September 24-28, 2012
Rio de Janeiro, Brazil

Congreso Latino-Iberoamericano
de Investigación Operativa

Simpósio Brasileiro
de Pesquisa Operacional

USING PRINCIPAL COMPONENTS FOR ESTIMATING POLYTOMOUS QUADRATIC LOGISTIC REGRESSION

INÁCIO ANDRUSKI-GUIMARÃES

DAMAT – DEPARTAMENTO ACADÊMICO DE MATEMÁTICA

UTFPR – UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

Av. Sete de Setembro, 3165 CEP 80230-901 Curitiba Paraná Brasil

andruski@utfpr.edu.br

ABSTRACT

Many papers on logistic regression have only considered the logistic regression model with linear discriminant functions, but there are situations where quadratic discriminant functions are useful, and works better. However, the quadratic logistic regression model involves the estimation of a great number of unknown parameters, and this leads to computational difficulties when there are a great number of independent variables. This paper proposes to use a set of principal components of the explanatory variables, in order to reduce the dimensions in the problem, with continuous independent variables, and the computational costs for the parameter estimation in polytomous quadratic logistic regression, without loss of accuracy. Examples on datasets taken from the literature show that the quadratic logistic regression model, with principal components, is feasible and, generally, works better than the classical logistic regression model with linear discriminant functions, in terms of correct classification rates.

KEYWORDS. Polytomous Logistic Regression, Quadratic Logistic Regression, Principal Components Analysis, Polytomous Response.

Main area EST - Statistics

1. Introduction

Is well known that the logistic regression model are a powerful method widely applied for modeling the relationship between a categorical - or ordinal - dependent variable and a set of explanatory variables, or covariates, both continuous or discrete. One advantage of using logistic model for discriminant analysis, rather than Linear Discriminant Analysis, is that it is relatively robust. Furthermore the accuracy of the Classical Logistic Regression Model has been reported in many studies involving bankruptcy prediction and cancer classification, among others applications. Previous papers on logistic regression have only considered the logistic regression model with linear discriminant functions, but there are situations where quadratic discriminant functions are useful, and works better. However, the quadratic logistic regression model involves the estimation of a great number of unknown parameters, and this leads to computational difficulties when there are a great number of independent variables. Furthermore, a great number of parameters should be avoided, because of the risk of over-fitting. Another problem is that, while the logistic regression model work well for many situations, may not work when the data set has no overlapping. Moreover, the logistic model becomes unstable when there is dependence, or multicollinearity, between the explanatory variables. In this paper we propose to use a set of principal components of the explanatory variables, in order to reduce the dimensions in the problem, with continuous independent variables, and the computational costs for the parameter estimation in polytomous quadratic logistic regression, without loss of accuracy.

2. Classical Logistic Regression Model

Let us consider a sample of n independent observations, available from the groups G_1, G_2, ..., G_s. Let $\underline{\mathbf{x}}$ the vector of explanatory variables, $\underline{\mathbf{x}}^T = [x_0, x_1, \ldots, x_p]$, where $x_0 \equiv 1$, for convenience. Let Y denote the polytomous dependent variable with s possible outcomes. We will summarize the n observations in a matrix form given by:

$$X = \begin{bmatrix} 1 & \ldots & x_{p1} \\ \ldots & \ldots & \ldots \\ 1 & \ldots & x_{pn} \end{bmatrix}$$

The Classical Logistic Regression (CLR) model assumes that the posterior probabilities have the form:

$$P(G_k|\underline{\mathbf{x}}) = \frac{exp\left(\beta_{k0} + \sum_{j=1}^{p} \beta_{kj} x_j\right)}{\sum_{i=1}^{s} exp\left(\beta_{i0} + \sum_{j=1}^{p} \beta_{ij} x_j\right)}$$

where $k = 1, 2, \ldots, s-1$, and $\underline{\mathbf{B}}_s = \mathbf{0}$. In this paper the group s is called reference group. The model involves $(s-1)(p+1)$ unknown parameters. The conditional likelihood function is:

$$L(\underline{B}|Y,\underline{\mathbf{x}}) = \prod_{i=1}^{n} \prod_{k=1}^{s} \left[P(G_k|\underline{\mathbf{x}}_i)\right]^{Y_{ki}}$$

where $Y = [Y_1, \ldots, Y_n]^T$ and $Y_i = (Y_{1i}, \ldots, Y_{si})$, with $Y_{ki} = 1$, if $Y = k$, and $Y_{ki} = 0$, otherwise. Taking the logarithm, the log-likelihood function is

$$L(\underline{B}|Y,\underline{\mathbf{x}}) = \sum_{i=1}^{n} \sum_{k=1}^{s} Y_{ki} ln\left[P(G_k|\underline{\mathbf{x}}_i)\right]$$

The Maximum Likelihood Estimator (MLE) is the maximizer of the log-likelihood function in

relation to **B**. The mostly used iterative method is the Newton-Raphson. In practice, the estimation of unknown parameters is affected by the data's properties. Albert and Anderson (1984) suggested a sample classification into three categories: complete separation, quasi-complete separation and overlap. They also proved that the MLE do not exist for complete and quasi-complete separation. Different approaches to deal with separation can be found in Heinze and Schemper (2002), Rousseeuw and Christmann (2003), for binary response, and Andruski-Guimarães and Chaves-Neto (2009), for polytomous response.

Our approach to solve the multiple group problem, when there are complete separation, is to provide a simple and direct generalization of the Hidden Logistic Regression Model (HLR), a robust estimation method presented by Rousseeuw and Christmann (2003). We consider n unobservable independent variables $T_1, \dots, T_n$, where each T_i has s possible values, $\gamma_1, \dots, \gamma_s$. Thus, we observe $Y_i = j$ with a $P(Y_i = j | T_i = \gamma_k) = \delta_{jk}$ probability, where $\sum_{j=1}^{s} \delta_{jk} = 1$ and $\delta_{jj} = max_{k=1,\dots,s}\{\delta_{jk}\}$. The maximum likelihood estimator for T_i, if $Y_i = j$, is $\hat{T}_i = \gamma_j$. In a model with n responses y_{ij} ($i = 1, \dots, n$; $j = 1, \dots, s$) where $y_{ij} = 1$, if $Y_i = j$, and $y_{ij} = 0$, otherwise, we can define the variable given by $\tilde{y}_{ij} = \sum_{k=1}^{s} y_{ik}\delta_{kj}$. Let us keep in mind that in the CLR model, $\delta_{jj} = 1$ and $\delta_{jk} = 0$, if $j \neq k$. The purpose is to maximize:

$$L(\underline{\Theta}|\tilde{Y},\underline{x}) = \prod_{i=1}^{n}\prod_{j=1}^{s}\left[P(T_j|\underline{x}_i)\right]^{\tilde{y}_{ji}}$$

The log-likelihood function becomes:

$$l(\underline{\Theta}|\tilde{Y},\underline{x}) = \sum_{i=1}^{n}\left[\sum_{j=1}^{s-1}\tilde{y}_{ji}\gamma_j - ln\left(1 + \sum_{j=1}^{s-1}exp(\mu_j)\right)\right]$$

where $\mu_j = \theta_{j0} + \theta_{j1}x_1 + \cdots + \theta_{jp}x_p$, $j = 1, 2, \dots, s-1$.

The maximum likelihood estimators are the maximizers of the log-likelihood function, which is strictly concave. For more details about the maximization of log-likelihood function, the interested reader is referred to Andruski-Guimarães and Chaves-Neto (2009). In related literature different approaches to implement robust estimation methods are given by Kodzarkhia et al. (2001), Hubert and Van Driessen (2004) and Gervini (2005), to name just a few.

According to Rousseeuw and Christmann (2003), Copas (1988) found that accurate the estimation of δ_0 and δ_1 in the binary case is very difficult, unless n is extremely large. For a detailed explanation, and discussion, see Copas (1988), Hubert and Van Driessen (2004) and Rousseeuw and Christmann (2003). In this paper, we consider that the probability of observing the true status, which is given by:

$$P(Y_i = j | T_i = \mu_j) = \delta_{jj} ,$$

should be higher than 0.5, this is, $0.5 < \delta_{jj} < 1$, furthermore $\sum_{k=1,k\neq j}^{s} \delta_{jk} < \delta_{jj}$. Therefore, we cannot take the estimate given by $\bar{\pi}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$, $j = 1, \dots, s$, once $\bar{\pi}_j$ can be smaller than 0.5. Our default choice will be $\delta = 0.99$, and set $\delta_{jj} = \delta$ and $\delta_{jk} = \frac{1-\delta}{s-1}$.

3. Quadratic Logistic Regression Model

An extension of the CLR model is to include quadratic and multiplicative interaction terms. The Quadratic Logistic Regression (QLR) Model can be given by:

$$Q(G_k|\underline{\mathbf{x}}) = \frac{exp\left(\underline{\chi_k}\right)}{\sum_{i=1}^{s} exp\left(\underline{\chi_i}\right)}$$

where $\underline{\chi_k} = \alpha_{k0} + \sum_{i=1}^{p} \alpha_{ki}x_i^2 + \sum_{i=p+1}^{\binom{p}{2}} \alpha_{ki}x_{j'}x_{j''} + \sum_{i=\binom{p}{2}}^{\binom{p}{2}+p} \alpha_{ki}x_j$, $k = 1$, 2 , $\dots$, $s-1$, $\chi_s = \mathbf{0}$ and j , $j'' = 1$, $\dots$, p , $j' = 1$, 2 , $\dots$, $p-1$.

The model involves $[(s-1)(p+1)]\left(1 + \frac{p}{2}\right)$ unknown parameters and the estimation of these parameters follows the same lines as that taken by the Classical Logistic Regression Model (CLR). However, for a large number of independent variables, the number of extra parameters can be render an unworkable problem, so that a reduction dimension method can be useful to way out of this problem. Furthermore, a large number of parameters should be avoided, because of the risk of over-fitting. As pointed out by Anderson (1975), the quadratic term also can be written as:

$$\underline{\chi_k} = \alpha_{k0} + \underline{\mathbf{x}}^T \Omega_k \underline{\mathbf{x}} + \alpha_k^T \underline{\mathbf{x}}$$

where $\Omega_k = \mathbf{V}_k^{-1} - \mathbf{V}_s^{-1}$, and $\mathbf{V}_k$ is the dispersion matrix in G_k , $k = 1$, $\dots$, $s - 1$. An approximation, proposed by Anderson (1975), gives a quadratic term with a reduced number of parameters. This approximation is given by the spectral decomposition:

$$\Omega_k = \sum_{j=1}^{p} \lambda_{jk} l_{jk} l_{jk}^T$$

where the λ_{jk} are the eigenvalues of Ω_k , in decreasing size, $\lambda_{1k} \geq \lambda_{2k} \geq \cdots \geq \lambda_{pk}$, and l_{jk} are the corresponding eigenvectors. In this case, Ω_k can be given by:

$$\Omega_k \cong \lambda_k l_k l_k^T$$

In the sequence, each $l_j^T = (l_{j1}, \dots, l_{jp})$ is normed with the constraints:

$$\sum_{k=1}^{p} l_{jk}^2 = 1$$

Since this approach is not always convenient for computing, an alternative parameterization is suggested:

$$\underline{\chi_k} = \alpha_{k0} + \mu_k (d_k^T \underline{\mathbf{x}})^2 + \alpha_k^T \underline{\mathbf{x}} ,$$

where $\mu_k = sgn(\lambda_k), k = 1$, $\dots$, $s-1, d_{kj} = \frac{l_{kj}}{\sqrt{|\lambda_k|}}$, $j = 1$, $\dots$, p.

The log-likelihood function is maximized with respect to the α_{kj} and d_{kj} unrestrictedly $2^{(s-1)}$ times for $\mu_k = \pm 1$ and to take as maximum likelihood estimates those values of the parameters which give the greatest of these $2^{(s-1)}$ values of the log-likelihood function. With this approximation, there are $(s-1)$ unknown parameters. However, this approach is not always applicable. If the independent variables are binary, the diagonal terms of Ω are zero. In this paper

we propose to use as covariates the principal components of the $(s-1)(p+1)$ matrix $I(\underline{\mathbf{B}})$ whose elements are given by:

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_{jm} \partial \beta_{jm'}} = -\sum_{i=1}^{m} x_{m'i} x_{mi} [Q(G_j|\underline{\mathbf{x}})][1 - Q(G_j|\underline{\mathbf{x}})]$$

and

$$\frac{\partial^2 L(\mathbf{B})}{\partial \beta_{jm} \partial \beta_{j'm'}} = -\sum_{i=1}^{m} x_{m'i} x_{mi} [Q(G_j|\underline{\mathbf{x}})][Q(G_{j'}|\underline{\mathbf{x}})]$$

where j, $j' = 1, 2, \ldots, (s-1)$ and $m, m' = 1, \ldots, p$.

4. Principal Components Analysis

The Principal Components Analysis (PCA) is a method to explaining the variance and covariance structure through linear combinations of the variates and may be considered a tool for reducing the dimensionality of the data, as well the multicollinearity among the independent variables.

Let us consider n observations of p continuous variables, given by the matrix $\mathbf{X}$, and let the sample covariance matrix:

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \cdots & \ddots & \cdots \\ s_{p1} & \cdots & s_{pp} \end{bmatrix}$$

The observations $\underline{\mathbf{x}}$ can be standardized, so that

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}.$$

The matrix $\mathbf{S}$ can be written as $\mathbf{S} = \mathbf{V}^T \mathbf{\Lambda} \mathbf{V}$, where $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ being orthogonal. Let $\mathbf{Z}$ the matrix whose columns are the principal components, given by $\mathbf{Z} = \mathbf{X}\mathbf{V}$, where $\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_p$ are the eigenvectors of the matrix $\mathbf{S}$, associated to the eigenvalues $\lambda_1, \ldots, \lambda_p$, so that the matrix of observations can be written as $\mathbf{X} = \mathbf{Z}\mathbf{V}^T$, where

$$x_{ij} = \sum_{k=1}^{p} z_{ik} v_{jk}$$

Furthermore, matrices $\mathbf{Z}$ and $\mathbf{V}$ also can be written as:

$$\mathbf{Z} = \begin{bmatrix} 1 & z_{11} & \cdots & z_{1(q+1)} & \cdots & z_{1p} \\ 1 & z_{21} & \cdots & z_{2(q+1)} & \cdots & z_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & z_{n1} & \cdots & z_{n(q+1)} & \cdots & z_{np} \end{bmatrix} = \left(\mathbf{Z}_{(q)} | \mathbf{Z}_{(r)}\right)$$

and

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ 1 & v_{11} & \cdots & v_{1(q+1)} & \cdots & v_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & v_{p1} & \cdots & v_{p(q+1)} & \cdots & v_{pp} \end{bmatrix} = \left(\mathbf{V}_{(q)} | \mathbf{V}_{(r)}\right)$$

In order to improve the parameter estimation under multicollinearity, and to reduce the dimension of the problem, Aguilera, Escabias and Valderrama (2006) propose to use as covariates of the logistic regression model a reduced set of optimum principal components of the original covariates. This approach, called Principal Component Logistic Regression (PCLR) model, provide an accurate estimation of the parameters in the case of multicollinearity. Furthermore, cf. Barker and Brown (2001), estimates obtained via principal components can have smaller mean square error than estimates obtained through standard logistic regression. Although, it is well known, cf. McLachlan (2004), that the estimates of the eigenvalues of $\mathbf{S}$ are biased. This bias is most pronounced when the eigenvalues of $\mathbf{S}$ tend toward equality, being less severe when they are highly disparate.

The generalization of the PCLR model for polytomous responses does not require a complex formulation. We begin by computing the covariance matrix $\mathbf{S}$. Then the matrix $\mathbf{X}$ can be written as:

$$x_{ik} = \sum_{j=1}^{p} z_{ij} v_{kj}$$

so that

$$P\big(G_t \big| \mathbf{Z}\underline{\mathbf{v}}_i\big) = \frac{exp\big(\beta_{t0} + \sum_{k=1}^{p}\sum_{j=1}^{p} z_{ij} v_{kj}\beta_{tk}\big)}{\sum_{m=1}^{s} exp\big(\beta_{m0} + \sum_{k=1}^{p}\sum_{j=1}^{p} z_{ij} v_{kj}\beta_{mk0}\big)}$$

where $i = 1, \ldots, s$; $j = 0, \ldots, p$; $t = 1, \ldots, s$ and $\beta_{sj} = 0$.

Making $\gamma_{tj} = \sum_{k=1}^{p} v_{kj}\beta_{tk}$, the PCLR model, with linear discriminant functions, extended to polytomous responses is given by:

$$\big(G_t \big| \mathbf{Z}\underline{\mathbf{v}}_i\big) = \frac{exp\big(\beta_{t0} + \sum_{j=1}^{p} z_{ij}\gamma_{tj}\big)}{\sum_{i=1}^{s} exp\big(\beta_{i0} + \sum_{j=1}^{p} z_{ij}\gamma_{mj}\big)}$$

The Principal Components Quadratic Logistic Regression (PCQLR) is given by:

$$Q\big(G_k \big| \mathbf{Z}\underline{\mathbf{v}}_i\big) = \frac{exp\big(\underline{\chi}_k\big)}{\sum_{i=1}^{s} exp\big(\underline{\chi}_i\big)}$$

where $\underline{\chi}_k = \chi_{k0} + \sum_{i=1}^{p} z_{ij}\gamma_{kj}^2 + \sum_{i=p+1}^{\binom{p}{2}} z_{ij}\gamma_{kj'}\gamma_{kj''} + \sum_{i=\binom{p}{2}+1}^{\binom{p}{2}+p} z_{ij}\gamma_{kj}$, $k = 1, \ldots, s-1$; $\underline{\chi}_s = \mathbf{0}$ and $j, j'' = 1, \ldots, p$; $j' = 1, \ldots, p-1$.

In order to estimate the principal components model's parameters, one can apply the Maximum Likelihood Method. In the dichotomous case, Aguilera, Escabias and Valderrama (2006) also proposes two methods to solve the problem of choosing the optimum principal components that should be included in the model. However, the interested reader should to take in mind that, according Jolliffe (1982), principal components with small eigenvalues can be as important as those with large eigenvalues.

In this paper the purpose is only to investigate the principal components model's classificatory performance in polytomous cases, using linear and quadratic forms, for practical purposes. In

order to formulate the model, the first step was to obtain the principal components of the covariates. We have used the first q principal components, with the largest variances, including principal components in the natural order, given by the explained variability. In the sequence, we fitted the quadratic logistic model, using the selected principal components as covariates. With respect to the QLR model, we propose to use as covariates the principal components of the $[(s-1)(p+1)]\left(1+\frac{p}{2}\right)$ matrix $\mathbf{I}(\chi)$.

5. Applications

In this section we consider two benchmark data sets, taken from the trade literature. Iris Data, taken from Fisher (1936), and Fatty Acid Composition Data, taken from Brodnjak-Vončina et al. (2005). We have applied the CLR model, PCLR model, QLR model and PCQLR model to both data sets. The purpose is to compare the results provided by the four models, given by the Correct Classification Rate (CCR), defined as the percentage of observations that are correctly classified. A computer program which implements the approaches described previously has been written in *Visual Basic 6.0* and runs on the *HP Pavillion b1040br* computer. The results achieved, in terms of performance, are given in the sequence.

Example 1: Iris Data. There are three groups: Iris Setosa (G_1), Iris Versicolor (G_2) and Iris Virginica (G_3). For each group there are 50 observations and four independent variables: Sepal Length, Sepal Width, Petal Length and Petal Width, all measured in *mm*. The reference group is Iris Virginica. Is well known that two groups, Iris Versicolor and Iris Virginica, overlap and form a cluster completely separated from Iris Setosa. The correct classification rates for HLR and PCLR models are summarized in Table 1. Table 2 displays the principal components and their cumulative percentage of the total variance. The correct classification rates for QLR and PCQLR models are summarized in Table 3. From Table 3, we can conclude that both models have high classification capability in terms of the correct classification rates.

Table 1. Classification Matrix. Iris data. Linear Discriminant Functions.

Model	Observed Group	Allocated Group G_1	G_2	G_3
HLR	G_1	1.00	0.00	0.00
	G_2	0.00	0.98	0.02
	G_3	0.00	0.02	0.98
PCLR (1 p.c.)	G_1	1.00	0.00	0.00
	G_2	0.00	0.88	0.12
	G_3	0.00	0.10	0.90

Table 2. Iris data. Variances (eigenvalues)

Variance (λ)	2.9185	0.9140	0.1468	0.0207
Cumulative Percentage of Total Variance	72.96	95.81	99.48	100

Table 3. Classification Matrix. Iris data. Quadratic Discriminant Functions.

Model	Observed Group	Allocated Group G_1	G_2	G_3
QLR	G_1	1.00	0.00	0.00
	G_2	0.00	0.98	0.02
	G_3	0.00	0.02	0.98
PCQLR (3 p.c.)	G_1	1.00	0.00	0.00
	G_2	0.00	0.98	0.02
	G_3	0.00	0.02	0.98

Congreso Latino-Iberoamericano
de Investigación Operativa
Simpósio Brasileiro
de Pesquisa Operacional

September 24-28, 2012
Rio de Janeiro, Brazil

Example 2: Fatty Acid Data. There are 120 observations, five groups and seven variables, representing the percentage levels of seven fatty acids, namely palmitic (x_1), stearic (x_2), oleic (x_3), linoleic (x_4), linolenic (x_5), eicosanoic (x_6) and eicosenoic (x_7) acids. In this paper we consider five groups: rapeseed (G_1), sunflower (G_2), peanut (G_3), corn (G_4) and pumpkin (G_5) oils. In this paper the reference group is (G_5) (pumpkin oil). The original data set have eight groups, and the complete table of the original data can be found in Brodnjak-Vončina et al. (2005). There are a high correlation between oleic and linoleic acids ($r = -0.9565$). Table 4 displays the classification matrix for the HLR and PCLR models. Table 5 displays the principal components and their cumulative percentage of the total variance. Table 6 displays the classification matrix for the QLR and PCQLR models.

Table 4. Classification Matrix. Fatty acid data. Linear Discriminant.

Model	Observed Group	Allocated Group				
		G_1	G_2	G_3	G_4	G_5
HLR	G_1	0.64	0.00	0.00	0.00	0.36
	G_2	0.00	0.95	0.00	0.00	0.05
	G_3	0.00	0.00	1.00	0.00	0.00
	G_4	0.00	0.00	0.00	1.00	0.00
	G_5	0.15	0.00	0.05	0.05	0.75
PCLR (6 p.c.)	G_1	0.64	0.00	0.00	0.00	0.36
	G_2	0.00	0.95	0.00	0.00	0.05
	G_3	0.00	0.00	0.96	0.00	0.04
	G_4	0.00	0.00	0.00	0.80	0.20
	G_5	0.17	0.06	0.03	0.06	0.68

Table 5. Fatty acid data. Variances (eigenvalues)

Variance (λ)	3.9092	1.0842	0.9325	0.7866	0.2053	0.0811	0.0001
% of Total Variance	55.85	71.84	84.66	95.90	98.83	99.99	100

Table 6. Classification Matrix. Fatty acid data. Quadratic Discriminant.

Model	Observed Group	Allocated Group				
		G_1	G_2	G_3	G_4	G_5
QLR	G_1	0.82	0.00	0.00	0.00	0.18
	G_2	0.00	1.00	0.00	0.00	0.00
	G_3	0.00	0.00	1.00	0.00	0.00
	G_4	0.00	0.00	0.00	1.00	0.00
	G_5	0.00	0.00	0.00	0.00	1.00
PCQLR (6 p.c.)	G_1	0.73	0.00	0.00	0.00	0.27
	G_2	0.00	1.00	0.00	0.00	0.00
	G_3	0.00	0.00	1.00	0.00	0.00
	G_4	0.00	0.00	0.00	0.90	0.10
	G_5	0.00	0.03	0.00	0.05	0.92

6. Conclusion

The purpose with this job is to develop and implement a simple and direct generalization for the Quadratic Logistic Regression Model, for polytomous response, which allows the reduction of the dimensions in the problem, and to explore the performance of the model when compared to the Classical Logistic Regression model with linear discriminant functions. In order to solve the problem that arises with the great number of unknown parameters, we have used the Principal Components Analysis (PCA), as well a generalization of the Hidden Logistic Regression Model, to deal with the complete separation. We can see that the PCA allows the reduction of the number of dimensions in a polytomous Quadratic Logistic Regression Model,

with continuous variables and avoiding the multicollinearity of these variables. For practical purposes, the main advantage of the HLR model is the existence and uniqueness of estimators. Furthermore, there are not computational difficulties to implement the referred approaches. With respect to the performance, we can see that the Quadratic Logistic Regression Model can provide better classification rates than the Classical Logistic Regression Model.

In the future we intend to study the behaviour of the models that were approached with respect to aspects such as their performance regarding data sets with a reduced number of observations and the bias of the estimators that were obtained.

References

Aguilera, A. M., Escabias, M., Valderrama, M. J. (2006), Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis* 55, 1905-1924.

Albert, A. and Anderson, J.A. (1984), On the existence of maximum likelihood estimates in logistic regression models, *Biometrika*, 71, pp. 1-10.

Anderson, J. A. (1975)**,** Quadratic logistic discrimination, *Biometrika*, 62, pp. 149-154.

Andruski-Guimarães, I. and Chaves-Neto, A. (2009), Estimation in polytomous logistic model: comparison of methods, *Journal of Industrial and Management Optimization*, 5, pp. 239-252.

Barker, L. and Brown, C. (2001)**,** Logistic regression when binary predictor variables are highly correlated, *Statistics in Medicine*, 20 (9-10), pp. 1431-1442.

Brodnjak- Vončina, D., Kodba, Z.C. and Novic,C. (2005), Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. *Chemometrics and Intelligent Laboratory Systems*, 75, pp. 31-43.

Copas, J. B. (1988), Binary regression models for contaminated data. With discussion. *Journal of Royal Statistical Society B*, 50, 225-265.

Fisher, R. A. (1936), The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 3, pp. 179-188.

Gervini, D. (2005), Robust adaptive estimators for binary regression models. *Journal of Statistical Planning and Inference,* 131, 297-311.

Heinze, G. and Schemper, M. (2002), A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21, 2409-2419.

Hubert, M. and van Driessen, K. (2004), Fast and robust discriminant analysis. *Computational Statistics & Data Analysis*, 45, 301-320.

Jolliffe, I. T. (1982), A note on the use of principal components in regression. *Applied Statistics*, 31, 3, 300-303.

Kodzarkhia, N., Mishra, G. D. and Reiersolmoen, L. (2004), Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference*, 98, 211-223.

McLachlan, G. J., *Discriminant Analysis and Statistical Pattern Recognition*, p. 130. John Wiley & Sons, Inc., Hoboken, New Jersey, U.S.A. (2004)

Rousseeuw, P. J. and Christmann, A. (2003), Robustness against separation and outliers in logistic regression, *Computational Statistics & Data Analysis* 43, pp. 315-332.