We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 3

Multivariate-Stepwise Gaussian Classifier (MSGC): A New Classification Algorithm Tested Over Real Disease Data Sets

Alexandre Serra Barreto

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74703

Abstract

In data mining, classification is the process of assigning one amongst previously known classes to a new observation. Mathematical algorithms are intensively used for classification. In these, a generalization is inferred from the data, so as to classify new cases, or individuals. The algorithm may misclassify an individual if the inference machine is not able to sufficiently discriminate it. Therefore, it is necessary to go further into the analysis of the information provided by the individual, until it can be sufficiently identified as belonging to a class. This chapter developed this idea for the improvement of a certain class of classifiers, using medical data sets to validate the new algorithm proposed here: The Multivariate-Stepwise Gaussian Classifier (MSGC). The results showed that MSGC is at least as competitive as the Gaussian Maximum Likelihood Classifier. MSGC attained the greatest accuracy rate in two of the data sets, and obtained identical results in the two remaining data sets. Concerning medical applications, once a classification method has been successfully validated considering a particular scope of data, the recommendable would be its use for the best diagnosis. Meanwhile, other algorithms could be tested until they proved to be effective enough to be put into practice.

Keywords: data mining, classification, algorithm, medical diagnosis and prediction

1. Introduction

Mankind has performed classification since remote years, as a part of daily life and survival. With human evolution, our motivation to classify has become more complex and wide, comprehending classification in a wide variety of fields like engineering, management, banking, marketing, psychology and medical diagnosis and prediction.

In the context of data mining, classification can be understood as the process of assigning to a new observation (sample) one among a set of previously known classes. In fact, the rapid increase in computational processing capacity, coupled with the low cost of storage, has contributed to the greater use of supervised or nonsupervised mathematical algorithms for computational classification. In these, in the learning phase, certain kind of generalization is inferred from the data, so that new cases, or individuals, can be classified by the inference machine.

It should be mentioned that in the medical field, there are several examples of researches applying successfully computational classification as an aid to the medical diagnosis. It can be referred, for instance, the research in [1], which apply a multivariate statistical analysis to explore the Dermatology Data Set (available in the UCI data repository, [2]) and construct a classifier, based only on the 12 clinical attributes, as an aid for the first medical consultation and diagnosis of erythemato-squamous dermatological diseases. The research results provide enhanced knowledge that can help to enrich dermatological diagnoses made by doctors. Also, the classifier developed using the linear discriminant analysis (LDA) obtains a high mean accuracy rate in relation to the six diseases (83.73% correct classifications). This rate means that patients have a good chance of being treated adequately, while biopsies may also be solicited to confirm diagnosis. A classification algorithm developed in [3] was tested over the Dermatology Data Set. This study reported mean accuracy rates (96.2 and 99.2% for a modified version of the algorithm). Note that it utilized all 34 features in the data set (clinical + histological attributes), which can certainly inform further the classifier, since it works knowing the biopsy results. In [4], an analysis is outlined attempting to classify the Dermatology Data Set by decision trees and employing all 34 features in the data set. The authors reported a 5.5 +/−1.46 error rate. A modified decision tree based on a genetic algorithm for attribute selection achieved a 4.2 +/−0.96 error rate. In [5], a classification algorithm is demonstrated, based on genetic algorithms that discovered comprehensible IF-THEN rules. The algorithm was submitted to all 34 features in the Dermatology Data Set and the result was 95% accuracy rate for classifications. By visiting the UCI data repository website, many other studies focusing several medical data sets are listed and can be accessed by the reader.

However, occasionally such generalization may not correctly classify an individual if the inference machine is not able to sufficiently discriminate it among the possible classes. Therefore, it is necessary to go more deeply into the analysis of the information provided by the individual being classified, until it can be sufficiently identified as belonging to a class.

This pursuing, moreover, may be analogous to the efforts made by physicians while performing their crucial diagnostics. In fact, medical theory and practice well acknowledge a basic foundation in medicine, that no two individuals are alike, either in health or illness. For this reason, more and more medical guidelines pursue this maxim, the individualities being considered in the midst of large numbers, examples being the programs of family physicians, homeopathy, psychoanalysis, encouragement of anamnesis rather than light and machine consultations and recent considerations involving slow medicine. Not to lengthen this subject too much, reference is made to the works ([6] p. 5–6, [7] p. 11–12], [8], and [9] p. 3). It could still be possible refer to a series of other initiatives that denote the search for health in its individual fullness, but what is important is that common sense says that such a foundation should also inform statistical methods and artificial intelligence applied to the classification of individuals.

http://dx.doi.org/10.5772/intechopen.74703

In this context, this chapter seeks to develop this idea for the improvement of a certain class of well-known classifiers; for this purpose, it uses real medical data sets for the validation of the algorithm proposed here.

2. Classifiers based on the assumption that the form for the underlying density function is known

In parametric classification techniques, we learn from data under the assumption that the form for the underlying density function is known. The most common procedure is to consider the normal distribution, as is the case of Gaussian Maximum Likelihood Classifier (GMLC). Suppose there are c distinct classes, given a sample vector $X^T = (x_1, x_2, \ldots, x_p)$ depicting p measurements made on the sample from p attributes, GMLC will assign to X the class h (h = 1,…, c) having the highest likelihood among the classes. GMLC assumes that the data follows the multivariate normal density function:

$$f(X|\mu_h, \Sigma_h) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_h|}} \exp\left[-1/2(X-\mu_h)^T \Sigma_h^{-1} (X-\mu_h)\right]. \quad (1)$$

In this equation, μ_h is the mean vector of class h, Σ_h is the covariance matrix for class h and $|\Sigma_h|$ is the determinant of Σ_h. Usually, these parameters are not known and must be estimated from training samples. The sample mean is typically the estimate for the density mean, and the covariance matrix is usually estimated via the sample covariance matrix or the maximum likelihood covariance matrix estimate. The sample mean and the maximum likelihood covariance matrix estimates maximize the joint likelihood of training samples, which are assumed to be statistically independent [10].

The depicted above is mostly the case of some well-known classifiers like linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and regularized discriminant analysis (RDA), which are trustworthy classifiers based on GMLC computations that reach good results in several data situations. A basic difference between these three classifiers is that in the case of LDA, it is assumed that each class h comes from a normal distribution with a class-specific mean vector and a common global covariance matrix. On the other hand, QDA provides a model that assumes as many covariance matrices (Σ_h) as there are classes (h). RDA provides a kind of mix of them by means of tuning parameters (λ, γ), which provides an optimal mix of sample covariance matrix, global covariance matrix and the identity matrix, for instance, if ($\lambda = 0$ and $\gamma = 0$) RDA will represent QDA, and if ($\lambda = 1$ and $\gamma = 0$) RDA will represent LDA. It is important to note that among them there is no method considered better. For instance, in [11], it is possible to see that the performance of a classification method varies according to the database considered. The reader could refer to [12], ([13] p. 331–335) and [14] for the accessing of LDA, QDA and RDA foundations.

However, for sure, these aforementioned methods have their shortcomings. Barreto [11] lists the more commonly identified shortcomings in the field literature, such as the fact that the mean and covariance estimates are optimal only asymptotically and can produce lower

classification accuracy when the training sample is small, actually, unless many more than $p+1$ samples are available, the true covariance matrix is poorly estimated. Also, the assumption of the knowledge about the form for the underlying density function may be suspicious in most applications. Furthermore, the method involves the inversion of Σ_h estimate and in some cases, this matrix can be ill-conditioned or even singular, making matrix inversion unfeasible. In spite of the research proposing improvements, specifically concerning the covariance matrix estimation, of which RDA is a legitimate representative, these approaches remain operating under the key assumption that the form for the underlying density function is known.

Beyond these problems, this chapter wants to discuss that these methods maximize $p(X|h)\times p(h)$ to predict the class for the vector of data X, that is, $p(h|X)$. But $p(X|h)$ is calculated on the basis of the density in Eq. (1), which involves the calculation of the well-known Mahalanobis distance from the multivariate mean $[(X\text{-}\mu_h)^T\Sigma_h^{-1}(X\text{-}\mu_h)]$, which is a positive measure. Formally, the Mahalanobis distance represents a dimensionless multivariate measure of the distance between the multivariate vector X**,** with p dimensions, and the class mean μ_h, that also has the same p characters. The smaller the distance with respect to a specific class mean, say μ_c, the more the probability that X belongs to class c. Therefore, by the inspection of Eq. (1), it is easy to see that this mathematical density will have problems in classifying a sample that presents close values for its distances of Mahalanobis considered in relation to the means of the involved classes, a particular situation that induces misclassification errors.

The solution to fix this is to benefit both from the training set and from information proportioned by the new sample itself to be classified. Doing this, the classifier can take into account new information that will improve the overall generalization proportioned by these traditional methods. Therefore, the proposal in this chapter is to make the classification algorithm able to identify and provide treatment to the sample cases presenting close values for its Mahalanobis distances until it can reveal more clearly its actual class for the Gaussian classifier.

3. The Multivariate-Stepwise Gaussian Classifier: A new classification algorithm

What is proposed is a new classification method: The Multivariate-Stepwise Gaussian Classifier (MSGC). MSGC theoretically works on the basis of the already depicted GMLC method. Its contribution is to treat individually a sample to be classified if this sample presents close values for its Mahalanobis distances with respect to the class means involved in the classification, so that the discrimination made by the classifier is, in thesis, inconclusive. In this case, the algorithm will work employing dimensionality reduction by disregarding, one by one, in a stepwise process, the p dimensions involved in the calculation of the Mahalanobis distances until the calculated distances are dissimilar enough to give greater accuracy (likelihood) to the classification made by the method.

The key question is: what would be the best numerical dissimilarity between the distances of Mahalanobis obtained from a sample and the class means so that its classification is optimal? It can be anticipated that this response depends on the database to be focused, which will require the previous calibration of the method proposed here.

3.1. Description of the algorithm

Given c predefined classes and n sample vectors $X_i^T = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, depicting p measurements (dimensions) made on the sample from p attributes, x_{ij} means the jth measurement, $j = 1, \ldots, p$, for the ith sample. So, let **X** be a data matrix of type (n, p) with the measurements of data (x_{ij}) as elements, $j = 1, \ldots, p$ and $i = 1, \ldots, n$. The MSGC algorithm functions as depicted below, considering **X** as a training set [with p attributes (variables), n instances (training samples) and c classes], x_{hij} as an element of **X** belonging to a class h, $h = 1, \ldots, c$, μ_h, $h = 1, \ldots, c$, being the class mean vector for the training set, and the matrix **U** of type (v, p) as a new unknown set [with measurements of data (u_{sj}), $s = 1, \ldots, v$, $j = 1, \ldots, p$, as elements, with p attributes (variables), v cases (unknown samples) and c classes, each line of matrix **U** being a sample vector $U_s^T = \left(u_{s1}, u_{s2}, \ldots, u_{sp}\right)$], these unknown sample vectors having to be classified by the algorithm. Also, consider the density $f\left(X|\mu_h, \Sigma_h\right)$ and its complete description in Eq. (1). Besides, note below that MD_{hs} refer to 'the Mahalanobis distance for $U_s^T = \left(u_{s1}, u_{s2}, \ldots, u_{sp}\right)$ in relation to the class mean vector μ_h', and Δ is the 'the numerical dissimilarity between the distances of Mahalanobis (MD_{hs}, $h = 1, \ldots, c$) calculated from a sample and the class means μ_h.' The Multivariate-Stepwise Gaussian Classifier (MSGC) algorithm pseudocode is (considering $c = 2$):

---

(0) begin algorithm (initialize variables and counters, $s = 1$);

(1) while $s < v + 1$ do:

(1.1) $j = p$;

(1.2) while $j > 0$ do:

(1.2.1) calculate the mahalanobis distances MD_{hs} for $U_s^T = \left(u_{s1}, u_{s2}, \ldots, u_{sj}\right)$ in relation to each class mean vector μ_h of the training data, $h = 1, 2$;

(1.2.2) calculate $f\left(U_s|\mu_h, \Sigma_h\right)$ for each class h, $h = 1, 2$;

(1.2.3) if $j = p$ then do:

$f\left(U_s|\mu_h, \Sigma_h\right) = ff_h$ for each class h, $h = 1, 2$;

end if;

(1.2.4) if MD_{1s}-$MD_{2s} < \Delta$ then do:

$j = j$-1;

else.

$j = 0$;

end do (referring to step 1.2);

(1.3) if MD_{1s}-$MD_{2s} < \Delta$.

(1.3.1) if $ff_1 > ff_2$ then do:

assign to sample U_s the class $h = 1$;

end if;

(1.3.2) if $ff_2 > ff_1$ then do:

assign to sample U_s the class $h = 2$;

end if;

else.

(1.3.3) if $f(U_s \mid \mu_1, \Sigma_1) > f(U_s \mid \mu_2, \Sigma_2)$ then do:

assign to sample U_s the class $h = 1$;

end if;

(1.3.4) if $f(U_s \mid \mu_2, \Sigma_2) > f(U_s \mid \mu_1, \Sigma_1)$ then do:

assign to sample U_s the class $h = 2$;

end if;

(1.4) $s = s + 1$;

end do (referring to step 1);

end of the algorithm.

Note that for simplicity of exposition, the above pseudocode was written for $c = 2$, but steps (1.2.4) and (1.3) can be expanded for any value for c. Another important observation to be made about the pseudocode is that if in steps (1.2.4) and (1.3) Δ is set to zero, then the new algorithm will function strictly as a GMLC.

Finally in this section, it should be added that recent literature involving classifiers which are in some way based on the GMLC method makes no mention of an algorithm that works like MSGC. See [15–28].

The Multivariate-Stepwise Gaussian Classifier (MSGC) algorithm was implemented by means of The R Program for Statistical Computing [29] (version 2.14.0).

4. Comparing MSGC with traditional GMLC method

4.1. Methodology

Some real data sets from the UCI repository [2] (available from: http://archive.ics.uci.edu/ml/datasets/) are used to compare MSGC to GMLC method.

A 10-fold cross validation is widely used in the related literature like [13, 30] to present a more stable estimate of the performance of a classification method. Then, it was used here.

So to calibrate the MSGC algorithm and define the best value for Δ to be applied to each validation bloc, a previous 10-fold cross-validation was performed for each of the 10 training blocs. In the calibration process, the chosen criterion was the greatest accuracy rate reached over the 10-fold cross-validation. In this process were considered Δs starting from 0 up to 1.1, with increments of 0.1 at each iteration. If there was a tie during the process of calibration, the chosen Δ was the lower one. Thereafter, MSGC was submitted to each validation bloc (once adjusted with the best Δ regarding the corresponding training bloc and its proper process of 10-fold cross-validation).

For comparison GMLC was also implemented in the R program and applied to exactly the same blocs generated by the depicted 10-fold cross-validation process.

4.2. Presentation of data sets and comparison of classification results

Pima Indians Diabetes Data Set comprises 768 entries (8 medical and demographical attributes and a class variable), 550 of the entries classified as 0 and 268 classified as category 1. Attribute information: (1) number of times pregnant, (2) plasma glucose concentration a 2 hours in an oral glucose tolerance test, (3) diastolic blood pressure (mm Hg), (4) triceps skin fold thickness (mm), (5) 2-hour serum insulin (mu U/ml), (6) body mass index (weight in kg/(height in m)^2), (7) diabetes pedigree function, (8) age (years) and (9) class variable (0 or 1). Ten mutually exclusive folds were randomly sampled from Pima Indians Diabetes Data Set (9 validation folds including 77 entries and the tenth fold comprising 75). The key importance involved in the classification of Pima Indians Diabetes Data Set lies in the possibility of diagnosing diabetes disease, considering the numerical attributes, since class 1 is interpreted as tested positive for diabetes.

Breast Cancer Winsconsin (Original) Data Set comprises 699 entries (9 attributes and a class variable), 458 of them classified as category 2 "benign" and 241 classified as category 4 "malignant" (recoded as 0 and 1, respectively). Attribute information: (1) sample code number (id number), (2) clump thickness: 1–10, (3) uniformity of cell size: 1–10, (4) uniformity of cell shape: 1–10, (5) marginal adhesion: 1–10, (6) single epithelial cell size: 1–10, (7) bare nuclei: 1–10, (8) bland chromatin: 1–10, (9) normal nucleoli: 1–10, (10) mitoses: 1–10 and (11) class: (2 for benign and 4 for malignant). Ten mutually exclusive folds were randomly sampled from the Breast Cancer Winsconsin (Original) Data Set (9 validation folds including 69 entries and the tenth fold comprising 62). Sixteen original entries with missing data were removed. As for Breast Cancer Wisconsin (Original) Data Set, this data set can be used to predict the severity (benign or malignant) of a clump of cells in relation to the nine numerical attributes.

Haberman's Survival Data Set comprises 306 entries (three attributes and a class variable), 81 of them classified as category 2 and the remaining 225 classified as category 1 (recoded as 1 and 0, respectively). Attribute information: (1) age of patient at time of operation (numerical), (2) patient's year of operation (year-1900, numerical), (3) number of positive axillary nodes

detected (numerical) and (4) survival status (class attribute), 1 = the patient survived 5 years or longer, 2 = the patient died within 5 years. Ten mutually exclusive folds were randomly sampled from the Haberman's Survival Data Set (9 validation folds including 31 entries and the tenth fold comprising 27). The main interest in the classification task involving the Haberman's Survival Data Set would be the attempt to predict the life expectancy of patients undergoing breast cancer surgery, taking into account their age at the time of surgery and the number of axillary nodes removed.

Mammographic Mass Data Set presents discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age. It comprises 961 entries of data (five attributes and a class variable). The class associated with each record is the field 'severity,' 0 or 1. Attribute information: (1) BI-RADS assessment: 1–5 (ordinal), (2) age: patient's age in years (integer), (3) shape: mass shape: round = 1, oval = 2, lobular = 3, irregular = 4 (nominal), (4) Margin: mass margin: circumscribed = 1, microlobulated = 2, obscured = 3, ill-defined = 4, spiculated = 5 (nominal), (5) Density: mass density: high = 1, iso = 2, low = 3, fat-containing = 4 (ordinal) and (6) severity: benign = 0 or malignant = 1 (binominal). A total of 131 original entries with missing data were removed. Ten mutually exclusive folds were randomly sampled from the Mammographic Mass Data Set (all of them with 83 entries). In relation to the Mammographic Mass Data Set, [2] informs that "*Mammography is the most effective method for breast cancer screening available today. However, the low positive predictive value of breast biopsy resulting from mammogram interpretation leads to approximately 70% unnecessary biopsies with benign outcomes. (…) This data set can be used to predict the severity (benign or malignant) of a mammographic mass lesion from BI-RADS attributes and the patient's age*."

To illustrate the 10-fold cross-validation process for MSGC calibration, **Table 1** summarizes the values for Δ that gave the greatest accuracy rate (%) for all the 10 training blocs. Remembering that there are two classes in all the data sets considered in the process. Afterward, these best settings for Δ (in **Table 1**) were applied to steps (1.2.4) and (1.3) in MSGC algorithm in order to classify the corresponding validation blocs.

From **Table 1**, it is possible to see that best values for Δ= 0.0 imply that MSGC optimally will work as a traditional GMLC for the Breast Cancer Winsconsin (Original) Data Set and Haberman's Survival Data Set classification.

DATA	Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5	Bloc 6	Bloc 7	Bloc 8	Bloc 9	Bloc 10
PI	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.1
BR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MA	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4

DATA SETS: PI = PIMA; BR = BREAST; HB = HABERMAN'S; MA = MAMMOGRAPHIC.

Table 1. Summary of the 10-fold cross-validation calibration process - The Δ settings giving best accuracy rate concerning training blocs.

Table 2 shows synoptically the accuracy rate mean and standard error for all data sets and methods (the best results for each data sets are highlighted in bold). Both methods were proficient in classifying data and obtained relatively similar results.

From **Table 2,**we can see that MSGC attained the greatest accuracy rate in two out of four data sets (PIMA and MAMMOGRAPHIC). For HABERMAN'S and BREAST, both methods achieved identical results since for these data sets MSGC was set with $\Delta = 0.0$. It has been seen that the Mammographic Mass Data Set source [2] reports that because of the low positive predictive value of the exam, about 70% of the biopsies are actually unnecessary as they end up showing benign lesions. With the practical use of classification algorithms such as MSGC and GMLC, a favorable new situation is achieved, with levels of diagnostic accuracy above 80%. This rate means that patients can be treated adequately, while biopsies may be subsidiarily requested.

Note that accuracy rate was chosen as the criterion for comparison, but in medicine, sometimes the physician needs to know other criteria like sensitivity, specificity or precision; in this case, the data analyst should take care to also calculate them based on the algorithm results.

We also have to remark the positive aspect that these results for MSGC algorithm are transcendent. Since GMLC is the basis on which other traditional classification methods (namely RDA, QDA and LDA) are based, an improvement made in GMLC, such as this obtained through the MSGC method, will probably imply improvements in performance also for RDA, QDA and LDA. Future research shall prove this.

DATA	MSGC	GMLC
PIMA	**73.67 (2.05)**	73.41 (2.25)
BREAST	94.92 (0.75)	94.92 (0.75)
HABERMAN'S	75.10 (2.42)	75.10 (2.42)
MAMMOGRAPHIC	**80.36 (1.27)**	80.00 (1.11)

SE: Standard error for accuracy rate mean.

Table 2. Classification results for real data sets—accuracy rate mean % (SE).

5. Conclusion

A new classification algorithm is presented in this chapter: The Multivariate-Stepwise Gaussian Classifier (MSGC).

MSGC theoretically works on the basis of the Gaussian Maximum Likelihood Classifier (GMLC) method. Its contribution is to treat individually a sample to be classified if this sample presents close values for its Mahalanobis distances with respect to the class means involved in the classification, so that the discrimination made by the classifier is, in thesis, inconclusive.

In this case, MSGC will work employing dimensionality reduction by disregarding, one by one, in a stepwise process, the p dimensions involved in the calculation of the Mahalanobis distances until the calculated distances are dissimilar enough to give greater accuracy to the classification made by the base method (GMLC).

For better performance, MSGC may be previously calibrated by means of a training set. A 10-fold cross-validation process was used to calibrate the algorithm.

MSGC was applied for data classification and its performance was compared with the traditional GMLC method considering four real medical data sets available in the UCI data repository. These data represent a range of different types of data dependence structure and dimensionality. The results showed that the performance of the MSGC algorithm is at least as competitive as GMLC. MSGC attained the greatest accuracy rate in two of the data sets (PIMA and MAMMOGRAPHIC). For HABERMAN'S and BREAST data sets, both methods achieved identical results. It was concluded that MSGC can be used as an effective classification tool in a wide range of data sets.

The presented results for the MSGC algorithm are transcendent. Since GMLC is the basis on which other traditional classification methods (namely RDA, QDA and LDA) are based, an improvement made in GMLC, such as this obtained through the MSGC algorithm, will probably imply improvements in performance also for RDA, QDA and LDA.

After reaching the conclusions, an additional discussion arises. With the emergence of the big data, as a robust successor to data mining emerged from the exponential development of computers and storage media since the 1990s, it has been a tendency to think of the intensive use of multiple algorithms simultaneously, in supervised or nonsupervised approaches, to analyze data and discover patterns. This certainly makes sense, as it has already been mentioned in this chapter that there is no one classification method or algorithm better than another. Beyond to a greater robustness or scalability of some methods over others, what concrete exists is a dependence of the results against the target database.

Therefore, in this context, and considering a matter as important as the medical clinic, once a classification method has been tested and successfully validated considering a particular scope of data, the most recommended would be its use for the best diagnosis. Meanwhile, if possible, already known or new algorithms could be tested for various diseases and symptoms data until they proved to be robust and effective enough to be put into medical practice.

Finally, it is important to remark that mathematical classifier serves as an aid to the crucial medical diagnosis made by the physician.

Acknowledgements

The author would like to thank the UCI Machine Learning Repository and the data donors for putting real data sets at the disposal of the scientific community and would also like to thank The R Foundation for Statistical Computing and its contributors for developing and making the R program available to the public. The Breast Cancer Wisconsin (original) Data Set was

obtained from the University of Wisconsin Hospitals, Madison, from Dr. William H. Wolberg, and the author would like to thank him.

Nomenclature

GMLC	Gaussian Maximum Likelihood Classifier
LDA	linear discriminant analysis
MSGC	Multivariate-Stepwise Gaussian Classifier
QDA	quadratic discriminant analysis
RDA	regularized discriminant analysis

Author details

Alexandre Serra Barreto

Address all correspondence to: alexsbdr@gmail.com

Ministry of Finance (Brazil), Brasília-DF, Brazil

References

[1] Barreto AS. Multivariate statistical analysis for dermatological disease diagnosis. In: Proceedings of the IEEE International Conference On Biomedical And Health Informatics (BHI '2014); 1-4 June 2014; Valencia. New York: IEEE; 2014. p. 500-504

[2] Dua D, Karra Taniskidou E. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. 2017. http://archive.ics.uci.edu/ml

[3] Demiroz G, Guvenir HA, Ilter N. Learning differential diagnosis of Eryhemato-squamous diseases using voting feature intervals. Artificial Intelligence in Medicine. 1998;**13**(3):147-165. DOI: 10.1016/S0933-3657(98)00028-1

[4] Pappa GL, Freitas AA, Kaestner CAA. Attribute selection with a multi-objective genetic algorithm. In: Proceedings of 16th Brazilian Symposium on Artificial Intelligence (SBIA); 11–14 November, 2002; Recife: Springer; 2002. p. 280-290

[5] Fidelis MV, Lopes HS, Freitas AA. Discovering comprehensible classification rules with a genetic algorithm. In: Proceedings of the 2000 Congress on Evolutionary Computation (CEC00). 16-19 July, 2000; La Jolla. New York: IEEE; 2000. p. 805-810

[6] Grandgeorge D. The Spirit of Homeopathic Medicines. Berkeley: North Atlantic Books; 1998

[7] Freud S. The Essentials of Psichoanalisys. London: Penguin Books; 1986

[8] Ali N, Khuwaja A, Kausar S, Patients NK. Evaluations of family practice care and attributes of a good family physician. Radcliffe Publishing, Quality in Primary Care. 2012;**20**: 375-383

[9] Rakel R. Family physician. In: Rakel R, Rakel D, editors. Textbook of Family Medicine. 8th ed. Philadelphia, PA: Elsevier Saunders; 2011. p. 1-14. DOI: 10.1016/B978-1-4377-1160-8.10001-6.ch1

[10] Hoffbeck JP, Landgrebe DA. Covariance matrix estimation and classification with limited training data. Pattern Analysis and Machine Intelligence. 1994;**18**(7):763-767. DOI: 10.1109/34.506799

[11] Barreto AS. Weighted correlation matrix similarity: A new classification algorithm. In: Proceedings of the IADIS European Conference on Data Mining 2012: Part of the IADIS Multi Conference on Computer Science and Information Systems 2012 (IADIS DATA MINING 2012); 17–23 July 2012; Lisbon. Lisbon: IADIS; 2012. p. 79-90. ISBN: 978-972-8939-69-4

[12] Friedman JH. Regularized discriminant analysis. Journal of the American Statistical Association. 1989;**84**(405):165-175. DOI: 10.1080/01621459.1989.10478752

[13] Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002. 495 p. ISBN: 0-387-95457-0

[14] Rpubs. Classification: Linear Discriminant Analysis [Internet]. 2014. Available from: https://rpubs.com/ryankelly/LDA-QDA [Accessed: 2017-12-24]

[15] Skolidis G, Sanguinetti G. Bayesian multitask classification with Gaussian process priors. IEEE Transactions on Neural Networks. 2011;**22**(12):2011-2021. DOI: 10.1109/TNN.2011.2168568

[16] Maugis C, Celeux G, Martin-Magniette M-L. Variable selection in model-based discriminant analysis. Journal of Multivariate Analysis. 2011;**102**(10):1374-1387. DOI: 10.1016/j.jmva.2011.05.004

[17] Hyun-Chul Kim Z, Ghahramani Z. Bayesian Gaussian process classification with the EM-EP algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; **28**(12):1948-1959. DOI: 10.1109/TPAMI.2006.238

[18] Opper M, Winther O. Gaussian processes for classification: Mean-field algorithms. Neural Computation. 2000;**12**(11):2655-2684. DOI: 10.1162/089976600300014881

[19] Ye J, Janardan R, Li Q, Park H. Feature reduction via generalized uncorrelated linear discriminant analysis. IEEE Transactions on Knowledge and Data Engineering. 2006; **18**(10):1312-1321. DOI: 10.1109/TKDE.2006.160

[20] Guo P, Jia Y, Lyu MR. A study of regularized Gaussian classifier in high-dimension small sample set case based on MDL principle with application to spectrum recognition. Pattern Recognition. 2008;**41**:2842-2854. DOI: 10.1016/j.patcog.2008.02.004

[21] Halbe Z, Aladjem M. Regularized mixture discriminant analysis. Pattern Recognition Letters. 2007;**28**:2104-2115. DOI: 10.1016/j.patrec.2007.06.009

[22] Ji S, Ye J. Kernel uncorrelated and regularized discriminant analysis: A theoretical and computational study. IEEE Transactions on Knowledge and Data Engineering. 2008; **20**(10):1311-1321. DOI: 10.1109/TKDE.2008.57

[23] Licheng J et al. An organizational coevolutionary algorithm for classification. IEEE Transactions on Evolutionary Computation. 2006;**10**(1):67-80

[24] Liu J, Chen S, Tan X. A study on three linear discriminant analysis based methods in small sample size problem. Pattern Recognition. 2008;**41**:102-116. DOI: 10.1016/j.patcog.2007. 06.001

[25] Lu H, Plataniotis KN, Venetsanopoulos AN. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. IEEE Transactions on Neural Networks. 2009;**20**(1):103-123. DOI: 10.1109/TNN.2008.2004625

[26] Lu H, Plataniotis KN, Venetsanopoulos AN. Regularized discriminant analysis for the small sample size problem in face recognition. Pattern Recognition Letters. 2003;**24**:3079-3087. DOI 10.1016/j.patcog.2007.06.001

[27] Peng J, Zhang P, Riedel N. Discriminant learning analysis. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics. 2008;**38**(6):1614-1625. DOI: 10.1109/TSMCB. 2008.2002852

[28] Xu P, Brock G, Parrish R. Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. Computational Statistics & Data Analysis. 2009;**53**:1674-1687. DOI: 10.1016/j.csda.2008.02.005

[29] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. URL: http://www.R-project.org/.2011

[30] Jiao L, Liu J, Zhong W. An organizational coevolutionary algorithm for classification. IEEE Transactions on Evolutionary Computation. 2006;**10**(1):67-80. DOI: 10.1109/TEVC. 2005.856068

IntechOpen