Journal of the Nigerian Association of Mathematical Physics
Volume **18 (May, 2011), pp 373 – 380**

Efficient Data-Driven Rule for Obtaining an Optimal Predictive Function of a Discriminant Analysis

[1]Osemwenkhae J. E. and [2]Iduseri A.

[1]Department of Mathematics,
University of Benin, Nigeria
[2]Department of Mathematics and Statistics,
Ambrose Alli University, Nigeria.

Abstract

This paper proposes a rule for optimizing a predictive discriminant function (PDF) in discriminant analysis (DA). In this study, we carried out a sequential-stepwise analysis on the predictor variables and a percentage-N-fold cross-validation on the data set obtained from students' academic records in a university system. The hit rates, $P^{(a)}$ result obtained for the optimized PDF, $Z_{(OPT)}$ calibrated on training and validation sets, when compared with that of PDF, Z obtained using the conventional rule, showed a significant improvement in terms of how well each PDF classifies cases into values of the categorical dependent. It was also discovered that the optimized PDF, $Z_{(OPT)}$ produces consistent high hit rates with little variability, thereby reducing the problem of overfitting.

Keywords: Optimal Predictive function, Overfitting, sequential-stepwise analysis, percentage-N-fold cross-validation

1. Introduction:

Discriminant analysis (DA) is a classical statistical method used to classify observations into predefined groups with respect to several underlying variables. Typically, a discriminant function is developed using observations with known group membership and this is then used to classify observations with unknown group membership. DA computes an optimal transformation by minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum class discrimination. DA has two sets of procedures: they are predictive discriminant analysis (PDA) and descriptive discriminant analysis (DDA). These two procedures are based on purpose of analysis. In PDA, the focus is on classifying subjects into one of several groups, whereas in DDA the focus is on revealing major difference among the groups [18].

Professionals in various fields of human endeavor are regularly faced with the problem of making prediction. When the criterion for prediction involves one or more predictor variables alongside with a categorical criterion, such prediction will call for the use of predictive discriminant analysis. Over the years, various methods of optimizing PDF (or maximizing the actual hit rate, $P^{(a)}$) and/or predicting the fit of a PDF to a hypothetical validation sample have been developed. These include stepwise methods [4, 10, 12, 14, 20, 21], all-possible subset approach [9, 13, 17, 20], hold-out sample method [3, 19], repeated random sub-validation method [15, 22, 23], leave-one-out cross-validation [1, 6, 7, 8, 19] and the K-fold cross-validation [2, 6]. In most cases, these methods produce an overly optimistic estimate of the success of classification otherwise known as overfitting, an indication that the PDF obtained by these methods are often less than optimal.

In discriminant analysis, a predictive discriminant function can be optimized by sorting/getting the best training sample from the historical sample, D_n which involves cleaning the historical sample to remove potential errors such as outliers. This procedure is analogous to optimizing a decision trees (classification trees in particular) which involves decreasing the level of impurity, which results in having terminal node (or rule) with only one response value- see [16]. In the same vein, using computation (as in data-driven procedures) in place of mathematical analysis to obtain empirical estimates of performance in PDA can also be used in predictive discriminant analysis to fine tune the discriminant weights over and above training sample. While the sheer volume of the data-driven methods developed over the years may be impressive, the application of these methods in PDA in particular, mainly focused on either obtaining an optimal combination of predictor variables in order to maximize hit rate, $P^{(a)}$or in predicting the fit of a PDF to a hypothetical validation set.

In this study, we propose a new rule that will optimize the PDF using computation in place of mathematical analysis. The predictive power of the optimized PDF, $Z_{(OPT)}$ was determined by comparing its predictive performance with the PDF, Z obtained using the conventional rule.

[2]Corresponding authors: ***Iduseri A: E-mail:*** augustineiduseri@yahoo.com, Tel: (+2348023713354 – O. J. E.)

2.0 The Proposed Rule

The proposed rule we present in this work involves a two stage analysis. The second stage of the proposed rule involves carrying out a percentage-N-fold partitioning of the data set and will therefore be referred as the percentage-N-fold cross-validation rule (NFCV$_{\text{-P}}$). In the first part of the analysis, for a data set that consists of N sample $(x_i,\ y_i)$, where $x_i \in \{1,\ 2,\ldots,\ p\}$ denote the corresponding predictor variable label, $y_i \in \{1,\ 2,\ldots,K\}$ denote the corresponding group label, p is the number of predictor variables, and K is the number of groups. The data matrix is given as

$$X = [x_1,\ x_2,\ .\ .\ .,x_P] \tag{1}$$

We obtain the data matrix for the potential predictor variables (control variable) as

$$X^* = [x^*{}_1,\ x^*{}_2,\ .\ .\ .,x_P{}^*] \tag{2}$$

The difference in the hit rate, $P^{(a)}$ between the control variables, X^* and the test variables, X-X^* indicates the explanatory effect of the test variable over and above the set of control variables. The subsets of the historical sample, D_N with the highest hit rate, $P^{(a)}$ will now be chosen as the new historical sample, D_N^* . From the historical sample, D_N^* we compute a PDF, Z given by

$$\begin{aligned} Z &= u_1\, X_1 + u_2\, X_2 + \ .\ .\ .\ + u_p\, X_p \\ &= \eta\ (D_N^*) \end{aligned} \tag{3}$$

where Z is the linear discriminant function, u_i is the discriminant weight, X_i is the predictor variables and $\eta\ (D_N^*)$ indicates that the PDF is calibrated on selected predictor variables from a pool of identified predictor variables.

For the second part of the analysis, a percentage-N-fold cross-validation on the data set, which is a modification of the K-fold cross-validation [6] was carried out in order to fine tune the discriminant weights or make stable the PDF over and above the historical sample D_N^* .

The outline of the percentage–N-fold cross-validation (NFCV$_{\text{-P}}$) procedure is given as follows:

Step1: Obtain a training set, $I^{(t)}$ as a percentage of the historical sample, D_N^*

Step2: For each training sample, $D_N^{(t)}$ obtained in step1, compute $Z = \eta\left(D_N^{(t)}\right)$ and obtain it's $P^{(a)}$ on the Historical sample, D_N^*

Step3: Repeat steps 1-2 using percentage values of 60, 70, 80 and 90 respectively. The optimize predictive function, $Z_{(OPT)}$ is the function with discriminant weights, u^* having the best matching performance on its training sample, $D_N^{(t)}$ and the historical sample, D_N^* . The optimized PDF is given as

$$\begin{aligned} Z_{(OPT)} &= u^*{}_1\, X_1 + u^*{}_2\, X_2 + \ .\ .\ .\ + u^*{}_p\, X_p \\ &= \eta^*\ (D_N^{(t)}) \end{aligned} \tag{4}$$

where the $u's$ are the discriminant weights with the best matching performance. This discriminant weights, u^* becomes the estimate of the true values that would be found if the observation set were comprised of all members of the population.

The work of [8, 11] opined that predictive functions with few predictor variables, relative to D_N, yield relatively more accurate and more precise estimators. Building a series of predictive functions and measuring the performance of each predictive function, we can see which predictive function resulted in the best matching performance [16].

3.0 Method of Evaluation of The Proposed Rule

In order to determine the efficiency of the rule, two predictive discriminant functions (PDF$_S$) are built. The first PDF was built using the conventional rule and the second was built using our proposed rule. The data (historical sample, D_N) for this study were obtained from students' academic records for 100 and 200 levels, in the Department of Statistics, in a University system as shown in Table 1 (extract from [5]). In the first stage of data collection, two groups of students in terms of their graduating class of degree were formed, and nine predictor variables, including the following: overall GPA for 100 level, grades in all the statistics and mathematics core courses.

In using the conventional approach, the task of optimizing the PDF (maximizing hit rate, $P^{(a)}$) begins with the researcher obtaining an optimal combination of the predictor variables when the number of predictor variables is more than two. Using the forward stepwise analysis, we found that the GPA and STA 202 made significant independent and combined contributions. In order to confirm the GPA and STA 202 as the best subsets of the predictor variables using the forward stepwise analysis, we then used an "all-possible subsets" approach which gave the same result [9, 20]. We use an arbitrary linear combination given by

$$Z = u_1 X_1 + u_2 X_2 \tag{5}$$

To determine the vector of discriminant weights, u in equation (5), we compute the inverse of the pooled sum of squares and cross products matrix, W given by

$$W^{-1} = \frac{1}{|W|} C \text{, where } C \text{ is the adjoint of W}$$

$$= \begin{bmatrix} 0.031770626 & -0.000662576 \\ -0.000662576 & 0.000097068 \end{bmatrix}$$

The deviation of Group 2 centroid from Group 1 centroid is given by

$$d = \begin{bmatrix} X_{1G1} - & X_{1G2} \\ X_{2G1} - & X_{2G2} \end{bmatrix}$$

$$= \begin{bmatrix} 1.1265 \\ 16.0167 \end{bmatrix}$$

So that the estimate of the vector of discriminant weights, u becomes

$$u = W^{-1} d = \begin{bmatrix} 0.025177329 \\ 0.000808317 \end{bmatrix}$$

While the standardized vector of discriminant weights, u_S is given by

$$u_S = \frac{u_i}{\sqrt{\underline{U}^T S \underline{U}}} = \begin{bmatrix} 1.345 \\ 0.043 \end{bmatrix}$$

Substituting the above values of u_S into equation (5), we obtain the PDF, Z given by

$$Z = 1.345\,(GPA) + 0.043\,(STA201) \qquad (6)$$

In using our proposed rule, the analysis involves two stages just as we have in the conventional approach. The first stage which is a pre-analysis to the second stage, involves carrying out a sequential stepwise analysis on the predictor variables, in order to obtain optimal combination of the predictor variables without dropping an important predictor variable(s). For the data set in Table 1, the students GPA and STA 202 were also the two predictor variables that made significant independent and combined contributions when a sequential stepwise analysis was done on the predictor variables. The second stage involves carrying out a percentage-N-fold cross-validation ($NFCV_{-P}$) on the data set. Using the outline of our proposed rule ($NFCV_{-P}$), we then build the optimized PDF, $Z_{(OPT)}$ by carrying out the following steps:

Step1: Obtain a training set, $I^{(t)}$ as a percentage of the historical sample, D_N^*

Step2: For each training sample, $D_N^{(t)}$ obtained in step1, compute $Z = \eta\left(D_N^{(t)}\right)$ and obtain it's $P^{(a)}$ on the Historical sample, D_N^*

Step3: Repeat steps 1-2 using percentage values of 60, 70, 80 and 90 respectively

Tables 2 present value of discriminant weights, u and cutoff points, Z_C obtained for different percentage values of the historical sample in Table 1. While Table 3 present summary of hit rates, $P^{(a)}$ results for each PDF calibrated on the training set, $I^{(t)}$ obtained from different percentage values of the historical sample in Table 1.The number outside the bracket in each row of the first column of Tables 2 and 3 indicate the percentage value and the number of training samples obtained based on the different percentage values used. The optimize predictive function, $Z_{(OPT)}$ is the function with discriminant weights, u^* (whose values are written in bold case in Table 2) having the best matching performance on its training sample, $D_N^{(t)}$ and the historical sample, D_N^* .The optimized PDF is given as

$$Z_{(OPT)} = 1.277\ (GPA) + 0.045\ (STA\ 201) \qquad (7)$$

To assess the predictive power of the optimize PDF, $Z_{(OPT)}$ (7) obtained using our proposed rule over above the PDF, Z (6) obtained using the conventional rule, the two PDF_S were validated using the $I^{(t)}$, and $I^{(V)}$. Table 4 present summary of hit rates, $P^{(a)}$ for both the PDF_S , while Figures 1 to 4 present the line graphs for the two sets of values in Tables 4.

4.0 Discussion of Results

Figure 1 reveals that the validation hit rates represented by the red line are more dispersed or characterized with high degree of overfitting compared to the training hit rates represented by the blue line. This is an indication that the PDF, Z obtained using the conventional rule produces an overly optimistic estimate of hit rates, $P^{(a)}$ when tested on the data that gave its birth. If we look intently at Figure 1, we found that this overfitting problem was more obvious for validation samples 2, 11, 15, 17 and 21,

compared to that of their corresponding training samples. In Figure 2, the overfitting problem was completely eliminated for validation samples 2, 15, and 21, when PDF, $Z_{(OPT)}$ was tested on these samples. Also in Figure 1,
we observed that the hit rates for validation samples 2, 11, 15, 17 and 21 fell below the 80 percent hit rate mark when the PDF, Z was tested on these validation samples. While in Figure 2, only two validation samples hit rates fell below 80 percent hit rate mark when PDF, $Z_{(OPT)}$ was tested on the validation sets,

This reduction in the number of validation samples with low hit rates is due to the fact that overfitting was kept at minimal when our proposed rule was used in building the PDF, $Z_{(OPT)}$. When we tested the two PDF_S on the training sets and validation sets separately as shown in Figure 3 and Figure 4, the two had the same predictive performance when tested on the training sets, except for sample 19 as shown in Figure 3. Judging from this same predictive performance obtained for the two PDF_S when tested on the training set, as shown Figure 3, one would have expected the same results for validation sets in Figure 4, but the reverse was the case. When we validated the two PDF_S using a test sample (see Table 1) independent of the training sample, the optimized function, $Z_{(OPT)}$ produced a higher hit rate of 82.4, compared to that of the PDF, Z with a hit rate of 76.5 as shown in Table 4 written in bold case. This goes to further prove that the optimized predictive function, $Z_{(OPT)}$ that was build using our proposed rule did not only produce consistent high hit rate with little variability, but perform better in terms of predictive performance.

5.0 Conclusions

Despite the sheer volume of contributions to DA in the literature, it is obvious that the major contributions to DA (especially in PDA), are variable selection methods, validation, and cross-validation methods. None of these methods or rules here are able to keep the problem of overfitting at minimal. In this work, a rule for obtaining an optimal PDF, which is stable enough to produce consistent high hit rates with little variability, thereby reducing the problem of overfitting was presented. The predictive performance of the optimized PDF, $Z_{(OPT)}$ alongside the PDF, Z obtained using the conventional rule was evaluated using the training and validation sets obtained from the historical sample, D_N^*, as well as an independent test sample. It was determined that the optimized PDF, $Z_{(OPT)}$ built from our proposed rule was better in terms of the precision with which it correctly classifies set of observations or future samples from the same population.

Table 1: Historical Sample

HISTORICAL SAMPLE												TEST SAMPLE					
VALUES OF GPA AND STA 201 FOR TWO GROUPS												VALUES OF GPA AND STA 201 FOR TWO GROUPS					
GROUP 1 (n_1= 60)						GROUP 2 (n_2 = 60)						GROUP 1(n_1=22			GROUP 2 (n_2 =12)		
NO	GPA	STA 201	NO	GPA	STA 201	NO	GPA	STA 201	NO	GPA	STA 201	NO	GPA	STA 201	NO	GPA	STA 201
1	2.11	34	**31**	3.30	56	**1**	1.57	34	**31**	2.19	28	**1**	2.89	80	**1**	2.11	58
2	3.41	57	**32**	2.02	65	**2**	1.54	50	**32**	1.86	47	**2**	3.62	76	**2**	2.19	61
3	2.44	46	**33**	2.86	64	**3**	2.27	40	**33**	2.32	44	**3**	3.46	78	**3**	1.49	60
4	2.65	42	**34**	3.30	67	**4**	1.08	15	**34**	3.08	40	**4**	3.38	75	**4**	2.57	40
5	235	50	**35**	3.38	32	**5**	1.11	21	**35**	1.22	27	**5**	2.57	57	**5**	1.86	40
6	3.08	70	**36**	2.72	55	**6**	1.56	40	**36**	1.33	24	**6**	3.46	63	**6**	1.97	43
7	3.35	70	**37**	2.37	54	**7**	2.19	47	**37**	2.78	46	**7**	3.00	63	**7**	2.08	49
8	3.14	56	**38**	2.03	71	**8**	1.76	64	**38**	1.81	52	**8**	1.78	71	**8**	1.86	58
9	4.00	67	**39**	2.14	53	**9**	2.57	40	**39**	1.81	52	**9**	3.30	68	**9**	2.63	50
10	2.89	40	**40**	3.78	62	**10**	2.35	46	**40**	2.97	61	**10**	2.32	67	**10**	2.21	41
11	2.70	53	**41**	2.05	58	**11**	2.11	40	**41**	2.14	40	**11**	2.46	42	**11**	1.41	24
12	3.05	57	**42**	3.59	53	**12**	1.76	64	**42**	2.20	46	**12**	2.92	66	**12**	1.89	61
13	2.38	43	**43**	3.35	64	**13**	1.06	21	**43**	1.89	43	**13**	3.41	55			
14	3.46	73	**44**	2.13	56	**14**	1.97	44	**44**	2.56	41	**14**	2.11	67			
15	3.92	69	**45**	2.81	63	**15**	2.78	45	**45**	2.81	36	**15**	1.68	69			
16	2.57	53	**46**	2.32	67	**16**	1.14	46	**46**	1.59	37	**16**	2.41	62			
17	3.95	79	**47**	4.11	66	**17**	2.43	50	**47**	1.92	40	**17**	1.59	77			
18	3.73	73	**48**	4.08	63	**18**	2.51	48	**48**	1.97	43	**18**	3.37	53			
19	3.68	65	**49**	3.27	60	**19**	2.00	33	**49**	2.05	40	**19**	2.89	65			
20	3.11	40	**50**	3.78	53	**20**	2.16	41	**50**	1.87	40	**20**	1.70	56			
21	3.19	45	**51**	2.51	56	**21**	1.33	40	**51**	1.64	28	**21**	2.43	79			
22	3.08	59	**52**	3.41	62	**22**	2.27	24	**52**	1.97	50	**22**	2.00	59			
23	2.81	53	**53**	3.49	60	**23**	1.22	40	**53**	2.81	53						
24	2.89	60	**54**	2.35	48	**24**	1.49	41	**54**	1.68	62						
25	3.51	60	**55**	2.00	49	**25**	1.95	52	**55**	1.81	48						
26	3.59	64	**56**	3.22	52	**26**	1.43	42	**56**	2.03	41						
27	3.59	57	**57**	3.32	46	**27**	1.78	41	**57**	1.89	35						
28	3.59	62	**58**	3.19	71	**28**	2.08	56	**58**	2.49	45						
29	2.54	51	**59**	3.16	58	**29**	1.24	29	**59**	1.19	40						
30	2.43	56	**60**	3.86	62	**30**	1.76	25	**60**	1.12	41						

Source: Extract from Erimafa et al. (2009)

Table 2: Values of discriminant weights and cutoff marks for different percentage values of the historical sample

Percent of Historical Sample	**Discriminant Weights** **Function 1**	**Cutoff Marks**	**Percent of Historical Sample**	**Discriminant Weights** **Function 1**	**Cutoff Marks**
60 (2)	GPA 1.074 STA 202 0.077	6.5526	**90 (10)**	GPA 1.301 STA202 0.054	6.0130
	GPA 1.843 STA202 0.007	4.9057		GPA 1.204 STA202 0.051	5.4732
70 (3)	GPA 1.205 STA202 0.072	6.5655		GPA 1.400 STA202 0.048	5.8096
	GPA 1.257 STA202 0.041	5.2228		GPA 1.275 STA202 0.044	5.3787
	GPA 1.654 STA202 0.023	5.1775		GPA 1.293 STA202 0.045	5.4477
80 (5)	GPA 1.137 STA 0.063	6.0075		GPA 1.401 STA202 0.038	5.3853
	GPA 1.328 SAT202 0.045	5.6615		GPA 1.485 STA202 0.036	5.2172
	GPA 1.343 STA202 0.040	5.3479		GPA 1.496 STA202 0.032	5.2746
	GPA 1.691 STA202 0.022	5.2510		GPA **1.277** STA202 **0.045**	5.3834
	GPA 1.262 STA202 0.044	5.3063		GPA 1.333 STA202 0.042	5.4042

Table 3: Summary of classification results for each PDF calibrated on training set obtained from different percentage values of the historical sample

% of Historical Sample	**% of Training Sample Correctly Classified**	**% of Historical Sample Correctly Classified**
60 (2)	90.3	86.7
	86.1	80.8
70 (3)	90.5	85.8
	82.1	85.8
	89.3	84.2
80 (5)	88.5	85.0
	86.5	87.5
	84.4	85.8
	88.5	84.2
	87.5	87.5
90 (10)	88.0	85.8
	85.2	86.7
	88.0	86.7
	86.1	87.5
	86.1	87.5
	86.1	85.8
	88.0	85.8
	85.2	85.0
	87.9	87.5
	87.0	87.5

Table 4: Summary of hit rate results for PDF, Z and PDF, $Z_{(OPT)}$ calibrated on training and validation sets obtained from historical sample

PDF, Z		PDF, $Z_{(OPT)}$	
Results for Training Set	**Results for Validation Set**	**Results for Training Set**	**Results for Validation Set**
87.5	87.5	87.5	87.5
87.5	43.8	87.5	87.5
89.3	83.3	89.3	83.3
83.3	97.2	83.3	97.2
90.5	80.6	90.5	80.6
88.5	83.3	88.5	83.3
86.5	91.2	86.5	91.7
85.4	95.8	85.4	95.8
89.6	79.2	89.6	79.2
87.5	87.5	87.5	87.5
89.8	66.7	89.9	66.7
86.1	100.0	86.1	100.0
88.0	83.3	88.0	83.3
86.1	100.0	86.1	100.0
86.1	50.0	86.1	100.0
87.0	91.7	87.0	91.7
88.0	66.7	88.0	66.7
87.0	91.7	87.0	91.7
88.0	83.3	87.0	83.3
87.0	91.7	87.0	91.7
	76.5		**82.4**

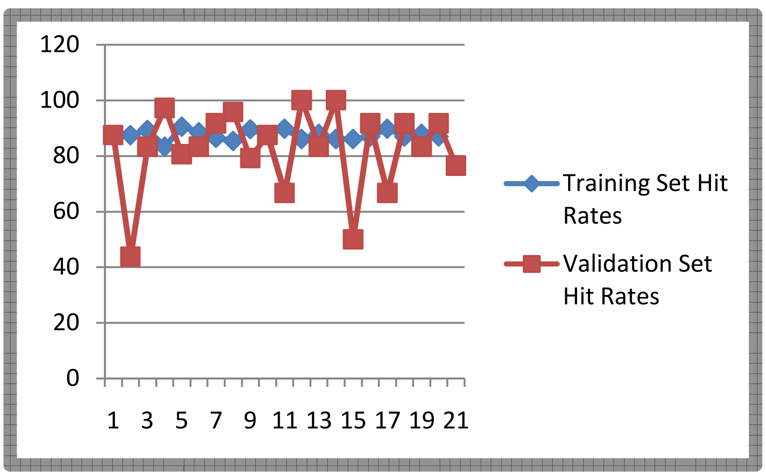

Figure 1: Graphical representation of hit rates for PDF, Z obtained using the conventional rule calibrated on training and validation sets as shown in Table 3

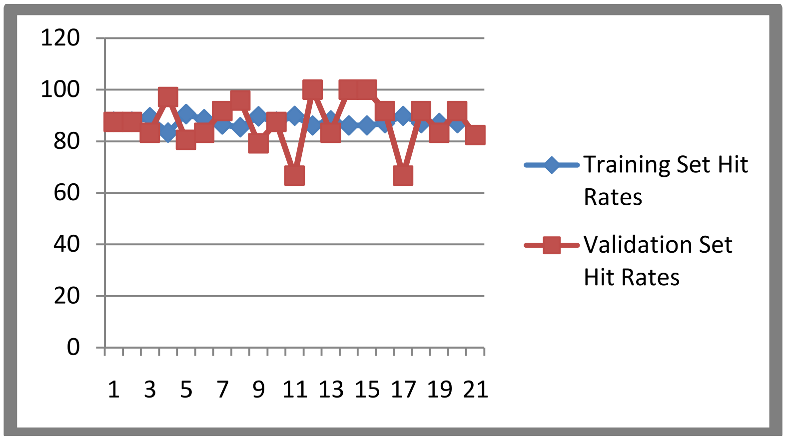

Figure 2: Graphical representation of hit rates for the optimal PDF, $Z_{(OPT)}$ obtained using the percentage N-fold cross-validation approach calibrated on training and validation sets as shown in Table 3

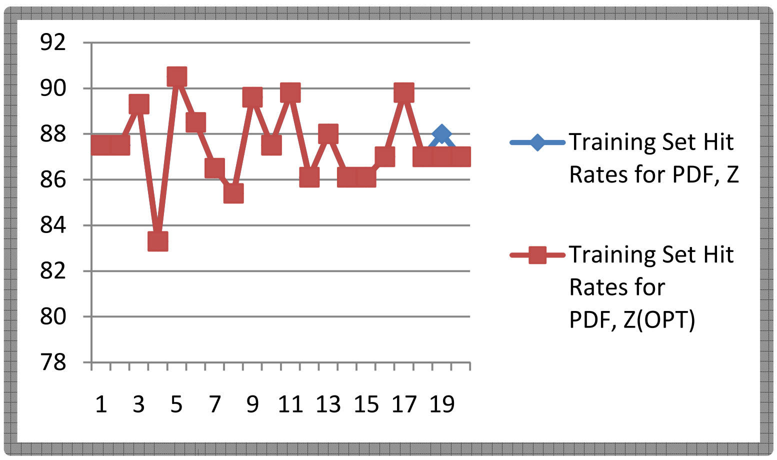

Figure 3: Graphical representation of hit rates for both PDF_S calibrated on training sets as shown in column 1 and 3 of Table3

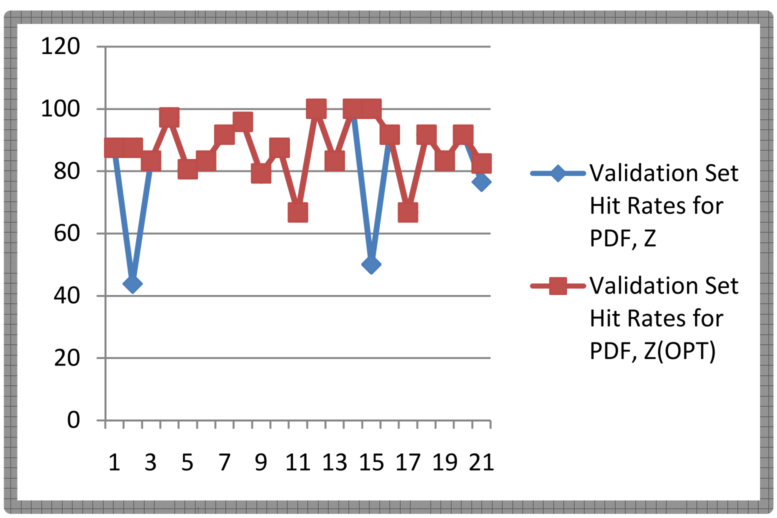

Figure 4: Graphical representation of hit rates for both PDF_S calibrated

References

[1] **Allen, D. M. (1974).** The Relationship Between Variable Selection and Data Augumentation and a Method for Prediction. Technometrics, 16: 125-127. MRO343481
[2] **Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984).** Classification and Regression Trees. Wadsworth Statistics/Probability Series. Wadsworth Wadsworth Advanced Books and Software, Belmont, CA. MR0726392
[3] **Devroye, L. and Wagner, T. J. (1979).** Distribution-Free Performance Bounds for Potential Function Rules. IEEE Transaction in Information Theory, 25(5): 601-604. MR0545015
[4] **Efroymson, M. A. (1960).** Multiple regression analysi.. In A. Ralston & H. S. Wilf (Eds.). Mathematical methods for digital computers, New York: Wiley, 191-203
[5] **Erimafa, J. T., Iduseri, A. and Edokpa, I. W. (2009).** Application of Discriminant Analysis to Predict the Class of Degree for Graduating Students in a University System. International Journal of Physical Sciences. Vol. 4(1), pp. 016-021.
[6] **Geisser, S. (1975).** The Predictive Sample Reuse Method with Applications. J.Amer. Statist. Assoc., 70: 320-328
[7] **Glick, N. (1978).** Additive Estimators for Probabilities of Correct classification. Pattern Pattern Recognition, 10(297, 302, 303): 211-227.
[8] **Hora, S. C., and Wilcox, J. B. (1982).** Estimation of error rates in several population discriminant analysi., Journal of Marketing Research, 19(302,303,304,336): 57-61
[9] **Huberty, C.J. (1989).** Problems with Stepwise Methods: Better Alternatives. In B.Thompson (ED), Advances in Social Science Methodology ,1: 43-70). Greenwich, CT: JIA Press.
[10] **Huberty, C.J. (1994).** Applied discriminant Analysis. New York: Willey and Sons.
[11] **Huberty, C. J., and Stephen, O. (2006).** Applied manova and discriminant analysis. Hoboken, New Jersey. John Wiley and Sons, Inc.
[12] **Jennrich, R. I. (1997).** Stepwise Regression. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.). Statistical Methods for Digital Computers, 3: 58-75. New York: Wiley
[13] **Jesus, T. (2000).** Variable Selection Strategies in Discriminant Analysis. Paper presented at the annual meeting of the Southwestern Psychological Association, Dallas
[14] **Klecka, W. R. (1980).** Discriminant Analysis. Quantitative Application in Social Sciences Series. 19. Thousand Oaks, CA: Sage Publications.
[15] **Kohavi, R. (1995).** A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Procedings of the Fourteenth International JointConference on Artificial Intelligence 2 (12): 1137-1143.
[16] **Myatt, G. J. (2007).** Making Sense of Data: A Practical Guide to Exploratory data Analysis and Data Mining. A John Willey and Sons, Inc., Publication, U. S. A.
[17] **Neter, J., Kutner, M, H., Nachtsheim, C. J., and Wasserman, W. (1996).** Applied Linear Statistical Models (4th ed.) Chicago, IL: Irwin
[18] **Stevens, J. (1996).** Applied Multivariate Statistics for the Social Sciences. (3rd ed.)," Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
[19] **Stone, M. (1974).** Cross-Validation Choice and Assessment of Statistical Predictions. J. Roy. Statist. Soc. Ser. B, 36. Pg. 111-147. With discussion and a reply by the authors. MRO35637
[20] **Thompson, B. (1995).** Stepwise Regression and Stepwise Discriminant Analysis need Not Apply Here: A guidelines editorial. Educational and Psychological Measurement, 55(4): 525-534.
[21] **Whitaker, J.S. (1997).** Use of Stepwise Methodology in Discriminant Analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, January, 1997.
[22] **Wold, S. (1976).** Pattern Recognition by means of Disjoint Principal Component Models. Pattern Recognition 8, pp. 127-139. [ISI]
[23] **Wold, S. (1978).** Cross-validation Estimation of the Number of Components in Factor and Principal Component Model., Technometrics 20, pp. 397-405 [ISI]