = I_v = 0$ and $I_t > 0$, for all $t = u+1, u+2, \ldots, v-1$. The following lemma represents the so called *inventory decomposition property*, which is easily proven by contradiction.

Lemma 4.1. *Consider an optimal feasible production plan X. Suppose that X has an inventory regeneration point at the end of period t for some t, $0 < t < T$. Then this optimal solution could also have been found by independently finding solutions for the first t periods and the last $T - t$ periods, given I_t.* □

From this together with $I_0 = I_n = 0$ it follows that any optimal production plan can be decomposed into one or more subplans. Given all of the regeneration points, one could determine the optimal production plan by solving the subproblems between each pair of consecutive regeneration points. Unfortunately, the optimal regeneration points are not known a priori. However, if the solutions for the subproblems are known for all possible pairs of regeneration points, then one can select the best combination of regeneration points.

Let $g(u, v)$ denote the cost of a minimum cost production plan for periods $u+1, \ldots, v$, given $I_u = I_v = 0$ and $I_t > 0$ for $t = u+1, \ldots, v-1$, i.e., an optimal subplan. Let the regeneration points be represented in a network as nodes. Then $g(u, v)$ represents the cost of traversing the arc from node u to node v. A production plan is a path from node 0 to node n, since $I_0 = I_n = 0$ holds. Since backlogging is not permitted, the network is acyclic. Hence, the problem can be formulated as finding the shortest path in an acyclic network, which is easily solved by dynamic programming and requires $O(n^2)$ computations.

In the now following, a number of properties of optimal production plans are given, which facilitate an efficient algorithm for finding optimal subplans. For a detailed handling of the problem and the proofs the reader is referred to [Stehouwer, Aarts & Wessels, 1995].

Theorem 4.1. *There exists an optimal an optimal production plan X having the properties that each subplan $(X_{u+1}, X_{u+2}, \ldots, X_v)$ of X*

(i) *contains not more than one production period t, with $X_t \neq C$, and*

(ii) $X_i \leq X_j$*, for all production periods i, j, $u+1 \leq i < j \leq v$.* □

We introduce the notion *cumulative demand axis* as described in [Chung & Lin, 1988]. Instead of giving each period an equal length on a time axis, each period is represented by an interval of length proportional to the demand in that period, and demand is spread uniformly over a period. The origin is used to indicate the beginning of period 1. We then mark the points $B_1 = 0$ and $B_t = \sum_{i=1}^{t-1} d_i$, for $t = 2, \ldots, n+1$. Each point B_t refers to the end of period $t-1$ and the beginning of period t, hence the interval from B_t to B_{t+1} represents the demand in period t. We define the notation $(i_1, i_2, \ldots, i_k)$, with $u+1 \leq i_k < \ldots i_2 < i_1 \leq v$ to denote a subplan $(X_{u+1}, X_{u+2}, \ldots, X_v)$ with production periods $i_1, i_2, \ldots, i_k$. Using the cumulative demand axis, it follows that production in period i_1 is used to meet the demand from $B_{v+1} - X_{i_1}$ to B_{v+1}. Production in period i_2 is used to meet the demand from $B_{v+1} - X_{i_1} - X_{i_2}$ to $B_{v+1} - X_{i_1}$, and so on. In a subplan the production in each period can only be used to meet present or future demand and inventory must be positive, therefore we shall require $B_{v+1} - X_{i_1} > B_{i_1}$, $B_{v+1} - X_{i_1} - X_{i_2} > B_{i_2}$ and so on. Clearly any subplan can be represented as such.

Theorem 4.2. *Consider an optimal production plan X. Let $(X_{u+1}, X_{u+2}, \ldots, X_v)$ be a subplan of X, with production periods $(i_1, i_2, \ldots, i_k)$, for some $k \in \{1, 2, \ldots, v-u\}$. Define $i_0 := v+1$. Let d_{uv} denote the cumulative demand for periods $u+1, \ldots, v$. Then the* *following holds*

$(i)\quad i_k = u+1$

$(ii)\quad d_{uv} > (k-1)C$

$$(iii)\quad \begin{cases} X_{i_k} = d_{uv} & \text{if } k=1 \\ X_{i_k} = X_{i_{k-1}} \cdots = X_{i_2} = C \ \wedge\ X_{i_1} = d_{uv} - (k-1)C & \text{if } k>1 \ \wedge\ d_{uv} \geq kC \\ X_{i_k} = d_{uv} - (k-1)C \ \wedge\ X_{i_{k-1}} = X_{i_{k-1}} \cdots = X_{i_1} = C & \text{otherwise} \end{cases}$$

$(iv)\quad i_n = \max\{j \mid u+1 < j < i_{n-1} \wedge B_{v+1} - \sum_{m=1}^{n} X_{i_m} > B_j\}, n = 1, 2, \ldots, k-1.$

□

Theorem 4.2 implies that for determining an optimal subplan $(X_{u+1}, X_{u+2}, \ldots, X_v)$ only $(v-u)$ values of k have to be examined. For each value of k the corresponding subplan can be determined in at most $(v-u)$ steps. Therefore, an optimal subplan can be found in $O((v-u)^2)$ steps. Since there are $O(n^2)$ arcs, an optimal production plan can be found in $O(n^4)$ steps. In [Stehouwer, Aarts & Wessels, 1995] it is shown that only subplans of limited length have to be examined, which in many cases reduces the number of computations drastically.

5 The on-line infinite horizon problem

In any sensible planning strategy for the on-line infinite horizon problem, decisions are based on at least the available future demand information. When applying a pure myopic approach this is the only information that is used. An MLP constructed for this problem will have at least n input units, one for each known future demand, and one output unit, representing the first period's lot size decision. Such an MLP can be trained by supervised learning with examples of past decision situations. Examples could be generated by taking past demand sequences of length n and solving the corresponding finite horizon problems. Training an MLP with such examples is equivalent to learning the MLP a standard rolling horizon procedure, i.e., applying a pure myopic approach. However, there is no need to be myopic when generating learning examples. An Example can also be generated by taking a demand sequence of length $m \geq n$ from demand history and by solving the corresponding m period finite horizon problem. The input part of the example are the first n demands of this demand sequence and the desired output part is the first period's production lot. In fact an MLP trained with such examples predicts the optimal production lot for the first period for a m period problem give the first n demands.

We prefer the use of MLPs as classifiers, not only for their classifying abilities, but also for their analyzability. To identify a classification problem here, we take a look at the standard rolling horizon procedure. Remark that the first period's production lot is obtained by solving a deterministic n period problem. From the results in Section 4, we know that the size of this lot is determined by the position of the first regeneration point in the optimal production plan. Given this regeneration point, the size of this lot can easily be calculated.

This gives rise to an approach in which the lot size is determined in two stages. In the first stage, the optimal first regeneration point $r \in \{1, 2, \ldots, n\}$ is determined by the MLP. In the second stage, the outcome of the first stage is used to calculate the first period's production lot by solving $g(0, r)$ and finding an optimal subplan. This decomposition is justified by the inventory decomposition property; see Lemma 4.1. The first stage can be reformulated as a classification problem (Ω, L, Γ) as follows

$$\Omega = \{d \in \mathbb{R}^n \mid d_t \geq 0,\ t = 1, 2, \ldots, n\},$$
$$\Gamma = \{\Omega_l \mid l \in L\},$$
$$L = \{1, 2, \ldots, n\},$$

where Ω denotes a set of objects that must be classified, L denotes a set of labels, and Γ denotes a collection of subsets of Ω, one for each label. For a given demand vector $(d_1, \ldots, d_n) \in \Omega$ the problem is to find a label $l \in L$ such that $d \in \Omega_l$. For $l = 1, 2, \ldots, n$, Ω_l can be characterized as

$$\Omega_l = \{d \in \Omega \mid g(0, l) + f(l, n) \leq g(0, r) + f(r, n), \text{ for all } r = 1, 2, \ldots, n\},$$

and solving (Ω, L, Γ) becomes equivalent to finding l such that the costs $g(0, l) + f(l, n)$ are optimal, i.e., to determine the first regeneration point in the optimal solution. Here $f(u, v)$ denotes the cost of an optimal production plan for the periods $u+1, u+2, \ldots, v$, provided that $I_u = I_v = 0$.

When determining first regeneration points for training examples, again there is no need to be myopic and demand sequences of length $m \geq n$ can be taken. It is possible that such a first regeneration point lies beyond period n. In the situations we considered this was not the case. How such situations can be handled is described in [Stehouwer, Aarts & Wessels, 1995].

In the next section we present some results of MLPs trained with such examples. These network function as follows. There are n inputs, one for each known demand. There are n outputs, one for each class label in L. An example exists of an input part, the first n demands, and an output part. This output part is vector of length n with $n-1$ zeros and 1 one. On input of an example the MLP classifies this example as that class label of which the output of the corresponding output unit has maximum response.

6 Numerical results and discussion

To validate the proposed approach, a number of representative examples of planning situations are considered. Therefore, we use the parameter settings which are derived in [Dixon, Elder, Rand & Silver, 1983]. In this paper we report on the case $r = 0.1$, $S = 60$, and $C = 200$, with a horizon of length $n = 6$. A complete description of the numerical results can be found in [Stehouwer, Aarts & Wessels, 1995]. We consider two types of demand patterns. In all cases the average demand is set at 200 units per period.

1. Level with noise - Demands for individual periods are generated independently from an uniform distribution with a mean of 200 units and standard deviations of 10 and 50 units.
2. Seasonal with noise - A sine with an amplitude of 100 units and a cycle of $n = 6$ periods is generated. To this the average level of 200 units as described in the level case is added.

For both patterns we generate three sets of 2895 examples, one for training (learning set), one for determining the best MLP during training (test set), and one for validation of the best MLP (validation set). Each example is obtained by taking a demand sequence of length $m = 100$, and solving the corresponding finite horizon problem as explained in the previous section. This value of m is determined empirically. Adding more demands information seems not to have any impact on the first regeneration points.

We train two-layered MLPs with $n = 6$ input units, $n = 6$ output units, and 3, 6, 12, 15, and 18 units in the hidden layer with standard backpropagation. We use a learning rate of 0.1 and a momentum term of 0.9. To average out the dependency on the initial weights, for every number of hidden units, we average over 10 MLPs trained with different randomly chosen initial weights.

After and during training of an MLP, its performance on the different sets is measured in two different manners. The first manner is to measure the percentage correct classifica-

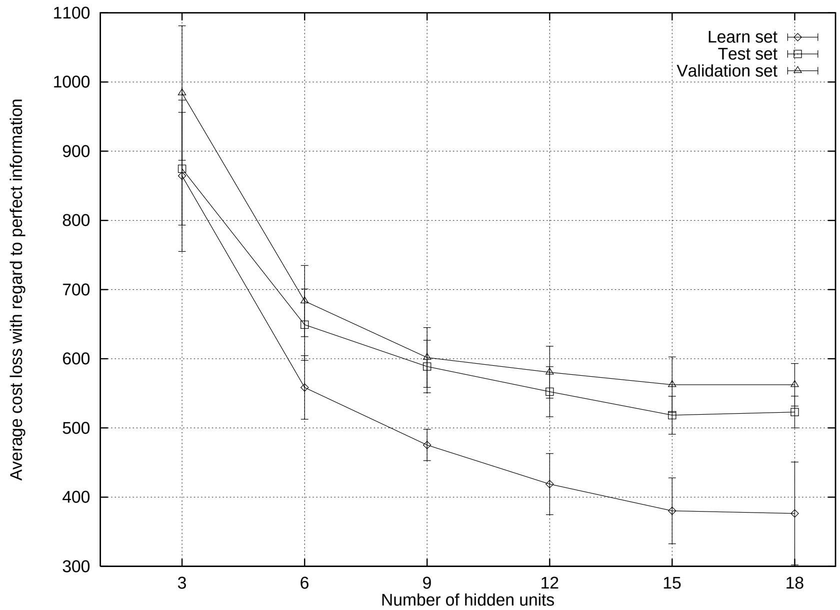

Figure 1: Performance on the sets averaged over 10 trained MLPs for demand pattern 2 with an uniform noise level with a mean of 200 and a standard deviation of 50 units.

tion on these sets. However, this is not a clear indiction of the actual performance in terms of production and inventory costs. In the second manner, for each example we compare the decision made by the MLP with the decision that would have been optimal with respect to the m period horizon of which the example originated in case of perfect demand information. Suppose for a particular demand sequence $d_1, \ldots, d_m$, on presentation of the first n demands $d_1, \ldots, d_n$ to the network, the MLP classifies it as having class label r. This means that, in the second stage, we calculate our production plan by solving $g(0, r)$. Assume that from period r through m an optimal plan is calculated by calculating $f(r, m)$. In case r is the first regeneration point of the m period problem this corresponds with an optimal decision. In case the classification was not correct and r is not the first regeneration point, a loss of $g(0, r) + f(r, m) - f(0, m)$ is incurred. However, there is no need to stick to the subplan structure here. Since $f(0, r) \leq g(0, r)$ by definition, it is better to calculate our production plan by solving $f(0, r)$. The corresponding loss then is equal to $f(0, r) + f(r, m) - f(0, m)$. The total performance of an MLP on a set of examples is obtained by summing the losses over the entire set.

It turns out that the MLPs are able to perform more than 90 percent correct classification on all sets for all tested patterns. A typical effect of increasing the number of hidden units on the performance on the different sets is shown in Figure 1. Remark that there is an optimal number of hidden units with respect to the networks generalization ability measured on the validation set. For this case this is 15 hidden units. Adding more units does only improve the results on the learning set. This effect is called *over-fitting*.

We compare our results with an ordinary rolling horizon approach. The outcome of this comparison is tabulated in Table 1. From this table it is clear that our approach outperforms this pure myopic approach on both percentage correct classification and cost losses for all patterns. The MLP approach shows a robust behavior with respect to the demand structure. Noteworthy is the excellent performance compared with the myopic approach in case of patterns in which the amount of noise is small compared with the deterministic part of the pattern. It is in these cases that the pattern recognition abilities of MLPs are exhibited. A drawback of the use of MLPs is the large amount of time required for training.

Demand		Approach				
		MLP			Myopic	
Pattern	Standard Deviation Noise	Average % Correct	Average Loss	Standard Deviation Loss	% Correct	Loss
1	10	97.9	72.2	0.0	97.9	72.2
2	10	92.5	361.3	9.3	78.2	2326.3
1	50	92.1	514.7	22.2	89.0	1346.0
2	50	92.1	518.4	27.5	91.9	870.4

Table 1: Results for the different demand patterns

Acknowledgement

The first author is supported by the National Science Foundation of the Netherlands under contract 611-304-032.

References

AGGARWAL, A. AND J.K. PARK [1993], Improved algorithms for economic lot size problems, *Operations Research* **41**, 549–571.

ARBIB, M.A. [1987], *Brains, Machines, and Mathematics*, Springer-Verlag, Berlin.

BAHL, H.C., L.P. RITZMAN, AND J.N.D. GUPTA [1987], Determining lot sizes and resource requirements: a review, *Operations Research* **35**, 329–345.

BLACKBURN, J.D. AND R.A. MILLEN [1980], Heuristic lot-sizing performance in a rolling-schedule environment, *Decision Sciences* **11**, 691–701.

CHUNG, C.-S. AND C.-H.M. LIN [1988], An $O(T^2)$ algorithm for the NI/G/NI/ND capacitated lot size problem, *Management Science* **34**, 420–426.

DIXON, P.S. [1980], *Single-item lot-sizing with limited regular and overtime capacity*, Working paper, Department of Finance and Management Science, Saint Mary's University, Halifax, Canada.

DIXON, P.S., M.D. ELDER, G.K. RAND, AND E.A. SILVER [1983], A heuristic algorithm for determining lot sizes of an item subject to regular and overtime production capacities, *Journal of Operations Management* **3**, 121–130.

FLORIAN, M. AND M. KLEIN [1971], Deterministic production planning with concave costs and capacity constraints, *Management Science* **18**, 12–20.

HUANG, W.Y. AND R.P. LIPPMANN [1988], Neural nets and traditional classifiers, in: D.Z. Anderson (ed.), *Neural Information Processing Systems*, New York, American Institute of Physics, 387–396.

LEE, Y.Y., B.A. KRAMER, AND C.L. HWANG [1991], A comparative study of three lot-sizing methods for the case of fuzzy demand, *International Journal of Operations & Production Management* **11**, 72–80.

PAO, Y.H. [1989], *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley.

RUMELHART, D.E., J.L. MCCLELLAND, AND R.J. WILLIAMS [1986], Learning internal representations by error propagation, in: D.E. Rumelhart and J.L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, Cambridge (MA), 318–362.

STEHOUWER, H.P., E.H.L. AARTS, AND J. WESSELS [1995], *Multi-Layered Perceptrons for on-line lot sizing*, Technical report, Eindhoven University of Technology, To appear as COSOR Memorandum.

TIJMS, H.C. [1986], *Stochastic Modelling and Analysis: A Computational Approach*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

WAGNER, H.M. AND T. WHITIN [1958], Dynamic version of the economic lot size model, *Management Science* **5**, 89–96.

ZWIETERING, P.J. [1994], *The Complexity of Multi-Layered Perceptrons*, Ph.D. thesis, Eindhoven University of Technology.

ZWIETERING, P.J., M.J.A.L. VAN KRAAIJ, E.H.L. AARTS, AND J. WESSELS [1991], Neural networks and production planning, *Proceedings of the Fourth International Conference on Neural Networks and their Applications*, 529–542.