ESTIMATION IN POLYTOMOUS LOGISTIC MODEL: COMPARISON OF METHODS

INÁCIO ANDRUSKI-GUIMARÃES
Departamento Acadêmico de Matemática
UTFPR - Universidade Tecnológica Federal do Paraná
Rua Sete de Setembro, 3165
80230-901 Curitiba Paraná, Brasil

ANSELMO CHAVES-NETO
Departamento de Estatística
UFPR - Universidade Federal do Paraná
Centro Politécnico Jardim das Américas Caixa Postal 19081
81531-980 Curitiba Paraná, Brasil

(Communicated by Andrés Medaglia)

Abstract. The logistic regression model is a powerful method for modeling the relationship between a categorical variable and a set of explanatory variables. In practice, however, the existence of maximum likelihood estimates is known to be dependent on the data configuration. In fact, the Maximum Likelihood Estimators (MLE) of unknown parameters exists if, and only if, there is data overlapping. The Hidden Logistic Regression (HLR) is an alternative model under which the observed response is related to the unobservable response. The Maximum Estimated Likelihood (MEL) method is also proposed, once it is immune to the complete or quasi-complete separation of data. The Principal Component Logistic Regression (PCLR) model is useful to reduce the number of dimensions of a logistic regression model with continuous covariates avoiding multicollinearity. In this paper we present an extension of the HLR and PCLR models as means for the solution of problems with polytomous responses. The purpose is to propose an alternative approach for the parameter estimation problem in polytomous logistic models when the data groups are completely separated. Simulations results resulting from databases taken from the literature show that the proposed approach is feasible.

1. Introduction. The logistic regression model is a powerful method for modeling the relationship between a categorical - or ordinal - dependent variable and a set of explanatory variables, involving both discrete and continuous variables. However the estimation of the logistic model parameters can be influenced by peculiarities in the data. Albert and Anderson [2] suggested a data classification into three categories: overlap, quasi-complete separation and complete separation, and proved

2000 Mathematics Subject Classification. Primary: 62H30; Secondary: 62J02, 68T10.
Key words and phrases. Polytomous Logistic Regression, Robust Estimation, Principal Component Analysis.
that the Maximum Likelihood Estimators (MLE) of unknown parameters exist if, and only if, there is data overlapping. Moreover, the estimation of parameters may not be very accurate when there is multicollinearity among the predictors. Another problem can be the large number of independent variables, or covariates, that is needed to explain the response, so a dimension reduction method may have to be used.

An alternative model, presented by [18], assumes that due to an additional stochastic mechanism the true response of a logistic regression model is unobservable, although there is an observable variable that is related to this response. Under this model, the true unobservable response is comparable to a hidden layer in a feed-forward neural network and hence the model is called the Hidden Logistic Regression (HLR) model. In the same paper those authors proposed the Maximum Estimated Likelihood (MEL) technique, and showed that it is immune to quasi-complete, or complete, separation. The MEL estimator is studied on data sets with dichotomous, or binary, response variables. In this paper we apply the HLR model to data with polytomous, or multinomial, response and compare the performance with those from the Classical Logistic regression (CLR) model and the Individualized Logistic Regression (ILR) model.

Another alternative method, called the Principal Component Logistic Regression (PCLR) model and presented by [1], proposes the use of a set of principal components of the predictors variables as covariates of the logistic model. The purpose is to improve the estimation of parameters under collinearity and to reduce the number of dimensions in the problem by means of continuous independent variables. In this paper we extend the approach [1] proposed for dichotomous responses to polytomous response, in order to investigate the performance of the PCLR model in multiple group cases.

This paper is organized as follows. Sections 2 and 3 gives an overview of the CLR and IRL models. Section 4 presents an overview of the HLR model for binary responses and proposes an extension to data sets with polytomous response. Section 5 has an overview of the PCLR model, as well as a generalization of the PCLR model for the polytomous response case. In Section 6 we apply the extended HLR and PCLR models on data sets taken from the literature and compare their performance with those from the CLR and ILR models. Section 7 gives a brief conclusion and makes suggestions for future studies.

2. **Polytomous logistic regression model.** We consider that a sample of \( n \) independent observations is available from the groups \( G_1, G_2, \ldots, G_s \). Let \( \mathbf{X} \) the vector of explanatory variables, \( \mathbf{X}^T = (x_0, x_1, \ldots, x_p) \), where \( x_0 = 1 \), for convenience. Let \( Y \) denote the polytomous dependent variable with \( s \) possible outcomes. Let us consider a set of \( n \) observations that we will summarize in a matrix form given by:

\[
\mathbf{X} = \begin{bmatrix}
1 & x_{11} & \cdots & x_{p1} \\
1 & x_{12} & \cdots & x_{p2} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{1n} & \cdots & x_{pn}
\end{bmatrix}
\]

or, equivalently,

\[
\mathbf{X} = \begin{bmatrix}
1 & x_1 & \cdots & x_p
\end{bmatrix}.
\]
2.1. **Classical logistic regression model.** The Classical Logistic Regression (CLR) model assumes that the posterior probabilities have the form:

\[
P_G k | \mathbf{X} = \frac{\exp \left( \beta k_0 + \sum_{j=1}^{p} \beta kj x_j \right)}{\sum_{i=1}^{s} \exp \left( \beta i_0 + \sum_{j=1}^{p} \beta ij x_j \right)}
\]

where \( k = 1, 2, \ldots, s - 1 \) and \( B_s = 0 \). In this paper the group \( s \) is called reference group. The model involves \( (s - 1)(p + 1) \) unknown parameters. The conditional likelihood function is given by:

\[
\ell (B | Y, \mathbf{X}) = \prod_{i=1}^{n} \prod_{k=1}^{s} [P (G_k | \mathbf{X}_i)]^{y_{ki}}
\]

where \( Y = (Y_1, \ldots, Y_n)^T \) and \( Y_i = (Y_{i1}, \ldots, Y_{is}) \), with \( y_{ki} = 1 \) if \( Y = k \), and \( y_{ki} = 0 \) otherwise. Taking the logarithm, the log-likelihood function is:

\[
L (B | Y, \mathbf{X}) = \sum_{i=1}^{n} \sum_{k=1}^{s} y_{ki} \ln [P (G_k | \mathbf{X}_i)]
\]

Thus:

\[
\frac{\partial}{\partial \beta kj} L (B | Y, \mathbf{X}) = \sum_{i=1}^{n} x_{ij} (y_{ki} - P (G_k | \mathbf{X}_i))
\]

The Maximum Likelihood Estimator (MLE) \( \hat{B} \) is obtained by setting the derivatives to zero and solving for \( B \). The mostly used iterative method is the Newton-Raphson. However, in practice, the estimation of unknown parameters is affected by the data’s properties. Albert and Anderson [2] 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. An algorithm to distinguish separation from multicollinearity is proposed by [15]. Another important reference is [11]. They argued that, in practice, monitoring the variance in the iteration process is sufficient to declare the separation.

3. **Individualized logistic regression model.** For the polynominal logistic regression [4] proposed an alternative method for the maximum likelihood estimation, called Individualized Logistic Regression (ILR) model. For a response variable with \( s \) groups, they suggested fitting \( s - 1 \) binary logistic regressions and, each time, comparing a group \( d \), \( d = 1, \ldots, s - 1 \), with the reference group. The coefficients for the polynominal logistic model are obtained from the \( (s - 1) \) separately fit logistic models. According the authors, the estimators that are obtained are consistent and will be approximately those from the CLR model. The criterion to evaluate the performance of estimators is the ARE, which is defined as the ratio of the estimators’ asymptotic variance. The authors showed that the estimates loss in ARE is small. According [12], the individualizing fit can be useful to select variables. By using this
method, a series of \((s - 1)\) conditional likelihood are separately maximized, where each likelihood is given by:

\[
L_d = \prod_{C_d} \theta_{dt}^{Y_{dt}} (1 - \theta_{dt})^{Y_{st}}
\]

where \(\theta_{dt} = P(Y_{dt} = 1 \mid x_i, Y_{si} + Y_{di} = 1)\), \(\theta_{st} = P(Y_{dt} = 0 \mid x_i, Y_{si} + Y_{di} = 1)\), and \(C_d = \{i : Y_{si} + Y_{di} = 1\}\).

The estimates \(\hat{\theta}_d\) are obtained by setting the derivatives to zero and solving for \(\theta\). A modification to the Begg and Greys method is suggested by [17]. Accordingly, the \(s(s - 1)/2\) pairwise binary logistic regressions are fitted and the estimators are obtained by weighted least squares. The modified procedure requires more computations, but it has a higher ARE than Begg and Greys method, according to [17].

4. Hidden logistic regression model. The CLR model assumes observable independent responses \(Y_i\) with Bernoulli distributions. The Hidden Logistic Regression (HLR) model is a robust estimation method presented by [18], and assumes that the true response is unobservable due to a stochastic mechanism. In the classical approach, with dichotomous response, the true response has two possible values, usually denoted as success and failure. The HLR model assumes that the true status \(T_i\), which can be \(s\) (success) or \(f\) (failure), is unobservable. However, there is an observable variable \(Y\) that is related to \(T\). If the true status is \(T = s\), then \(Y = 1\) with a \(P(Y = 1 \mid T = s) = \delta_1\) probability. Hence, the probability of misclassification is \(P(Y = 0 \mid T = s) = 1 - \delta_1\). If \(T = f\), then \(Y = 1\) with a \(P(Y = 1 \mid T = f) = \delta_0\) probability, and \(Y = 0\) with a \(P(Y = 0 \mid T = f) = 1 - \delta_0\) probability. The authors consider that \(0 < \delta_0 < 0.5 < \delta_1 < 1\).

According [18], [7] considered the general case as having \(n\) observations. In the notation Rousseeuw and Christmann’s adopt, there are \(n\) unobservable random independent variables \(T_i\) that result from a CLR model with a finite parameter vector given by \(\theta = (\beta_0, \beta_1, ..., \beta_p)^T\). The \(T_i\) response has a Bernoulli distribution with a success probability of \(\pi_i = \Lambda(x_i^T \theta), \Lambda(z) = [1 + \exp(-z)]^{-1}, x_i \in \mathbb{R}^p\). The \(Y_i\) observable responses are related to \(T_i\) through an entire mechanism that is called hidden logistic regression, because the \(T_i\) true status are hidden by the stochastic structure. The maximum likelihood estimator of \(T\) given \(Y = y\) is \(\hat{T}_{ML} (Y = 0) = f\) and \(\hat{T}_{ML} (Y = 1) = s\).

The conditional probability that \(Y = 1\) given the MLE of \(T\) is:

\[
P\left(Y = 1 \mid \hat{T}_{ML} \right) = \delta_0 \text{ if } y = 0
\]

\[
P\left(Y = 1 \mid \hat{T}_{ML} \right) = \delta_1 \text{ if } y = 1
\]

where \(y\) is the observed value of \(Y\), and denoting the probabilities by \(\hat{Y}, \hat{Y} = (1 - Y) \hat{\delta}_0 + Y \hat{\delta}_1\). In a model with \(n\) observations \(y_i\),

\[
\hat{y}_i = (1 - y_i) \hat{\delta}_0 + y_i \hat{\delta}_1,
\]
where the pseudo-observation $\tilde{y}_i$ is the success probability, conditioned by the estimate of $t_i$. Let us keep in mind that $t_i$ is the true status.

In order to fit a logistic regression to the pseudo-observations we can apply the maximum likelihood formula, given by:

$$L(\theta | \tilde{y}_1, \ldots, \tilde{y}_n) = \prod_{i=1}^{n} \left[ \Lambda (x_i^T \hat{\theta}) \right]^{\tilde{y}_i} \left[ 1 - \Lambda (x_i^T \hat{\theta}) \right]^{1-\tilde{y}_i},$$

which is called the Estimated Likelihood, because the true likelihood is unknown and it is dependent on the unobservable $t_1, \ldots, t_n$. In the Rousseeuw and Christmann terminology, the Estimated Likelihood function maximizer is called the Maximum Estimated Likelihood (MEL) estimator. Moreover, the authors showed that unlike the classical Maximum Likelihood Estimator, the MEL estimator always exists and is unique, even when the data set has no overlap.

4.1. Hidden logistic regression model applied to polytomous case. Our approach to solve the multiple group problem is to provide a simple and direct generalization of the HLR model. We consider $n$ unobservable independent variables $T_1, \ldots, T_n$, where each $T_i$ has $s$ possible values, $\gamma_i, \ldots, \gamma_s$. Thus, we observe $Y_i = j$ with a $P(Y_i = j | T_i = \gamma_k)$ probability, where $\sum_{j=1}^{s} \delta_{jk} = 1$ and $\delta_{jj} = \max_{k=1, \ldots, s} \{ \delta_{jk} \}$. The maximum likelihood estimator for $T_i$, if $Y_i = j$, is $\hat{T}_{ML,i} = \gamma_j$. In a model with $n$ responses $y_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, s$, 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:

$$\ell(\theta | \vec{Y}, \vec{X}) = \prod_{i=1}^{n} \prod_{j=1}^{s} [P(T_j | x_i)]^{\tilde{y}_{ij}}.$$  

The log-likelihood function becomes:

$$L(\theta | \vec{Y}, \vec{X}) = \sum_{i=1}^{n} \left[ \sum_{j=1}^{s-1} \tilde{y}_{ij} \mu_j - \ln \left( 1 + \sum_{j=1}^{s-1} \exp(\mu_j) \right) \right],$$

where $\mu_j = \theta_{j0} + \theta_{j1} x_1 + \theta_{j2} x_2 + \ldots + \theta_{jp} x_p$, $j = 1, 2, \ldots, s - 1.$

The MEL 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 [3]. In related literature different approaches to implement robust estimation methods are given by [10], [13] and [14], to name just a few.

4.2 Choice of $\delta$. According to [18], [6] found that accurate the estimation of $\delta_0$ and $\delta_1$, in the binary case, is very difficult, unless $n$ is extremely large. The symmetric approach is to chose a constant $\gamma > 0$ and to set $\delta_0 = \gamma$ and $\delta_1 = 1 - \gamma$, where $\gamma$ is small so that terms in $\gamma^2$ can be ignored, and $\delta_0 < \hat{\pi} < \delta_1$, where
\( \hat{\pi}, \delta_0 \) and \( \delta_1 \) are given by 
\[ \delta_1 = \frac{1 + \hat{\pi}}{1 + \delta}, \quad \delta_0 = \frac{\hat{\pi} \delta_1}{1 + \hat{\pi}}, \quad \hat{\pi} = \max\{\delta, \min(1 - \delta ; \bar{\pi})\}, \]
\( \bar{\pi} = \frac{1}{n} \sum_{i=1}^{n} y_i \). For a detailed explanation and discussion see [6], [13] and [18]. In this paper, we consider that the probability of observing the true status, which is given by 
\[ P(Y_i = j | T_i = \gamma_j) = \delta_{jj}, \]
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 
\[ \bar{\pi}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij}, \quad j = 1, \ldots, s \], once \( \bar{\pi}_j \) can be smaller than 0.5. Our default choice will be \( \delta = 0.99 \), and set \( \delta_{jj} = \delta \) and \( \delta_{jk} = 1 - \delta_{s-1} \).

5. **Principal component logistic regression.** In order to improve the parameter estimation under multicollinearity and to reduce the dimensions in the problem, [1] propose to use a reduced set of optimum principal components of the original covariates as covariates for the logistic regression model. This approach, which is called Principal Component Logistic Regression (PCLR) model, provides an accurate estimation of the parameters in the case of multicollinearity. Some features in this model, in binary case, are given below. Let us consider \( n \) observations of \( p \) continuous variables, given by the matrix:
\[ X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{12} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \ldots & x_{pn} \end{bmatrix} \]
The sample covariance matrix is given by:
\[ S = \begin{bmatrix} s_{11} & s_{12} & \ldots & s_{1p} \\ s_{21} & s_{22} & \ldots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{bmatrix} \]
And the observations can be standardized, so that:
\[ S = \frac{1}{n-1} X^T X. \]
It is well known that \( S \) can be written as \( S = V^T \Lambda V \), where \( \Lambda = diag(\lambda_1, \ldots, \lambda_p) \) and \( V \) being orthogonal. Let \( Z \) be the matrix whose columns are the principal components, which are given by \( Z = VX \), where \( \mathbf{v}_1, \ldots, \mathbf{v}_p \) are the eigenvectors of the matrix \( S \), associated to the eigenvalues \( \lambda_1 \geq \ldots \geq \lambda_p \), so that the observations matrix can be written as \( X = ZV^T \), where
\[ x_{ij} = \sum_{k=1}^{p} z_{ik} v_{jk}. \]
The binary logistic regression model will then be given by:
\[ P(Y | \mathbf{x}_i) = \frac{\exp\left( \beta_0 + \sum_{j=1}^{p} x_{ij} \beta_j \right)}{1 + \exp\left( \beta_0 + \sum_{j=1}^{p} x_{ij} \beta_j \right)}. \]
And this probability can be expressed as:
\[
P(Y | z_i) = \frac{\exp \left( \beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k \right)}{1 + \exp \left( \beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k \right)},
\]

where \( \gamma_k = \sum_{j=1}^{p} v_{jk} \beta_j \), \( k = 1, \ldots, p \), and \( z_{ik} \) are the elements of \( Z = XV \), \( i = 1, \ldots, n \), \( k = 1, \ldots, p \). Furthermore, matrices \( Z \) and \( V \) can be written as:

\[
Z = \begin{bmatrix}
1 & z_{11} & \cdots & z_{1s} & z_{1(s+1)} & \cdots & z_{1p} \\
1 & z_{21} & \cdots & z_{2s} & z_{2(s+1)} & \cdots & z_{2p} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & z_{n1} & \cdots & z_{ns} & z_{n(s+1)} & \cdots & z_{np}
\end{bmatrix} = (Z_{(s)}|Z_{(r)})
\]

and

\[
V = \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
1 & v_{11} & \cdots & v_{1s} & v_{1(s+1)} & \cdots & v_{1p} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & v_{p1} & \cdots & v_{ps} & v_{p(s+1)} & \cdots & v_{pp}
\end{bmatrix} = (V_{(s)}|V_{(r)})
\]

Considering that \( r = p - s \), the PCLR model in terms of the first \( s \) principal components of matrix \( X \), is given by:

\[
P_{(s)}(Y | Z_{(s)}x_{i}) = \frac{\exp \left( \gamma_0 + \sum_{j=1}^{s} z_{ij} \gamma_j \right)}{1 + \exp \left( \gamma_0 + \sum_{j=1}^{s} z_{ij} \gamma_j \right)}, \ i = 1, \ldots, n.
\]

According to [1], the principal components with the largest variance are not necessarily the more efficient predictors, because principal components with small variances could be correlated with the response variable. There are different methods to include principal components in the regression method and the interested reader should see [9] for a detailed explanation in linear case.

5.1. Principal component logistic regression model extended to polytomous response. The generalization of the PCLR model for polytomous responses does not require a complex formulation. We begin by computing the covariance matrix \( S \). Then the matrix \( X \) can be written as:

\[
x_{ik} = \sum_{j=1}^{p} z_{ij} v_{kj},
\]

so that
\[
\begin{align*}
P(G_i | Z_{Y,i}) &= \frac{\exp \left( \beta_{t0} + \sum_{k=1}^{p} \sum_{j=1}^{p} z_{ij} v_{kj} \beta_{tk} \right)}{\sum_{m=1}^{s} \exp \left( \beta_{m0} + \sum_{k=1}^{p} \sum_{j=1}^{p} z_{ij} v_{kj} \beta_{mk} \right)},
\end{align*}
\]

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 extended to polytomous responses is given by:

\[
\begin{align*}
P(G_i | Z_{Y,i}) &= \frac{\exp \left( \beta_{t0} + \sum_{j=1}^{p} z_{ij} \gamma_{tj} \right)}{\sum_{i=1}^{s} \exp \left( \beta_{i0} + \sum_{j=1}^{p} z_{ij} \gamma_{mj} \right)},
\end{align*}
\]

In order to estimate the principal components model’s parameters, one can apply the Maximum Likelihood Method. In the dichotomous case, \([1]\) also proposes two methods to solve the problem of choosing the optimum principal components that should be included in the model. In this paper the purpose is only to investigate the PCLR model’s classificatory performance in polytomous cases.

6. **Examples.** In this section we consider three benchmark data sets. Iris Data, taken from \([8]\), Fatty Acid Data, taken from \([5]\), and Insulating Oil Data, taken from \([16]\). The purpose is to compare the results provided by the four models. We applied the CLR, ILR, HLR and PCLR models to the data sets. The computer program that implements the approaches previously described was written in Visual Basic 6.0 and was run on a HP Pavilion b1040br computer. The results that were obtained in terms of estimates and performance are described below.

**Example 1: Iris Data.** This data set includes 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. It is well known that two groups (the Iris Versicolor and Iris Virginica ones) overlap and form a cluster that is completely separated from Iris Setosa group. There is no MLE for the CLR model. For the ILR model there is only the MLE \(\tilde{\beta}_{21}\), which corresponds to the Iris Versicolor group. Table 1 displays the estimates and their standard errors, between brackets, obtained by the ILR model and the MEL estimates for the HLR model. One can see that the coefficients that were obtained from the ILR model are close to those from the HLR model. Table 2 displays the principal components and their cumulative percentage of the total variance.
Table 1. Iris data variables. Estimates and standard errors, ILR and HLR.

| Model | Variables     | Function 1                  | Function 2                  |
|-------|---------------|-----------------------------|-----------------------------|
|       |               |                             | -42.6378 (25.7077)          |
| ILR   | Intercept     |                             | 0.2465 (0.2394)             |
|       | Sepal length  | -x-                         | -0.6681 (0.4480)            |
|       | Sepal width   | -x-                         | 0.9429 (0.4737)             |
|       | Petal length  | -x-                         | 1.8286 (0.9743)             |
|       | Petal width   | -x-                         | -42.6378 (25.7077)          |
| HLR   | Intercept     | 39.1964 (94.6970)           | -41.5222 (24.9085)          |
|       | Sepal length  | 0.4553 (2.7904)             | -0.2464 (6.2583)            |
|       | Sepal width   | 1.0465 (2.0433)             | -0.6541 (0.4386)            |
|       | Petal length  | -1.4298 (2.5094)            | 0.9258 (0.4596)             |
|       | Petal width   | -2.2602 (4.7388)            | 1.7886 (0.9474)             |

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 | 100.00 |

Table 3 displays the estimates obtained by the PCLR model with the first principal component. The PCLR model with two, three or four principal components does not exist.

Table 3. Iris data. PCLR 1st. p.c., estimate and standard error.

| Function | Principal component | MLE ( SE ) |
|----------|---------------------|------------|
| 1        | Intercept           | 65.2140 (26.1361) |
|          | 1st. P.C.           | -1.3624 (0.6627)  |
| 2        | Intercept           | 34.5452 (8.1777)  |
|          | 1st. P.C.           | -0.5528 (0.1309)  |

The correct classification rates for HLR and PCLR models are summarized in Table 4. From Table 4, we can conclude that both models have high classification capability in terms of the correct classification rates.

Table 4. Classification matrix. Iris data.

| Model  | Observed group | Allocated Group |
|--------|----------------|-----------------|
|        | G 1 | G 2 | G 3 |
| PCLR   | 1.00 | 0.00 | 0.00 |
| HLR    | 1.00 | 0.00 | 0.00 |

Example 2: Fatty Acid Data. There are 120 observations, five groups and seven variables that represent the percentage levels of seven fatty acids, namely palmitic, stearic, oleic, linoleic, eicosanoic and eicosenoic acids. In this paper we consider five groups: rapeseed (G1), sunflower (G2), peanut (G3), corn (G4) and pumpkin (G5) oils. The original data set has eight groups and was used by [5] to develop and implement an automated method to classify vegetable oil samples. According to [5], the Principal Component Analysis (PCA) was successful to distinguish clusters of different oil samples, but it is not suitable for an automatic prediction of vegetable oil classes, because PCA requires a visual inspection and the final decision has to be made by an expert. The complete table of the original data can be found in the mentioned paper. In this paper we use the group 5 (pumpkin oil) as the reference
Table 5. Fatty acid data variables. Estimates and standard errors, ILR and HLR.

| Function | Variables | MLE ( SE ) | MEL ( SE ) |
|----------|-----------|------------|------------|
| 1        | Intercept | -98.9055 (107.1558) | -103.7708 (107.4063) |
|          | Palmitic  | 1.5720 (1.3175) | 1.6041 (1.3513) |
|          | Stearic   | -0.0941 (1.3175) | -0.1358 (1.4076) |
|          | Oleic     | 0.9256 (1.0708) | 0.9740 (1.0822) |
|          | Linoleic  | 0.9491 (1.1029) | 1.0097 (1.0903) |
|          | Linolenic | 1.2921 (1.0750) | 1.3325 (1.0468) |
|          | Eicosanoic| 3.5625 (4.2976) | 3.5678 (4.0833) |
|          | Eicosenoic| -1.4939 (3.6686) | -1.3585 (3.5988) |
| 2        | intercept | 1.2413 (153.8729) | 13.3611 (141.7017) |
|          | Palmitic  | 1.3511 (1.4740) | 1.2292 (1.3275) |
|          | Stearic   | 1.1142 (2.0490) | 1.1061 (2.0490) |
|          | Oleic     | -0.1526 (1.5591) | -0.2768 (1.4371) |
|          | Linoleic  | -0.2848 (1.5960) | -0.3764 (1.4627) |
|          | Linolenic | -1.2777 (1.8158) | -1.3835 (1.6955) |
|          | Eicosanoic| 1.5219 (4.1871) | 1.8051 (3.7158) |
|          | Eicosenoic| 1.5388 (4.7567) | 0.8945 (4.1551) |
| 3        | intercept | -x-         | -2579.9801 (1695.6943) |
|          | Palmitic  | -x-         | 23.0618 (14.9258) |
|          | Stearic   | -x-         | -14.6793 (14.4351) |
|          | Oleic     | -x-         | 23.5866 (15.8356) |
|          | Linoleic  | -x-         | 30.3859 (19.7337) |
|          | Linolenic | -x-         | 22.5048 (16.2788) |
|          | Eicosanoic| -x-         | -1.4245 (8.8502) |
|          | Eicosenoic| -x-         | 96.8737 (71.8125) |
| 4        | intercept | 408.8112 (418.0917) | 370.5995 (375.6840) |
|          | Palmitic  | -6.1084 (6.9035) | -5.5093 (6.2583) |
|          | Stearic   | -0.5833 (4.9011) | -0.6216 (4.8528) |
|          | Oleic     | -4.2459 (4.3936) | -3.8481 (3.9510) |
|          | Linoleic  | -4.3607 (4.4947) | -3.9576 (4.0427) |
|          | Linolenic | -1.7572 (2.8222) | -1.5179 (2.6002) |
|          | Eicosanoic| -10.6680 (10.2203) | -10.0232 (9.7806) |
|          | Eicosenoic| -2.7562 (3.9845) | -2.7033 (3.9584) |

Table 6. Fatty acid data. Variances (eigenvalues)

| Variance (λ) | 3.9092 | 1.0842 | 0.9325 | 0.7866 | 0.2653 | 0.0811 | 0.0001 |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| % of total variance | 55.85 | 71.84 | 84.66 | 95.90 | 98.83 | 99.99 | 100.00 |
Table 7. Fatty acid data. PCLR, estimates and standard errors.

| Function | Principal components | MLE ( SE ) |
|----------|----------------------|------------|
| 1        | Intercept            | 1.2539 (3.7960) |
|          | 1                    | 0.4637 (1.2132) |
|          | 2                    | 0.1935 (2.1425) |
|          | 3                    | 0.1214 (0.6125) |
|          | 4                    | 0.3026 (0.8387) |
|          | 5                    | -2.9400 (3.8016) |
|          | 6                    | -0.8097 (0.9521) |
| 2        | Intercept            | -11.0090 (4.4355) |
|          | 1                    | 0.4805 (1.2791) |
|          | 2                    | 2.1476 (2.0091) |
|          | 3                    | 0.9328 (0.5006) |
|          | 4                    | -1.1316 (0.7925) |
|          | 5                    | 1.4880 (3.5099) |
|          | 6                    | -0.3579 (1.3669) |
| 3        | Intercept            | -30.5484 (15.8682) |
|          | 1                    | -4.6451 (2.8750) |
|          | 2                    | 0.6804 (1.0748) |
|          | 3                    | -2.0162 (0.8480) |
|          | 4                    | 2.5532 (1.2140) |
|          | 5                    | 6.7826 (5.6999) |
|          | 6                    | -8.4422 (3.8623) |
| 4        | Intercept            | -1.5000 (10.2190) |
|          | 1                    | 0.01400 (1.6572) |
|          | 2                    | -5.8951 (7.1775) |
|          | 3                    | -0.1702 (1.6188) |
|          | 4                    | 2.5820 (2.3050) |
|          | 5                    | -2.3293 (6.6937) |
|          | 6                    | -2.3971 (4.1939) |

Table 8. Classification matrix. Fatty acid data.

| Model | Observed group | G 1 | G 2 | G 3 | G 4 | G 5 |
|-------|----------------|-----|-----|-----|-----|-----|
| PCLR  | 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|
| 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| 0.70| 0.30|
|       | G 5            | 0.15| 0.00| 0.05| 0.05| 0.75|

Example 3: Insulating Oil Data. This data set contains 2,567 observations, three groups and six variables that represent the physical and chemical characteristics that were observed in order to evaluate oil deterioration in power transformers: the neutralization number (NN), given in mg/KOH, the power factor (PF), given in percentages, the dielectric strength (DS), in kV, the interfacial tension (IFT) in
dyne/cm, water contents (WA), in p.p.m., and the oil temperature (OT), given in °C. Insulating oil that are in service are subject to deterioration due to mechanical and chemical conditions, and may oxidise producing acids and sludge. In order to evaluate oil conditions, oil samples should be taken from power transformers and tested. Insulating oils in service can be classified as: oils that are in satisfactory condition (G_1), oils that have low breakdown voltage and require reconditioning (G_2), and oils with low interfacial tension, high acidity and low electrical resistivity and in this case the oil should be replaced (G_3). In this paper we applied the four models in order to investigate their performances when data overlap. Table 9 displays the estimates, and standard errors that were obtained for the CLR, ILR and HLR models. The reference group is G_3 (oil to be replaced).

Table 9. Insulating oil data. Estimates and standard errors, CLR, ILR and HLR.

| Model | Variables | Function 1 | Function 2 |
|-------|-----------|------------|------------|
| CLR   | Intercept | -34.1505 (2.9944) | -41.8911 (3.6416) |
|       | NN        | -39.5000 (4.2585) | -38.8193 (8.3041) |
|       | PF        | -0.2095 (0.0489) | -0.3140 (0.0979) |
|       | DS        | 0.3654 (0.1902) | -0.5332 (0.2687) |
|       | IFT       | 18.1628 (1.3467) | 18.5252 (1.4319) |
|       | WA        | -0.4163 (0.1862) | 2.5440 (0.2505) |
|       | OT        | 0.1994 | 0.4498 (0.1968) |
| ILR   | Intercept | -30.0026 (2.7600) | -21.1716 (4.4459) |
|       | NN        | -32.1792 (3.8234) | -42.0228 (10.1668) |
|       | PF        | -0.2094 (0.0437) | -0.5447 (0.1628) |
|       | DS        | 0.1916 (0.1828) | -0.5588 (0.3477) |
|       | IFT       | 15.9302 (1.2279) | 10.5624 (1.7401) |
|       | WA        | 0.2527 (0.1823) | 1.3587 (0.3088) |
|       | OT        | 0.0709 (0.1242) | 0.5085 (0.3047) |
| HLR   | Intercept | -34.6608 (3.6381) | -42.4240 (3.6771) |
|       | NN        | -39.8560 (4.2918) | -38.6569 (8.2697) |
|       | PF        | -0.2127 (0.0492) | -0.3160 (0.0979) |
|       | DS        | 0.3697 (0.1914) | -0.5255 (0.2689) |
|       | IFT       | 18.4198 (1.3697) | 18.7933 (1.4537) |
|       | WA        | -0.4092 (0.1873) | 2.5427 (0.2505) |
|       | OT        | 0.2019 (0.1321) | 0.4489 (0.1967) |

Table 10 displays the principal components and their cumulative percentage of the total variance. Table 11 displays the estimates, and their standard errors, obtained by the PCLR model, with six principal components, and Table 12 summarizes the correct classification rates. Based on the results, we conclude that logistic regression provides an alternative to develop a computational tool that could evaluate the insulating oil condition.

Table 10. Insulating oil data. Variances (eigenvalues)

| Variance (λ) | 2.3863 | 1.3552 | 0.9047 | 0.6342 | 0.4544 | 0.2654 |
|--------------|--------|--------|--------|--------|--------|--------|
| Percentage of total variance | 39.77 | 62.36 | 77.44 | 88.01 | 95.58 | 100.00 |
Table 11. Insulating oil data. Estimates and standard errors, PCLR model.

| Model | Variables | Function 1         | Function 2         |
|-------|-----------|--------------------|--------------------|
| PCLR  | Intercept | -34.6608 (3.0381)  | -42.4240 (3.6771)  |
|       | 1st pc    | -29.7848 (2.4810)  | -28.0208 (4.0202)  |
|       | 2nd pc    | -18.4514 (1.6498)  | -20.3194 (2.8612)  |
|       | 3rd pc    | 2.7021 (0.2631)    | 2.0423 (0.3966)    |
|       | 4th pc    | 24.152 (2.3857)    | 23.9765 (4.4361)   |
|       | 5th pc    | -3.6795 (0.3548)   | -2.1049 (0.4817)   |
|       | 6th pc    | -9.8197 (2.3714)   | -8.5481 (5.1251)   |

Table 12. Classification matrix. Insulating oil data.

| Model | Observed group | Allocated Group |
|-------|----------------|-----------------|
|       | G 1            | G 2  G 3        |
| PCLR  |                |                 |
| G 1   | 0.98           | 0.00  0.02      |
| G 2   | 0.22           | 0.73  0.05      |
| G 3   | 0.05           | 0.00  0.95      |
| HLR   |                |                 |
| G 1   | 0.98           | 0.00  0.02      |
| G 2   | 0.24           | 0.75  0.01      |
| G 3   | 0.05           | 0.00  0.95      |

7. Conclusion. The application of the Classical Logistic Regression Model and the Individualized Logistic Regression Model has been reported in many studies involving bankruptcy prediction and cancer classification, among others applications. However, while these techniques work well for many situations, they may not work when the data set has no overlapping. Moreover, the logistic model becomes unstable when there is dependence between the explanatory variables.

The purpose with this job is to develop and implement a direct generalization for two alternative methods, the Hidden Logistic Regression and the Principal Component Logistic Regression models, for polytomous response, and to explore the performance of both models when compared to the Classical Logistic Regression and Individualized Logistic Regression models. The main advantage of the Hidden Logistic Regression model is that the MEL estimator always exists and it is unique, even when the data set has no overlapping. The Principal Component Logistic Regression model allows the reduction of the number of dimensions in a logistic regression model, with continuous variables and avoiding the multicollinearity of these variables. One can see that the model proposed by [18] and employed to model the relationship between a dichotomous response variable and a set of covariates, can be used with a few modifications when the response variable is polytomous. For practical purposes the main advantage is the existence and uniqueness of estimators. Furthermore, if the data set presents overlapping, the estimators are close to those obtained from the Classical Logistic Regression and Individualized Logistic Regression models. With respect to the model proposed by [1] we can see that this model, when applied to a polytomous response case, outperforms the Hidden Logistic Regression model at least in one simulation, although is not immune to the complete or quasi-complete separation.

In the future we intend to study the behavior 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. In the particular case of the HLR model the purpose is to study the influence that the
values that were stipulated to $\delta$ have over the performance of the model at issue, in relation to the classification rates.

Acknowledgments. We would like to thank very much Professor Andrés Medaglia and Professor Héctor Cancela, for their efforts, and the two anonymous referees, for their helpful and valuable comments and suggestions.

REFERENCES

[1] A. M. Aguilera, M. Escabias and M. J. Valderrama, Using principal components for estimating logistic regression with high-dimensional multicollinear data, Computational Statistics & Data Analysis, 50 (2006), 1905–1924.
[2] A. Albert and J. A. Anderson, On the existence of maximum likelihood estimates in logistic regression models, Biometrika, 71 (1984), 1–10.
[3] I. Andruski – Guimarães, Modelos de regressão logística oculto e de componentes principais para reconhecimento e classificação de padrões com variável resposta poliômica, DSc. thesis, Programa de Pós Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, (2006)
[4] C. B. Begg and R. Gray, Calculation of polychotomous logistic regression parameters using individualized regressions, Biometrika, 71 (1984), 11–18.
[5] D. Brodzijak – Vončina, Z.C.Kodba and M.Novič, Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids, Chemometrics and Intelligent Laboratory Systems, 75 (2005), 31–43.
[6] J. B. Copas, Binary regression models for contaminated data. With discussion, Journal of Royal Statistical Society B, 50 (1988), 225–265.
[7] A. Ekholm and J. Palmgren, A model for binary response with misclassification, In “Gil-Christ, R. (Ed.), GLIM 82, Proceedings of the International Conference on Generalized Linear Models,” Springer, Heidelberg (1982), 128–143.
[8] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics, 3 (1936), 179–188.
[9] T. Foucart, A decision rule for discarding principal components in regression, Journal of Statistical Planning and Inference, 89 (2000), 187–195.
[10] D. Gervini, Robust adaptive estimators for binary regression models, Journal of Statistical Planning and Inference, 131 (2005), 297–311.
[11] G. Heinze and M. Schepmer, A solution to the problem of separation in logistic regression, Statistics in Medicine, 21 (2002), 2409–2419.
[12] D. W. Hosmer and S. Lemeshow, “Applied Logistic Regression,” Wiley, New York, 2nd. Edition, 2000.
[13] M. Hubert and K. van Driessen, Fast and robust discriminant analysis, Computational Statistics & Data Analysis, 45 (2004), 301–320.
[14] N. Kodzarkhia, G. D. Mishra and L. Reiersolmoen, Robust estimation in the logistic regression model, Journal of Statistical Planning and Inference, 98 (2001), 211–223.
[15] E. Lesafre and A. Albert, Partial separation in logistic discrimination, Journal of Royal Statistical Society B, 51 (1989), 109–116.
[16] L. C. Paixão, Classificação de óleo isolante em transformadores elétricos de potência. MSc. Thesis. Programa de Pós Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná (2006).
[17] M. Rom and A. Cohen, Estimation in polytomous logistic regression model, Journal of Statistical Planning and Inference, 43 (1995), 341–353.
[18] P. J. Rousseuwen and A. Christmann, Robustness against separation and outliers in logistic regression, Computational Statistics & Data Analysis, 43 (2003), 315–332.

Received April 2007; 1st revision December 2007; final revision February 2009.

E-mail address: andruski@utfpr.edu.br
E-mail address: anselmo@ufpr.br