Iterative Multivariate Regression Model for Correlated Responses Prediction

S. Tom Au, Guangqin Ma, Rensheng Wang
AT&T Labs, Inc.- Research
180 Park Avenue
Florham Park, NJ 07932
{sau, gma, rw218j }@att.com

Abstract

In service oriented industry, a group of customers may be targeted for a set of marketing interests, and these interests are usually inter-correlated. For example, churn, upselling and appetency are often considered together, and decisions on how to retain customers, and to promote or to upgrade services are associated. Instead of predicting them separately as univariate models, we propose an iterative procedure to model multiple responses prediction into correlated multivariate predicting scheme. Our correlation factor derivations show that the exclusive case has more negative correlation factors, which is always favorable for responses separations in our multivariate prediction. We also point out that non-exclusive responses case can be reformed as another exclusive case via adding the overlapped positive response areas as new exclusive responses.

This proposed method combines partial least squares (PLS) method and logistic regressions, in which the former is used to extract the mutual information from correlations, while the latter is utilized to refine every single response prediction through auxiliary information from PLS predictions. In other words, not only with the given predictor matrix, but the predicted probability information from other correlated responses are also inserted to help every single response prediction. This hybrid regression modeling is implemented iteratively to refine the prediction gradually. More importantly, before every round of iteration, all the positive predictions from different responses compete each other and the highest values are kept for the only positive prediction and the others are changed to negative. Here we exploit the positive exclusive property (i.e., positive for one response means the negative for others) between multivariate responses. Numerical results show that with the aid of mutual information from other responses and the positive exclusion adjustment, our proposed scheme can improve the conventional regression models significantly.

1 Introduction

In customer care and marketing world, the same set of given customers will have multiple values that we are interested in, such as churn, upselling, appetency, etc. These values are reflected in their responses to a particular marketing design. These responses are inherently inter-correlated and the decisions on how to retain customers, and to promote or to upgrade services are usually considered together. For these kind of multi-responses predictions, most existing research works focus on univariate predictive modeling or algorithms [2] [3]. To capture the correlations between different responses, instead of predicting them separately as univariate models, multivariate model may help. The multivariate and multivariable predictive modeling scheme has been investigated in [6], which identifies the smallest combination of predictor variables accounting for a maximal proportion of the variation space of a given set of criterion variables and is based on additive multivariate measures of association. Unlike the scheme in [6], we begin with the investigation of the correlation factors between multivariate responses, and propose an iterative procedure to model multiple responses prediction into correlated multivariate predicting scheme. This iterative procedure utilizes the correlation structure of the predictors and the responses to enhance the prediction iteratively. Correlation related multivariate prediction has also been studied in [5], where the multinomial assumption is required in their logistic related models. Here we release this assumption and discuss a more general case. Our proposed method combines partial least squares (PLS) [9] [11] [7] method and logistic regressions, in which the former is used to extract the mutual information from correlations, while the latter is utilized to refine every single response prediction through auxiliary information from PLS predictions.

In particular, we are interested in how the proposed scheme outperforms conventional ones over different correlation structures and how to exploit the correlation property to direct our proposed algorithm adjustment for robust performance improvements. Our correlation factor derivations show that the positive exclusive responses (i.e., positive for one response means the negative for others) always have a negative correlation while the non-exclusive responses have either greater than zero or less than zero correlations depending on how much they overlap each other in terms of the positive responses. For multivariate response prediction, the negative correlations are always favorable since it helps the positive responses separation. To make use of the negative correlation in positive exclusive responses, we can reform the non-exclusive situations into exclusive ones by adding the overlapping responses as a new variable. As such, we only focus our discussion on the positive-exclusive situations in the following.

This proposed method is essentially an iterative fusion of PLS and logistic regressions. During each iteration, we utilize both the given predictor matrix and the predicted probability information from other correlated responses to help every single response prediction. More importantly, before we start a new round of iteration, all the positive predictions from different responses compete each other and the highest values are kept for the only positive prediction and the others are changed to negative (i.e., exploiting the aforementioned positive exclusive property). The positive and negative prediction exchange can gradually build up the regression accuracy during the consecutive iterations. Herein we use receiver operating characteristic (ROC) [4] and the area under ROC curve (AUC) [1] to evaluate our proposed scheme. The numerical results show that, when the correlation information and positive exclusive property have been adopted in the iterative procedures, the prediction accuracy can be significantly improved and the performance enhancement can be converged to stable level only after a few iterations.

The rest of paper is organized as below: In section 1, we formulate the problem we are interested in and introduce the necessary notations. Then, in section 2, for both exclusive and non-exclusive response situations, the definition and calculations of correlation factors between multivariate models are derived mathematically in terms of positive response probabilities. Following that, we present our proposed schemes as a combined iterative method with PLS and logistic regressions in section 3. Accordingly the numerical results are presented in section 4. In the end, we summarize and conclude our work in section 5.

Notation: Vectors (matrices) are denoted by boldface lower (upper) case letters respectively. All vectors are column vectors.

Let us assume $N \times p$ predictor matrix $\mathbf{X}$ denotes the N observations over p variables. Let $X_1, X_2, \cdots, X_p$ represent predictors and Y is the response. We have training data structure

$$\{(Y_i, x_{i1}, x_{i2}, \cdots, x_{ip}), \ \ i = 1, 2, \cdots, N\};$$

where the positive response $Y_i = 1$ or negative response $Y_i = 0$ indicating the outcomes for ith observation and x_{ij} is the jth explanatory variable for ith observation. For multiple outcomes situations, we have $N \times L$ responses matrix $\mathbf{Y}$, where L is the number of responses we are interested in. To simplify following discussion, without loss of generality, we focus our discussion with $L = 3$ and

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3],$$

where three $N \times 1$ vectors $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ are dependent and correlated responses. The problem of interest here is to investigate the impact of inherent correlation property among response vectors on the predictive modeling development and exploit this to enhance the multivariate predictive algorithm performance.

2 Correlation Between Responses

Let us define the correlation factor between response Y_i and response Y_j as ρ_{ij}. Mathematically,

$$\rho_{ij} = \frac{E\left[(Y_i - E[Y_i])(Y_j - E[Y_j])\right]}{\sigma_i \sigma_j} \tag{1}$$

where σ_i and σ_j are the standard deviations with Y_i and Y_j. Note response vectors $\mathbf{y}_i = \{Y_i\}$ and $\mathbf{y}_j = \{Y_j\}$ are bernoulli trial sequences, i.e., only containing elements $\{1\}$ or $\{0\}$. For N dimensional observation size, let us assume vector $\mathbf{y}_i$ and $\mathbf{y}_j$ separately have n_i and n_j positive responses $\{1\}$ (accordingly, $(N - n_i)$ and $(N - n_j)$ zero responses). With these notations, we note that the mean values are also the positive response probabilities,

$$E[Y_i] = n_i/N = Pr[Y_i = 1] \equiv p_i, \quad i = 1, 2, \cdots$$

Standard deviations σ_i and σ_j

$$\sigma_i = \frac{\sqrt{p_i(N - n_i)}}{\sqrt{(N-1)}} \approx \sqrt{p_i(1 - p_i)}, \quad i = 1, 2, \cdots$$

where we make use of the fact $(N-1) \approx N$ when N is large enough. These n_i and n_j positive responses with Y_i and Y_j can be overlapping or non-overlapping. For the former, it means for same observations, both responses are positive. For the latter, the positive responses within Y_i and Y_j are exclusive (i.e., if $Y_i = 1$ for one observation, then $Y_j = 0$ for sure with the same observation). We discuss both cases separately below.

2.1 Exclusive Responses

We first consider the non-overlapping, i.e., positive exclusive case, since it is relatively simple and meanwhile it can give some comparable information in the non-exclusive case.

$$E\left[(Y_i - E[Y_i])(Y_j - E[Y_j])\right] = -\, p_i p_j = -\frac{n_i n_j}{N^2}$$

Plugging the above results into (1), we have

$$\begin{aligned}\rho_{ij} =& \frac{-p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} \\ =& -\sqrt{\frac{p_i}{(1-p_i)}}\sqrt{\frac{p_j}{(1-p_j)}}\end{aligned} \tag{2}$$

- Remark 1: Correlation factors ρ_{ij} for exclusive responses are always less than zero due to the negative sign.
- Remark 2: For a fixed p_j, ρ_{ij} is a monotonic decreasing function with p_i, and vice versa. In particular, when p_i is getting small enough, then, ρ_{ij} is approaching to 0_- from negative side.

2.2 Non-Exclusive Responses

For non-exclusive responses, i.e., there are some $n_{ij}(n_{ij} < n_i, n_{ij} < n_j)$ observations where $Y_i = 1$ and $Y_j = 1$ happen at the same time. We define a related joint probability $p_{ij} \equiv Pr[Y_i = 1; Y_j = 1] = n_{ij}/N$. With introduction of the joint probability,

$$\begin{aligned}&E\left[(Y_i - E[Y_i])(Y_j - E[Y_j])\right] \\ =& p_{ij} - p_i p_j\end{aligned}$$

The standard deviations σ_i and σ_j stay the same as in exclusive responses case. Then, the correlation factor for non-exclusive responses,

$$\rho_{ij} = \frac{p_{ij} - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} \tag{3}$$

- Remark 1: Correlation factor ρ_{ij} positive or negative depends on joint probability p_{ij} greater or less than the product of probabilities $p_i p_j$.
- Remark 2: For a fixed p_j, ρ_{ij} is a monotonic decreasing function with p_i, and vice versa.

Comparing (2) and (3), one can see that for the same positive probability p_i and p_j, the exclusive situations have more negative correlation factors, which are more favorable for the response prediction separations. To exploit the negative correlation factors in exclusive situations, we can switch the non-exclusive situation into exclusive situation by introducing a new variable as the overlapping positive responses. After adding the new variable, all the positive responses are mutually exclusive. Therefore, in the following discussion, our proposed scheme assumes all the responses are positive exclusive and we will exactly exploit this exclusion property in our algorithm implementation.

3 Multivariate Logistic Regression

3.1 Conventional Logistic Regression

The conventional logistic regression [10] [8] is to predict individual vectors $\mathbf{y}_i$ from $\mathbf{X}$ with probability predictions. Mathematically, we will have estimated individual predictive vectors $\mathbf{p}_i$, $i = 1, 2, 3$;

$$\begin{aligned}\mathbf{p}_1 =& P_r(\mathbf{y}_1|\mathbf{X}) \\ \mathbf{p}_2 =& P_r(\mathbf{y}_2|\mathbf{X}) \\ \mathbf{p}_3 =& P_r(\mathbf{y}_3|\mathbf{X})\end{aligned} \tag{4}$$

where $\mathbf{p}_i$ is obtained from logistic regression, for given observations $\mathbf{X}$ with response $\mathbf{y}_i$. Here, we note that this method does not take advantage of the mutual information from other response vectors. For example, some response vectors will have exclusive relationship, i.e., one positive response in one vector can deduce the according responses in other vectors must be negative. On the other hand, one negative response in one vector does not mean anything with others for sure. This exclusive relationship brings partial correlations between the response vectors $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$. We consider to utilize the correlation information to improve the multivariate predictive modeling.

3.2 Fusion of PLS and Logistic Regression

Unlike the conventional logistic regression with individual considerations, partial least squares(PLS) regression takes advantage of both information from $\mathbf{X}$ and all vectors $\mathbf{y}_i$ at the same time for predictions. Specifically, PLS finds components from $\mathbf{X}$ that are also relevant for all responses $\mathbf{y}_i$. With notation $\mathbf{Y} = [\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3]$, PLS regression searches for a set of components (called latent vectors) that performs a simultaneous decomposition of $\mathbf{X}$ and $\mathbf{Y}$ with the constraint that these components explain as much as possible of the covariance between $\mathbf{X}$ and $[\mathbf{y}_1,\ \mathbf{y}_2,\ \mathbf{y}_3]$. This step generalizes principle components analysis (PCA). It is followed by a regression step where the decomposition of $\mathbf{X}$ is used to predict $\mathbf{y}_i$.

$$\hat{\mathbf{p}}_\mathrm{i} = P_r(\mathbf{y}_i|\mathbf{X}), \quad i = 1, 2, 3. \tag{5}$$

where $\hat{\mathbf{p}}_i$ is obtained from PLS regression, for given observations $\mathbf{X}$ with responses $\mathbf{y}_i$.

Obtaining $\hat{\mathbf{p}}_1$, $\hat{\mathbf{p}}_2$ and $\hat{\mathbf{p}}_3$ from (5), we can insert them into the conventional logistic regression in (4). This ends with a fusion method of PLS and logistic regression,

$$\begin{aligned}\tilde{\mathbf{p}}_1 =& P_r(\mathbf{y}_1|[\mathbf{X},\ \hat{\mathbf{p}}_2,\ \hat{\mathbf{p}}_3]) \\ \tilde{\mathbf{p}}_2 =& P_r(\mathbf{y}_2|[\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_3]) \\ \tilde{\mathbf{p}}_3 =& P_r(\mathbf{y}_3|[\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_2])\end{aligned} \tag{6}$$

where $\tilde{\mathbf{p}}_i$ is obtained from logistic regression, for given observations $\mathbf{X}$ and initial estimates from PLS with other two responses. The key point behind is that the estimated vector $\hat{\mathbf{p}}_i$ from PLS containing the correlated information of all response vectors $\mathbf{y}_i$, not a single one anymore. We would like to exploit the correlation to enhance the conventional logistic regressions. Moreover, the set of new probability vectors $\tilde{\mathbf{p}}_i$ can be inserted to the logistic regression (6) again to refine the regression precision. The iterative process can built up the predictive modeling performance gradually.

3.3 Positive Exclusive Responses

After obtaining the probability vectors $\hat{\mathbf{p}}_i$ from (5), we can slice then into to positive response part and negative response part via a properly selected threshold value T_i. For example, from the prior knowledge in the training data, (say, 30% positive responses), we can substitute the 30% highest value in the vector $\hat{\mathbf{p}}_i$ as "1" while the remaining as "0". We note that the prior knowledge, i.e., the threshold values selections are not critical in the following iteration procedures. The proposed scheme is robust to the T_i selections. One reason is that we will have positive response adjustment and the adjustment will be further reflected in each iteration. Following this thresholding process, we will create three binary vectors, $\hat{\mathbf{q}}_1,\ \hat{\mathbf{q}}_2,\ \hat{\mathbf{q}}_3$, by comparing $\hat{\mathbf{p}}_i$ with thershold T_i,

$$\hat{\mathbf{q}}_i = \begin{cases} 1 & \text{if } \hat{\mathbf{p}}_i > T_i \\ 0 & \text{if } \hat{\mathbf{p}}_i \le T_i \end{cases}$$

Again, one can not help adding the extra information into (6). That is,

$$\begin{aligned} \tilde{\mathbf{p}}_1 &= P_r(\mathbf{y}_1 | [\mathbf{X},\ \hat{\mathbf{p}}_2,\ \hat{\mathbf{p}}_3,\ \hat{\mathbf{q}}_2,\ \hat{\mathbf{q}}_3]) \\ \tilde{\mathbf{p}}_2 &= P_r(\mathbf{y}_2 | [\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_3,\ \hat{\mathbf{q}}_1,\ \hat{\mathbf{q}}_3]) \\ \tilde{\mathbf{p}}_3 &= P_r(\mathbf{y}_3 | [\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_2,\ \hat{\mathbf{q}}_1,\ \hat{\mathbf{p}}_2]) \end{aligned} \tag{7}$$

Our numerical results show that (7) can only bring negligible enhancement. Recall that $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ are positive exclusive responses. In the estimate set $\hat{\mathbf{q}}_1,\ \hat{\mathbf{q}}_2,\ \hat{\mathbf{q}}_3$, they do not satisfy the fact that one positive response in one vector means all negative responses in two other vectors. To utilize the positive exclusive property, we adjust the "1" and "0" values in $\hat{\mathbf{q}}_i$ according to the comparisons between $\hat{\mathbf{p}}_i$. Specifically, for n-th row elements of $\hat{\mathbf{q}}_i$, if there are more single "1", we need to compare the corresponding n-th row elements of $\hat{\mathbf{p}}_i$, and find the highest one. Then, accordingly to keep the related $\hat{\mathbf{q}}$ element single "1" and other two "0". For example, the n-th row elements of $\hat{\mathbf{q}}_i$ and $\hat{\mathbf{p}}_i$ are

$$\begin{aligned} \hat{\mathbf{q}}_i[n] &= [1, \quad 0, \quad 1] \\ \hat{\mathbf{p}}_i[n] &= [0.45, 0.72, 0.74] \end{aligned}$$

Then, the n-th row elements of $\hat{\mathbf{q}}_i$ should be adjusted as single "1" case for positive exclusive property since the third element of $\hat{\mathbf{p}}_i[n]$ has the largest value.

$$\hat{\mathbf{q}}_i[n] = [1, \quad 0, \quad 1] \quad \Rightarrow \quad \tilde{\mathbf{q}}_i[n] = [0, \quad 0, \quad 1]$$

Note that, this positive exclusive replacement only is implemented for any row with positive responses in $\hat{\mathbf{q}}_i$ since all negative responses do not reveal any extra information between multivariate responses.

Once the positive exclusive replacement is completed, $\hat{\mathbf{q}}_1,\ \hat{\mathbf{q}}_2,\ \hat{\mathbf{q}}_3$ will be tuned as new vectors $\tilde{\mathbf{q}}_1,\ \tilde{\mathbf{q}}_2,\ \tilde{\mathbf{q}}_3$, and the equation (7) can be updated as

$$\begin{aligned} \tilde{\mathbf{p}}_1 &= P_r(\mathbf{y}_1 | [\mathbf{X},\ \hat{\mathbf{p}}_2,\ \hat{\mathbf{p}}_3,\ \tilde{\mathbf{q}}_2,\ \tilde{\mathbf{q}}_3]) \\ \tilde{\mathbf{p}}_2 &= P_r(\mathbf{y}_2 | [\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_3,\ \tilde{\mathbf{q}}_1,\ \tilde{\mathbf{q}}_3]) \\ \tilde{\mathbf{p}}_3 &= P_r(\mathbf{y}_3 | [\mathbf{X},\ \hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_2,\ \tilde{\mathbf{q}}_1,\ \tilde{\mathbf{p}}_2]) \end{aligned} \tag{8}$$

Following that, the predictive probability vectors $\hat{\mathbf{p}}_1,\ \hat{\mathbf{p}}_2,\ \hat{\mathbf{p}}_3$ will be renewed as a refined set $\tilde{\mathbf{p}}_1,\ \tilde{\mathbf{p}}_2,\ \tilde{\mathbf{p}}_3$. Then, the next round iteration can be implemented so on and so forth.

We summarize the proposed scheme as below:

1. Use (5) and (6) to obtain the initial responses prediction $\hat{\mathbf{p}}_\mathrm{i}$.
2. User prior knowledge to convert $\hat{\mathbf{p}}_\mathrm{i}$ into $\hat{\mathbf{q}}_\mathrm{i}$ by threshold T_i.
3. Based on the ranks between $\hat{\mathbf{p}}_\mathrm{i}$, adjust the positive responses with $\hat{\mathbf{q}}_\mathrm{i}$ to $\tilde{\mathbf{q}}_\mathrm{i}$.
4. Use (8) to obtain the refined $\tilde{\mathbf{p}}_\mathrm{i}$.
5. Go to step 2 and refine the $\hat{\mathbf{p}}_\mathrm{i}$ with $\tilde{\mathbf{p}}_\mathrm{i}$ and start a new iteration.

4 Numerical Results

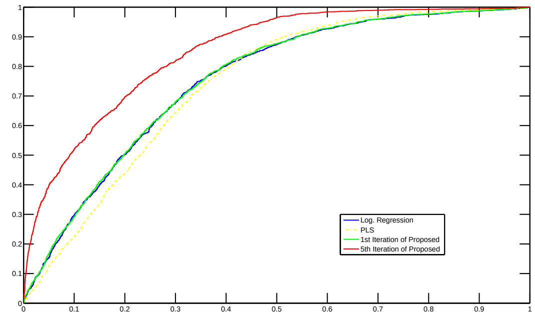
We use a real application data to verify the comparisons between the proposed multivariate PLS and LS hybrid method in (8) and the univariate method in (4). The data set is coming from KDD (knowledge discovery and data mining) CUP 2009. It contains 50,000 examples with the first 190 variables numerical and the last 40 categorical. There are in total three responses are of interest for comparisons, response 1 with 30% positive responses, response 2 with 6% and response 3 with 20%, and all three are positive exclusive. The compared performances are ROC curves and AUC with probabilities prediction.

We first examine the comparisons between conventional PLS, univariate logistic regressions and our proposed scheme with iterations for all three responses in Figure 1. Here Figure 1 (a) $\sim$ (c) show the individual comparisons of the ROC curves for the individual logistic regression, PLS and the iterative logistic regressions with aid of PLS. One can see from ROC comparisons that the hybrid logistic regression and PLS outperform both of them. In particular, when iteration is implemented to refine the procedures, the performance enhancement is even more obvious. To test the convergence of the iterative process, Figure 2 shows the converging performance of our proposed scheme for all three responses in terms of AUC values. As depicted in Figure 2, all three responses after the first 3 or 4 iterations have gain obvious performance improvements of AUC. As an additional remarker, we note that response 2 shows a relatively large range fluctuation due to the small size of positive responses (only 6%). Response 1 is almost 5 times large as response 2 and during the positive exclusive adjustment, response 2 needs 3 or 4 iterations before it converge to a stable performance gain. As a whole, all three sets results show consist convergency of AUC.

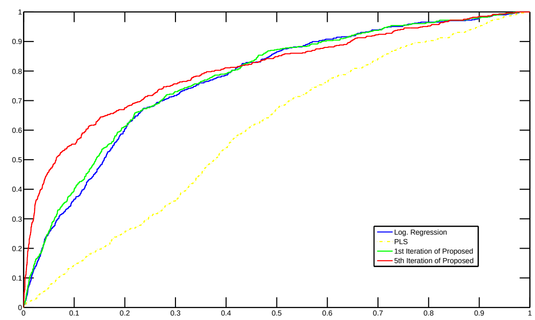
5 Conclusions

In this study, we propose an iterative procedure for predictive modeling with correlated multivariate responses data. This procedure utilizes the correlation structure and positive exclusive property of the responses and the predictors to enhance the prediction accuracy iteratively. Our algorithm development and performance analysis is based on the key point of correlation factors in multivariate modeling. In particular, we are interested in how the proposed scheme outperforms conventional ones over different correlation structures and how to exploit the correlation property to direct our proposed algorithm adjustment for robust performance improvements. Our correlation factor derivations show that the positive exclusive responses always have a negative correlation and are more favorable for prediction separations.

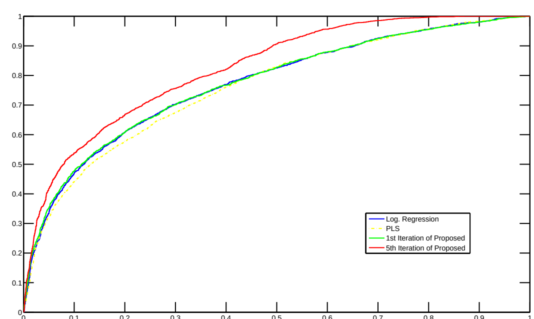
This proposed method combines partial least squares (PLS) method and logistic regressions, in which the former is used to extract the mutual information from correlations, while the latter is utilized to refine every single response prediction through auxiliary information from PLS predictions. The hybrid regression modeling is implemented iteratively to refine the prediction gradually. More importantly, before every round of iteration, all the positive predictions from different responses compete each other and the highest values are kept for the only positive prediction and the others are changed to negative. Numerical results show that with consideration of the correlation information and the positive exclusion adjustment, our proposed scheme can improve the performance of conventional regression models significantly.



(a) ROC curves for Response 1.



(b) ROC curves for Response 2.



(c) ROC curves for Response 3.

Figure 1. ROC curve comparisons with the conventional logistic regression, PLS, the first iteration and the 5th iteration of the proposed scheme. Response 1 is with 30% positive response, response 2 with 6% and response 3 with 20% and all of them are mutually positive exclusive.

References

[1] SAS Institute Inc. SAS-Graphics. *Cary, NC, USA,*

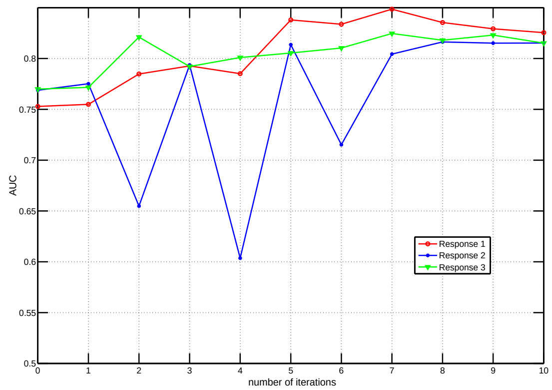


Figure 2. AUC comparisons of all 3 responses with respect to different iterations. Response 1 is with 30% positive response, response 2 with 6% and response 3 with 20% and all of them are mutually positive exclusive.

2001-2003.

[2] A. Afifi and V. Clark. Computer-Aided Multivariate Analysis. 1999, Chapman & Hall, Ltd. London, UK.

[3] S. Au, S. Li, and G. Ma. Applying and Evaluating Models to Predict Customer Attrition Using Data Mining Techniques . *Journal of Comparative International Management*, 6(1):10–22, 2003.

[4] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. John Wiley amd Sons, Inc, New York., 1966.

[5] O. Kuss. A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Computer Methods and Programs in Biomedicine*, vol.87, issue.3:262–269, September, 2007.

[6] M. W. S. Martin W. Gillo. Predictive Modeling of Multivariable and Multivariate Data. *Journal of American Statistical Association*, vol.69, No.347:646–653, 1974.

[7] B. D. Marx. Iteratively Reweighted Partial Least Squares Estimation for Generalized Linear Regression. *Technometrics*, vol.38, No. 4:374–381, 1996.

[8] S. Menard. *Applied logistic regression analysis*. Sage Pubns, 2002.

[9] P. NOTIONS. Partial least squares regression and projection on latent structure regression (PLS Regression). 2010.

[10] L. Ott and M. Longnecker. *An introduction to statistical methods and data analysis*. Duxbury Pr, 2008.

[11] S. Wold, A. Ruhe, H. Wold, and W. Dunn III. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5:735, 1984.