COMPARATIVE ANALYSIS OF TWO CLASSIFIERS IMPLEMENTING NOMINAL LOGISTIC REGRESSION

Hélio Radke BITTENCOURT[1]
Bárbara Patrícia Olberman PASINI[1]
Denis Altieri de Oliveira MORAES[2]
Bruno Duarte dos SANTOS[3]
Vitor HAERTEL[4]

- ABSTRACT: There is a wide range of statistical classification methods described in the scientific literature. Each method presents advantages and disadvantages that depend mainly on the probabilistic behaviour of the classes. This article presents two methods of classification based on logistic regression and shows results obtained with three real data sets. The first method is a single-stage classifier based on Nominal Logistic Regression. The second method is a hierarchical classifier, structured like a binary decision tree that uses the traditional logistic regression at each stage. Both methods are compared with traditional classifiers and the results are discussed.

- KEYWORDS: Nominal logistic regression; decision trees; classifier; pattern recognition.

1 Introduction

Statistical approaches to the pattern recognition problem have been extensively discussed in the scientific literature and implemented in a large number of commercial systems (Bittencourt and Clarke, 2003; Jain, Duin and Mao, 2000). In such approaches, each pattern is regarded as a p-dimensional random vector, where p is the number of characteristics used for classification. In a survey of papers published in IEEE Transactions on Pattern Analysis and Machine Intelligence, since 1979, 350 articles on pattern recognition were found, 300 of them concerned with aspects of statistical approaches to the problem (Jain, Duin and Mao 2000).

In statistical pattern recognition, each pattern is regarded as a random vector $\mathbf{x}$, denoted by $\mathbf{x} = (x_1, x_2, \ldots x_p)^T$ characterized by measurements of p variables, which can be represented in a metric space known as feature space (Aeberhard, Coomans and De Vel,

[1] Departamento de Estatística, Faculdade de Matemática, PUCRS, Av. Ipiranga, 6681 CEP: 90619-900, Porto Alegre, Brazil, and Programa de Pós-Graduação em Geografia, UFRGS, Brazil. E-mail: *heliorb@pucrs.br / barbarap@pucrs.br*
[2] Departamento de Estatística, UFSM, Av. Roraima, 1000, CEP: 97105-900, Santa Maria, RS, Brazil, E-mail: *d_altieri@yahoo.com.br*
[3] Bolsista de Iniciação Científica, Pibic– PUCRS, Porto Alegre, RS, Brazil. E-mail: *brunoduantos@hotmail.com*
[4] Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, UFRGS, Porto Alegre, RS, Brazil, Caixa Postal: 15044, CEP: 91501-970, Porto Alegre, RS, Brazil, E-mail: *victor.haertel@ufrgs.br*

1994; Duda, Hart and Stork, 2000; Gonzalez, Woods and Eddins, 2004). One assumes that different classes of patterns have different behaviour in the feature space, allowing the use of classification methods based on spatial location (Aeberhard, Coomans and De Vel, 1994). According to Jain, Duin and Mao (2000), classification is the main goal of Pattern Recognition.

The problem of classifying a pattern **x** in one of k possible classes w_1, w_2, ... , w_k assumes that each pattern can be allocated to one and only one class. Determining the rules of allocating, patterns in pre-defined classes can be regarded as a problem of discriminant analysis, in which the quantity and quality of available information on the patterns would determine the efficiency of rules for the correct discrimination between the respective classes of pattern (Bittencourt and Clarke, 2002).

The classification procedures can be divided in supervised or unsupervised; parametric or nonparametric ones. The supervised methods need training samples in which all patterns are labeled, that is, the classes which the patterns belong are known. The unsupervised ones, in turn, do not need it. Parametric classifiers assume that the underlying class distribution has a determined probabilistic behaviour, while the nonparametric ones do not make assumptions on the functional form of entry data (Cheng, Varshney and Arora, 2006).

This paper presents two supervised and partially parametric classifiers: Nominal Logistic Regression (NLR) and Decision Trees implementing Logistic Regression (DTLR). The results obtained (accuracies and Kappa coefficients) by using these two classifiers are compared between them and also with traditional methods.

1.1 Nominal logistic regression – single stage classifier

The logistic regression model was originally proposed meant to discriminate only two classes. An extension to k classes (k>2) – known as nominal logistic regression or NLR – is presented by Hosmer and Lemeshow (1989). The nominal model must be used when the classes w_1, w_2, ... , w_k are not hierarchically ordered. In other cases, the ordinal logistic model is more appropriate, but this method is not of interest here.

In the NLR model, the probability that a pattern **x** belongs to the class w_i is estimated directly from the following expression:

$$P(w_i \mid \mathbf{x}) = \frac{\exp(\beta_{i0} + \boldsymbol{\beta}_i^T \mathbf{x})}{1 + \sum_{j=1}^{k-1} \exp(\beta_{j0} + \boldsymbol{\beta}_j^T \mathbf{x})} \qquad i=1, 2, \ldots, k \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots x_p)^T$ is the vector characterized by measurements of p variables, β_{i0} is the constant and $\boldsymbol{\beta}_i$ is the vector of parameters associated to the p variables of vector $\mathbf{x}$.

The logistic model requires the estimation of k-1 vectors of parameters $\boldsymbol{\beta}_i$, corresponding to the k-1 classes to be discriminated. The *k-th* class is taken as a basis, from which the natural log of the ratio of the two probabilities become linear functions of the parameters. Hence, in a model to discriminate k classes, there are k-1 functions $g_i(\mathbf{x})$, as in (2).

$$g_i(\boldsymbol{x}) = ln\,(P(w_i \mid \boldsymbol{x}) / P(w_k \mid \boldsymbol{x})) = \beta_{i0} + \boldsymbol{\beta}^T_i \boldsymbol{x}. \tag{2}$$

The assumption of linearity is fundamental to the logistic approach, and for that reason it is called a partially parametric model, since only the linear functions are modeled. Despite the apparent strength of this assumption, (Anderson, 1992; Efron, 1975) state that the logistic model can be used with a wide range of probability distributions and, at least in theory, logistic discrimination has greater robustness than the method of Gaussian maximum likelihood. The classification rule in logistic discrimination is very simple, given by expression (3).

$$\boldsymbol{x} \in w_i \quad if \quad P(w_i \mid \boldsymbol{x}) > P(w_j \mid \boldsymbol{x}) \quad \forall\, j \neq i \tag{3}$$

It is obvious that the sum of all probabilities $P(w_i \mid \mathbf{x})$ to i=1,…, k results equal to one. This classification rule based on the probability is equivalent to the following simpler one:

$$\boldsymbol{x} \in w_i \quad g_i(\boldsymbol{x}) > g_j(\boldsymbol{x}) \quad \forall\, j \neq i \tag{4}$$

The classifier based on the rules (3) or (4) can be regarded as single stage methods because the patterns are initially clustered in a single group. Thus, the decision function is estimated for each class and the patterns **x** are allocated to the winner class. The Figure 1 shows the classification process used by a single stage classifier.

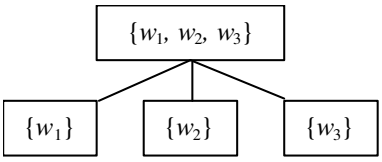

Figure 1 - Structure of single stage classifier in a three-class example.

1.2 Binary decision tree – multiple stage classifier

Another possible classification approach consists in the use of multiple stage classifiers – or decision tree classifiers – where the multi-class global problem is subdivided in smallest local problems. In this approach, the classification process of each vector **x** considers only a subset of classes at each stage. However, as the number of classes increases, so increases the number of possible decision trees, becoming more difficult to select an optimal structure to the classifier.

A common way to deal with the diversity of solutions to the tree is to use a binary type of tree. In this case, in each node only two classes – the most distinct of them – are considered, which originates two descending nodes (Breiman et al., 1984; McLachlan, 1992). The procedure to generate the tree and the classification process are as described in Moraes and Haertel (2008). First, the structure of the binary tree is found with the application of training samples, following the steps:

i) At the root node, considering all classes, the distances between all pairs of classes are computed. In this work, the Bhattacharyya distance (5) under the assumption of normality was used (Fukunaga, 1990). This distance is defined for two classes, where $\mathbf{M}_i$, $\mathbf{M}_j$, Σ_i and Σ_j are the mean vectors and the variance/covariance matrix of the pair

of classes w_i and w_j to be compared at the node. The estimation of the mean vectors and variance/covariance matrix are obtained by the training samples of each class at the node. The first term of the Bhattacharyya distance measures the contribution to the class separability due to the difference between the mean vectors, and the second term measures the contribution due to the compounded variance/covariance matrix of the two classes.

$$B = \frac{1}{8}\left(\mathbf{M}_i - \mathbf{M}_j\right)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} \left(\mathbf{M}_i - \mathbf{M}_j\right) + \frac{1}{2}\ln\left(\frac{\left|\frac{1}{2}\left(\Sigma_i + \Sigma_j\right)\right|}{\left|\Sigma_i\right|^{1/2}\left|\Sigma_j\right|^{1/2}}\right) \qquad (5)$$

$$i, j = 1, 2, ..., k \qquad \forall\ i \neq j$$

ii) Once defined the two most different classes at the node, they originate two branches and descending nodes, left and right, and all their samples are allocated in separated nodes, respectively.

iii) All the other classes are allocated simultaneously into both descending nodes for further comparisons, until the last but one level of the tree is reached.

iv) In the last but one level, which all nodes contain only two classes, are originated the tree leafs, or the terminal nodes. At this stage, considering three classes, a tree with the structure of Figure 2 is found.

The classification process of the test samples, or validation process, consists on testing all the patterns allocated at the root node individually. This way, the probability of each pattern belongs to a given class is computed by using the estimated logistic function for the node. This function is generated by the two most different classes (training samples) at the node and the pattern goes to the node that scores the highest value in probability. Once the pattern reaches a leaf of the tree, it is classified with the label of this terminal node, and the process continues with the next pattern at the root node until all validation samples are classified.

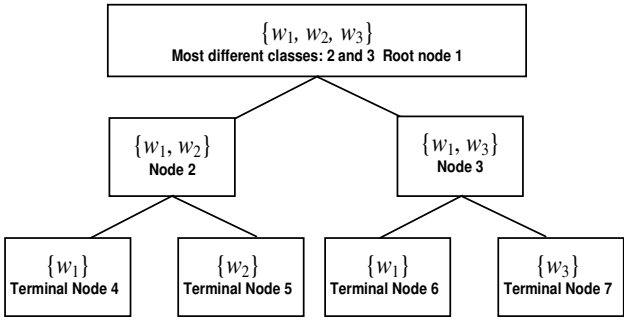

Figure 2 - Structure of binary tree classifier in a three-class case.

2 Methodology

The methodology of this study was inspired in Aeberhard, Coomans and De Vel (1994) in which the authors compared eight statistical classification methods using a wide range of artificial and real data sets. The classifiers used by them were: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Regularized Discriminant Analysis (RDA), Nearest Neighbour Method (1NN), Fisher's Discriminant Plane (FDP), Fisher-Fukunaga-Koonz transform (FF), Fisher Radius transform (FR), Fisher-Variance transform (FV). A brief description of these classifiers can be found in Aeberhard, Coomans and De Vel (1994).

In this paper the two considered classifiers – NLR and DTLR – are applied on three real data sets, two of which were also used by Aeberhard, Coomans and De Vel (1994). All real databases were randomly divided in training and validation sample (fifty-fifty), i.e. the model estimated in the training sample was used to classify another half (validation sample). This procedure was repeated 10 times and the average results were obtained. The results were organized in tables on which the accuracies and Kappa coefficients were calculated. MS Excel® was used to organize the real data sets and calculate the measures of distance. The SPSS® software was used to estimate the logistic regression parameters and get the Kappa coefficients with its confidence intervals. Kappa coefficient was proposed by Cohen (1960). It is a statistical measure of the agreement between two classifications more robust than a simple percent agreement calculation, having the advantage to allow the computation of confidence intervals.

When possible, the results obtained by using the two classifies based on logistic regression were compared with Aeberhard, Coomans and De Vel (1994).

2.1 Real datasets

A brief description of each data set is shown in this item in order to give a real sense of them. All databases were found in UCI (2007), more specifically in the Machine Learning Repository at University of California – Irvine. Table 1 shows the descriptive statistics for the three databases and their respective classes.

The first dataset – Balance Scale – was originally generated to model psychological experimental results. It has 625 observations of three classes that specify if the balance scale tip to the right, tip to the left, or if it is balanced. The four attributes are: left weight, left distance, right weight and right distance. According to UCI (2007) the correct way to find the class is the higher value between (left-distance*left-weight) and (right-distance*right-weight). If they are equal, it is balanced.

The second dataset – Wine Recognition – uses a chemical analysis to determinate the origin of wines. Three classes and thirteen continuous variables compose this dataset with 178 observations in the total.

The third database – Iris Fisher Data (Fisher, 1936) – is one of the most known databases to be found in the Pattern Recognition literature. The data set contains three classes of fifty observations each, where each class refers to a type of iris plant. Four quantitative variables are available to describe the plants. One class is linearly separable from the other two; the latter are not linearly separable from each other.

Table 1 - Descriptive statistics to the three databases per class

Dataset	*Variables*	*Mean*	*Std Dev.*	*Mean*	*Std Dev.*	*Mean*	*Std Dev.*
		Class 1:Left		***Class 2: Balanced***		***Class 3: Right***	
Balance Scale Weight	Left-Weight (LW)	1.056	0.236	1.167	0.389	1.093	0.291
	Left-Distance (LD)	4.000	1.188	2.917	1.505	2.602	1.427
	Right-Weight (RW)	1.611	0.916	2.000	1.348	3.204	1.331
	Right-Distance (RD)	1.611	0.916	2.000	1.348	3.315	1.309
		Class 1: Cultivar A		***Class 2: Cultivar B***		***Class 3: Cultivar C***	
Wine Recog.	Alcohol	13.745	0.462	12.279	0.538	13.176	0.547
	Malic acid	2.011	0.689	1.933	1.016	3.301	1.101
	Ash	2.456	0.227	2.245	0.315	2.437	0.183
	Alcalinity of ash	17.037	2.546	20.238	3.350	21.298	2.384
	Magnesium	106.339	10.499	94.549	16.753	99.878	11.479
	Total phenols	2.840	0.339	2.259	0.545	1.702	0.388
	Flavanoids	2.982	0.397	2.081	0.706	0.828	0.436
	Nonflavanoid	0.290	0.070	0.364	0.124	0.444	0.125
	Proanthocyanins	1.899	0.412	1.630	0.602	1.177	0.436
	Color intensity	5.528	1.239	3.087	0.925	7.360	2.300
	Hue	1.062	0.116	1.056	0.203	0.690	0.124
	OD280/OD315	3.158	0.357	2.785	0.497	1.729	0.418
	Proline	1,115.712	221.521	519.507	157.211	638.776	129.749
		Class 1: Setosa		***Class 2: Versicolor***		***Class 3: Virginica***	
Iris Fisher Data	Sepal length (cm)	5.006	0.352	5.936	0.516	6.588	0.636
	Sepal width (cm)	3.418	0.381	2.770	0.314	2.974	0.322
	Petal length (cm)	1.464	0.174	4.260	0.470	5.552	0.552
	Petal width (cm)	0.244	0.107	1.326	0.198	2.026	0.275

3 Results

For reasons of clarity, the results of the first database – Balance Scale Weight – are detailed, including an example of the parameters estimated in each classifier, the Bhattacharyya's distance matrix and the decision tree. Concerning to the other sets, only the results are shown. It is important to point out that the procedures were repeated ten times per classifier in each database, always using approximately 50% of the database to training and the another half to validation. The estimates reached by the NLR procedure in the first simulation are given by (6).

$$P(w_i \mid \mathbf{x}) = \frac{\exp\{g_i(\mathbf{x})\}}{1+\sum_{j=1}^{2} \exp\{g_j(\mathbf{x})\}}, \quad \text{where} \quad i=1, 2, 3 \quad \text{and}$$

$$\begin{aligned} g_1(\mathbf{x}) &= 15.83 + 15.18*LW + 10.24*LD - 30.10*RW - 26.15*RD \\ g_2(\mathbf{x}) &= 30.18 + 03.23*LW + 05.02*LD - 09.89*RW - 08.97*RD \\ g_3(\mathbf{x}) &= 0 \end{aligned} \tag{6}$$

The logistic regression procedure, available in the SPSS® software, estimates the k-1 functions at the same time, so the classifier NLR is a single stage method. When a binary tree is used to discriminate three classes (see Figure 2), three independent binary logistic regressions are performed. The parameters estimated are shown in (7) and the Bhattacharyya's distance matrix is shown in Table 2. As in (6), the $P(w_i|\mathbf{x})$ term in (7) allows to estimate the probability of the pattern $\mathbf{x}$ belongs to the class w_i

$$
\begin{gathered}
w_1 \text{ from } w_3 \text{ (root node 1)} \\
P(w_1|\mathbf{x}) = 1/(1+\exp\{4.02*LW+4.09*LD-7.75*RW-11.89*RD+3.312\}) \\
w_2 \text{ from } w_3 \text{ (node 2)} \\
P(w_2|\mathbf{x}) = 1/(1+\exp\{71.92*LW+14.59*LD-65.69*RW-122.75*RD+32.40\}) \\
w_1 \text{ from } w_2 \text{ (node 3)} \\
P(w_1|\mathbf{x}) = 1/(1+\exp\{7.35*LW+8.84*LD-12.26*RW-16.90*RD-19.90\})
\end{gathered}
\tag{7}
$$

Table 2 - Bhattacharyya's distance matrix for a training sample of Balance Scale Weight

	w_1	w_2	w_3
Class w_1 – Left	0	0.7109	1.1438
Class w_2 – Balanced	0.7109	0	0.6499
Class w_3 – Right	1.1438	0.6499	0

Table 3 shows the accuracies and the Kappa coefficients reached on ten validation samples. It is possible to visualize that there is no significant differences between the two classifiers because there are intersections among the confidence intervals. Therefore, the results suggest that the classifiers are equivalent concerning to accuracy.

Table 3 - Accuracies and Kappa coefficients reached in validation samples on the Balance Scale Weight database

Simulation	NLR				DTLR			
	Accuracy	Kappa	Kappa CI 95%		Accuracy	Kappa	Kappa CI 95%	
			Lower	Upper			Lower	Upper
1	88.89%	0.722	0.537	0.908	90.48%	0.762	0.584	0.941
2	96.83%	0.899	0.765	1.034	93.65%	0.813	0.640	0.986
3	88.16%	0.630	0.435	0.826	90.79%	0.712	0.517	0.907
4	91.67%	0.733	0.556	0.910	90.28%	0.699	0.515	0.883
5	92.54%	0.738	0.527	0.948	86.15%	0.583	0.409	0.757
6	88.61%	0.667	0.494	0.840	86.15%	0.656	0.463	0.849
7	91.14%	0.738	0.570	0.907	92.86%	0.794	0.639	0.949
8	92.50%	0.761	0.595	0.927	87.88%	0.651	0.467	0.835
9	86.15%	0.646	0.455	0.838	91.05%	0.729	0.526	0.931
10	90.79%	0.752	0.581	0.924	88.41%	0.691	0.519	0.864
Average	90.73%	0.729	0.551	0.906	89.77%	0.709	0.528	0.890

All values presented in Table 3 were bigger than 0.60. Landis and Koch (1977) affirms that Kappa coefficients bigger than 0.60 disclose substantial agreement.

The two following databases were also used by Aeberhard, Coomans and De Vel (1994) allowing a comparison between the results reached by several classification methods. Table 4 shows the results achieved by using NLR and DTLR classifiers (average results based on 10 validation samples) and Table 5 presents the Kappa coefficients.

Table 4 - Average accuracies based on ten validation samples and its comparison with other methods of classification

Classifier	**Iris Fisher Data (3 classes, p=4)**	**Wine Recognition (3 classes, p=13)**
Nominal Logistic Regression (NLR)	96.0%	95.7%
Decision Trees – Logistic Regression (DTLR)	94.2%	94.6%
Linear Discriminant Analysis (LDA)	97.3%	99.4%
Quadratic Discriminant Analysis (QDA)	98.0%	98.9%
Regularized Discriminant Analysis (RDA)	98.0%	100%
Nearest Neighbour Method (1NN)	96.7%	91.0%
Fisher's Discriminant Plane (FDP)	97.3%	88.2%
Fisher-fukunaga-Koonz transform (FF)	96.0%	99.4%
Fisher Radius transform (FR)	70.7%	93.8%
Fisher-Variance transform (FV)	98.0%	74.1%

Table 5 - Average Kappa coefficients based on ten validation samples and its confidence intervals 95%

Classifier	*Iris Fisher Data (3 classes, p=4)*	*Wine Recognition (3 classes, p=13)*
Nominal Logistic Regression (NLR)	0.940 (0.877 ; 1.000)	0.935 (0.872 ; 0.997)
Decision Trees – Logistic Regression (DTLR)	0.921 (0.844 ; 0.999)	0.909 (0.821 ; 0.998)

Conclusions

Both classifiers – NLR and DTLR – presented satisfactory results. The average accuracies were close to the obtained from other well known classifiers. The accuracies reached by the other classifiers (except FR) on the Iris Fisher Data were higher than NLR and DTLR ones. However, in the Wine Recognition database, the results reached by the methods based on logistic regression were better than 1NN, FDP, FR and FV ones. In the two databases, RDA presented the highest accuracies. Aeberhard, Coomans and De Vel (1994) point out that the RDA reached the best overall performances for eight of nine databases analyzed by them. The reason is that RDA uses a regularized covariance matrix and this makes it a more versatile technique than QDA and LDA. As RDA does not

perform reduction of dimensionality, as FDP, FF, FR and FV classifiers, the accuracies achieved by it tends to be higher.

Statistically speaking, the two classifiers based on logistic regression presented equivalent results because there were overlaps in the confidence intervals for the Kappa coefficients. When compared to the three databases, the single stage classifier (NLR) showed a slight superiority on the multiple stage method (DTLR). These results differ from the ones found by Bittencourt, Moraes and Haertel (2007) in the digital images classification. These authors found a slight superiority of the classifier structured as a decision tree on the single stage NLR classifier. This apparent contradiction is not surprising because the differences between both classifiers are not statistically significant. Also, Bittencourt, Moraes and Haertel (2007) presented results obtained on very high dimensional sets (p=190), different from the databases used here.

The results suggest that the both classifiers based on Logistic Regression can be used as an alternative in the pattern recognition problems.

Acknowledgements

The authors acknowledge the support received from the CNPq and PUCRS. They also thank to the anonymous referees for the very helpful comments and suggestions.

BITTENCOURT, H. R.; PASINI, B. P. O. ; MORAES, D. A. O.; SANTOS, B. D. ; HAERTEL, V. F. A. Análise comparativa dos dois métodos de classificação baseados em regressão logística. *Rev. Bras. Biom.*, São Paulo, v.27, n.1, p.115-124, 2009.

- *RESUMO: Existe uma ampla variedade de métodos de classificação descritos na literatura cientifica. Cada método apresenta vantagens e desvantagens que dependem principalmente do comportamento probabilístico das classes. Este artigo apresenta dois métodos de classificação baseados em regressão logística e mostra resultados obtidos com a utilização de três bancos de dados reais. O primeiro método e a classificador de estagio único baseado na regressão logística nominal. O segundo método é um classificador hierárquico estruturado como uma arvore de decisão binária, o qual usa regressão logística tradicional em cada estagio. Ambos métodos são comparados com classificadores tradicionais e os resultados são discutidos.*

- *PALAVRAS-CHAVE: Regressão logística nominal; árvores de decisão; classificador; reconhecimento de padrões.*

References

AEBERHARD, S.; COOMANS, D.; DE VEL, O. Comparative analysis of statistical pattern recognition methods in high dimensional settings. *Pattern Recognit*, Amsterdam, v.27, n.8, p.1065-77, 1994.

ANDERSON, J. A. Logistic discrimination. In: KRISHNAIAH, P. R.; KANAL, L. (ed.) *Handbook of statistics.* Amsterdam: North-Holland, 1992. v.2, p.169-191.

BITTENCOURT, H. R. CLARKE, R. T. Use of logistic discrimination to classify remotely-sensed digital images, In: PORTUGUESE CONFERENCE ON PATTERN RECOGNITION, 12., 2002, Aveiro. *Proceedings...* 2002.

BITTENCOURT, H. R.; CLARKE, R. T. logistic discrimination between classes with nearly equal spectral response in high dimensionality. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2003, Toulouse. *Proceedings…* p. 3748-3750.

BITTENCOURT, H. R.; MORAES, D. A.; HAERTEL, V. Single and multiple stage classifiers implementing Logistic Regression. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., 2007. *Proceedings* …p. 6431-6436.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R.A.; STONE, C. J. *Classification and regression trees*. Belmont, CA: Wadsworth, 1984.

COHEN, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.,* v.20, p.37–46, 1960.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*, 2. ed., Wiley, 2000.

EFRON, B. The efficiency of logistic regression compared to normal discriminant analysis. *J. Am. Stat. Assoc.*, New York, v.70, n.352, p.892-898, 1975.

FISHER, R. A. The use of multiple measures in taxonomic problems, *Ann. Eugenics*, Cambridge, v.7, n.2, p.179-188, 1936.

FUKUNAGA, K. *Introduction to statistical pattern recognition*. 2. ed. Boston: Academic Press, 1990.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. *Digital image processing using Matlab,* Prentice Hall, 2004.

HOSMER, D.; LEMESHOW, S. *Applied logistic regression.* New York: Wiley, 1989.

JAIN, A K.; DUIN R. P.; MAO, J. Statistical pattern recognition: a review*, IEEE Trans. Pattern. Anal. Mach. Intell.*, New York, v.22, n.1, p.04–37, 2000.

LANDIS, J. R; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics,* Washington*,* v. 33, p. 159-174, 1977.

McLACHLAN, G. J. *Discriminant analysis and statistical pattern recognition*. New York: Wiley, 1992.

MORAES, D. A. O.; HAERTEL, V. Estabilidade de classificadores de decisão em árvore binária para dados imagem em alta dimensão, *Rev. Inf. Teor. Apl.*, Porto Alegre, v.15, n.2, p.27-42, 2008.

QI CHENG; VARSHNEY, P. K.; ARORA, M. K.; Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Trans. Geosci. Remote Sens. Lett.* New York, v.3, n.4, p.491–494, 2006.

UCI Machine Learning Repository: University of California, School of Information and Computer Science, 2007, Irvine, CA. Disponível em: <http://www.ics.uci.edu/~mlearn/MLRepository.html>. Acesso em: 20 out 2007.

Recebido em 24.10.2008.

Aprovado após revisão 22.03.2009.