NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

SELECTION OF CORE PREDICTIVE FEATURES OF STUDENTS' ACADEMIC GRADES USING BRIDGE ESTIMATION METHOD

ALABI Peter A., OLALEYE Olalekan A., ANKELI Uchechi C., ALADE A. F., OLATIDOYE Waheed Y. and BAMIGBOYE Jonah. S.

Statistics Department, Federal Polytechnic, Ede, Osun State, Nigeria

hezekiaholaleye@gmail.com, justsaypeter001@yahoo.com

Abstract

It is frequently a key concern of tertiary institution administration and academics in particular to understand the predetermining factors that are most related to student's academic performance. This study tends to examine the efficiency of some sparse estimators (i.e., Sparse Group-LASSO, Group-lasso & Group-Bridge) in predicting the student's final grades by selecting the core predictive factors. The grades of the students (Failed/Passed) are the target variable of interest in the data, and the predictors include seven (7) variables: gender, number of absence days, relations, programs, parent's educational background, Course units, and Grade points. The original set of data was split (30:70) to training-set comprises of 2800 observations with a test set comprises of 1200 observations. The three (3) estimators were assessed for classification and prediction of their probability of a given target variable, using the Sensitivity (SEN), Specificity (SPEC), Misclassification Error Rate (MER), Precision (PREC), Negative Predictive value (NPV) and Positive Predictive value (PPV). It was discovered that the Group-Bridge estimator selected Grade points, Gender, and Parental Education Background as the core relevant factors to the students' academic achievement, while the baseline estimator (Group-LASSO) and Sparse Group-LASSO selected Subject units and Grade points. Also, the study found that Sparse Group-LASSO is much more efficient the other two estimators in selecting and predicting the core relevant predictors.

Keywords: Students' Academic Performance, Predictive factors, Sparse Group – LASSO, Training set and Test Set.

1.0 Introduction

It is frequently a major curiosity of administration of tertiary institution along with academics in particular to be able to determine which variable (if any) in terms of courses of study by students are mainly germane to determine the academic good standing together with final grades of their students. It will, therefore be meaningful and helpful as a scientist to proffer a meaningful solution to this supervised statistical learning problem by suggesting statistical method that are capable of selecting best model subset(s) for efficient selection of core relevant predictors of students' academic grades. A

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

sparse statistical model possesses just a minute amount of nonzero parameters or weights; thus, it is easier to calculate approximately and deduce than a dense model. Statistical Learning through Sparsity: Lasso and Generalizations give methods that utilize sparsity to help recuperate the fundamental signal in a data set.

Professionals in this fast-growing field, in computation, authors explained lasso for linear regression with a easy coordinate descent algorithm. They argued the relevance of ℓ_1 penalties to summarized linear models and sustain vector machines, extending to generalized penalties, which include elastic net, as well as group lasso, with review of numerical methods used in optimization. Moreover, they delve into statistical inference methods for fitted (lasso) models, which incorporate the bootstrap, Bayesian methods, plus some newly developed approaches.

Hoerl & Kenmard (1970a, 1970b), projected ridge regression, that minimizes RSS student to a constraint $\sum|\beta_1|^2 \leq t$. The ridge regression reduces OLS estimator towards zero and produces a biased estimator. Frank &Friedman (1993) unveiled bridge regression, that often reduces RSS subject to a constraint $\sum|\beta_1|^q \leq t$ with $q > 0$. The estimator comprises of two significant unique cases. The first is the familiar ridge estimator, i.e. when $q = 2$ (Hoerl & Kennard, 1970). The second is the LASSO estimator, i.e. when $q = 1$ (Tibshirani, 1996), which was introduced as a variable selection and shrinkage method. According to Fu (1998), for differed value of q, the guarded area is dissimilar in the parameter space. However, while Frank & Freedom (1993) for any given $q > 0$ didn't work out the estimator of bridge regression, they then stated that optimizing the parameter q is worthy of note.

Adding to bridge estimators, various penalization methods was anticipated with the intension of shrinkage estimation alongside simultaneous variable selection. However, for SCAD penalty, with handful number of finite parameter, Fan &Li (2001) delve into understanding some characteristics of penalized likelihood methods asymptotically. Huang *et al*., (2005) being mindful of some particular regularity situation, properly proved the existence of local maximizers of a penalized likelihood which possesses an oracle. Moreover, Fan & Peng (2004) as well analyzed the same work from the angle of the divergence of such parameters in use.

The characteristics of oracle is simply revealing that the local maximizes can accurately choose the non-zero coefficients having the probability that converges to 1 and that estimators of the non-zero coefficients are normal asymptotically having equal means and covariance which they would have of the zero coefficients were known in advance.

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

Moreover, HimelMallick and Nengjun Y (2018) proposed from the perspective of Bayesian, a bridge regression. Distinct from classical bridge regression which sums up the inference employing a single point estimate, the anticipated Bayesian method gives vague estimates of the regression factors, permitting rational inference in the course of the posterior distribution. Portnoy (1984, 1985) made available situation on the rate of growth of p_n which are adequate for steadiness along with the asymptotic normality of class of the M-estimators of regression factors. Nonetheless, in sparse models, Portnoy didn't work on the penalized regression of collection of variables. Bair*et al*., (2004) established the stability of managed principal components study subjected to a partial orthogonality circumstance. However, they too didn't work on penalized regression.

Therefore, in this study, we incorporate the bridge estimator to be used on linear regression; comparing the performance of Bridge estimator, Spars Group lasso, and Group lasso in selecting best model subset(s) for efficient selection of the relevant predictors of some set of data.

2.0 Materials with Methods

2.1 *Data Source with Description*

The data sets utilized for this work are primary and secondary types. The predictor data which include gender of the students (Male/ Female), Number of absence days (Below 7 days/ Above 7 days), Relation of the student (Father/ Mother/ Both), programs of the students (Full-Time/ Daily Part-Time/ Regular Part-Time), parent's education background (No Education or Primary sch./Secondary/ HND/BSc. / MSc. / PhD.) and grade of the students (Failed/ Passed), were collected using the medium of Google forms (online questionnaire) from the selected students of Federal Polytechnic Ede, Osun State. The second data were online extracted from MIS (Management Information System) data archive of the Federal Polytechnic, Ede, Osun state, Nigeria which span a period of five years covering National Diploma (ND) sets. The data sets include univariate quantitative response y involving students' subject units and grade points. The predictor matrix X include continuous variables involving raw scores of students in courses offered from year one through year two of their ND programs as the case may be.

2.2 ***Group Bridge Estimator***

Considering a Bridge Estimator, we let $x_k = (x_{1k}, \ldots x_{nk})'(k = 1, \ldots, d)$ be the design sectors and $y = (y_{i,} \ldots, y_n)'$ being a response vector inside (1), the linear regression model is given below:

$$y_i = x_{i1}\beta_i +, \ldots . + x_{id}\beta_d + \varepsilon_i \ (i = 1, \ldots, n) \qquad (1)$$

where

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

y_i = response variable

$x_{i1}, \cdots, x_{id}$ = covariate variables,

β_j's = regression coefficients

$\varepsilon_i s$ = error terms with an error vector $\varepsilon = (\varepsilon_{i,} \ldots \ldots, \varepsilon_n)$.

Let $A_{1,} \ldots . A_j$ be subsets of $\{1, \cdots, d\}$ which stand for identified groupings of the design vectors, which also signify the regression coefficients in the j^{th} group by $\beta_{Aj} = (\beta_{k,} k\epsilon A_j)$.′. For any mx1 vector a, denoted its L_1, norm by$\|a\|_j = |a_1| + \cdots + |a_m|$ we look into the objective function.

$$L_n(\beta) = \left\| y - \sum_{k=1}^{d} x_k \beta_k \right\|_2^2 + \lambda_n \sum_{j=1}^{J} c_j \|\beta_{Aj}\|^{\gamma} \quad (2)$$

where

$\lambda_n > 0$ = penalty level

c_j = constants for modification of the various dimension of β_{Aj} A simple choice is $c_j \propto |A_j|^{1-\gamma}$,

$|A_j|$ = cardinality of A_j.

In equation (2), bridge penalty is used for the L_1 norms of the fractioned coefficients. Thus, the $\widehat{\beta_n}$. that minimizes (2) is known as a group bridge estimator.

Note: When;

$|A_j| = 1$ $(j = 1, \cdots, J)$ equation (2) reduces to standard bridge measure.

γ = 1, equation (2) stands for lasso criterion that can merely select individual variable.

0 <γ< 1, we have the group bridge criterion of equation (2) is capable of variable selection at the cluster with each variable levels concurrently.

A straight reduction of $L_n(\beta)$ is difficult, because all collection bridge penalty is a convex function for $0 < y < 1$. A corresponding minimization function which is easier to work out computationally for $0 < y < 1$, define

$$S_{1n}(\beta, \theta) = \left\| y - \sum_{k=1}^{d} x_k \beta_k \right\|_2^2 + \sum_{j=1}^{J} \theta^{1-1/\gamma} c_j^{1/\gamma} \left\| \beta_{A_J} \right\|_1 + \tau_n \sum_{j=1}^{J} \theta_j \quad (3)$$

where $\tau_n =$ penalty parameter

2.2.1 *Group Lasso*

This is stated as

$$\tilde{\beta}_n = \arg\min \left\| y - \textstyle\sum_{k=1}^{d} x_k \beta_k \right\|_2^2 + \lambda_n \textstyle\sum_{j=1}^{j} \|\beta_{Aj}\|_{Kj,2} \quad (4)$$

Where $k_j$is a positive define matrix and $\|\beta_{Aj}\|_{Kj,2} = (\beta'_{Aj}K_j\beta_{Aj})^{1/2}$

According to Yuan & Lin (2006), $K_j = |A_j|I_J$ where I_j, is the $|A_J| \times |A_j|$ identify matrix.

Let τ_n be a penalty parameter which is define as following

$$S_{2n}(\beta, \theta) = \left\|y - \sum_{k=1}^{d} x_k\beta_k\right\|_2^2 + \sum_{j=1}^{j} \theta_j^{-1} \|\mathrm{B}_{Aj}\|_{Kj,2}^2 + \tau_n \sum_{j=1}^{j} \theta_j \qquad (5)$$

2.2.2 *Sparse Group Lasso*

Group Lasso merely possesses sparsity between groups, but within the cluster, there exists no sparsity. The sparsity between groups means that at least a single set of coefficients is non-zero. But when there is sparsity in the coefficients vector components, then we talk about sparsity within cluster. Nevertheless, in numerous useful problems, the impacts of variables within clusters tend to be different, that can reduce its application. As suggested by Simon *et al.* (2013), sparse group Lasso decisive factor is:

$$\min_\beta \frac{1}{2n}\left\|y - \sum_{l=1}^{m} X^{(l)}\beta^{(l)}\right\|_2^2 + (1 - \alpha)\lambda \sum_{l=1}^{m} \sqrt{p_i}\,\|\beta^{(l)}\|_2 + \alpha\lambda\|\beta\|_1 \qquad (6)$$

Where $X_{N\times p}$ is segmented into m sub-matrix $X^{(1)}, X^{(2)}, \cdots, X^{(m)}$, $X^{(i)}$ represents a sub-matrix of X, $\beta^{(l)}$ stands for coefficient vector in group i having p_i length. $\alpha \in [0,1]$, a penalty of it is convex mixture of Lasso together with group Lasso penalty. Standing on Nesterov's technique in generalizing gradient descent, algorithm is formed to compute $\beta^{(l)}$.

3.0 Data Analysis

To carry-out the necessary analysis, the study used a data set on the final grading of the students in Federal Polytechnic Ede. In the data set, there are 4000 binary observations indicating whether a student passed or failed in the final result, and 7 integer-value variables. All the features are log-transformed before applying the estimators. We split the data into a training set having 2800 observations with a test set having 1200 (ratio of 70:30). This implies that, 70 of the data set was used to train our model/ estimator, while the test set was used to test set (unseen set) the performance of the estimators in the classification of the classes of the grade (Failed & Passed). The estimators are tuned on the training set using 5-fold cross validation before predicting on the test set.

4.0 Results and Discussion

Table 1: Frequency Table (Performance distribution)

Partition Ratio	Failed (%)	Passed (%)	Total

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

Train set (70%)	480 (17.1%)	2320 (82.9%)	2800
Test set (30%)	199 (16.6%)	1001 (83.4%)	1200

The table indicates that less than one-quarter of the students performed poorly (failed) in the final grading of the 4000 students selected for this study, while a large number of the students performed well in the selected course for both the train and test sets.

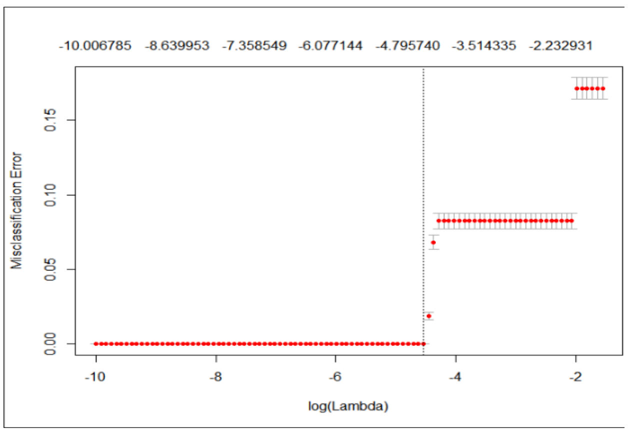


Fig. 1: *Plot of the Log. of lambda (λ) against Misclassification errors(5-fold) for Group-LASSO*

From Figure 1, the plot exhibits the misclassification errors in accordance with lambda logarithm. The vertical line that is dashed signifies that log of optimal value of lambda is closed to that which minimizes the misclassification error.

Table 2: Result of the baseline model (group-LASSO)

Variables	**Coefficients**
(Intercept)	-67.1267874
Subject Units	0.02406601
Grade Points	31.2734801
Gender	-
Absence Days	-
Programs	-
Parental Education Background	-

From Table 2 which presents the results of the Group-LASSO estimator, we realized that the estimator only fitted a model containing only two of the predictors (i.e., Subject/Course units and Grade points). This implies that the remaining predictors or features were shrinking to zero. In other words, it shows that the remaining features do not significantly contribute to the final grades of a student. Noticeably, the results show that the selected features contributed positively to the final grades of the students.

Table 3: Confusion matrix (Group-LASSO)

	Actual Grade		
Predicted Grade		Failed	Passed
	Failed	101	0
	Passed	98	1001

From Table 3, we realized that the True Positive (TN) = 1001, True Negative (TN) =

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

101, False Positive (FP) = 98, and the False Negative (FN) = 0. The results indicate that the estimator correctly classified 1001 out of the 1200 students to passed and correctly classified 101 students to failed in their final grades. Meanwhile, the estimators falsely predict 98 students to passed in their final grade.

Table 4: Model Assessment criteria

	SEN	SPEC	MER	ACC	PPV	NPV	PREC
Train	0.51875	1.0000	0.75938	0.9175	1.00	0.90945	0.9031
Test	0.50754	1.0000	0.75377	0.9183	1.00	0.91083	0.9084

Table 4 depict the statistical measures in assessing the performance of the estimator on both the Train and Test set data, this include the Sensitivity (SEN), Specificity (SPEC), Misclassification Error Rate (MER), Precision (PREC), Negative Predictive value (NPV) and Positive Predictive value (PPV). A little difference in the performance of the estimator on the Training set compare to the performance in the Test set was realized.

Table 5: Result of the Sparse Group-LASSO

Variables	Coefficients (λ=lambda.1se = 7.375024e-05)
(Intercept)	-9.2702780
Subject Units	0.2061579
Grade Points	4.5121608
Gender	-
Absence Days	-
Programs	-
Parental Education Background	-

Comparing the result from the Sparse Group-LASSO (Table 5) estimator to that of Group-LASSO in Table 4, we noticed that they both selected two features (i.e., Subject units and Grade points), but not surprisingly, we could see how the Sparse Group-LASSO estimator further shrink the coefficient for Grade points more toward zero and we noticed that the intercept also reduced, meanwhile, the coefficients for the subject units was found to increase just a little from 0.02406601 (for Group-LASSO) to 0.2061579 in Sparse Group-LASSO estimator output in Table 5.

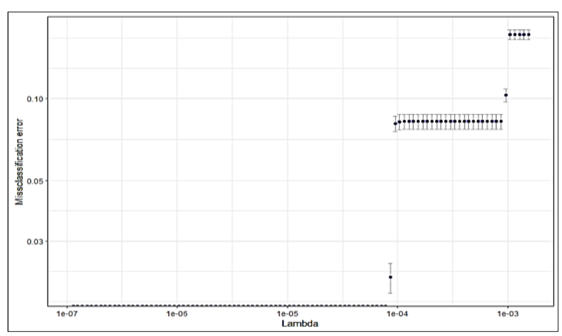


Fig. 2: *Plot of the Log. of lambda (λ) against Misclassification errors(5-fold) for Sparse Group-LASSO*

In Figure 2, we have the plot of the of the Log of lambda (λ) against Misclassification errors

NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

(5-fold) from the Sparse Group-LASSO. We could notice that, right from the left to the right, the first significant misclassification error was reported at lambda around 1e-04 which was around the value for the lambda used in fitting the Sparse Group-LASSO in Table 5.

Table 6: Confusion matrix (Sparse Group-LASSO)

		Actual Grade	
		Failed	Passed
Predicted Grade	Failed	102	0
	Passed	97	1001

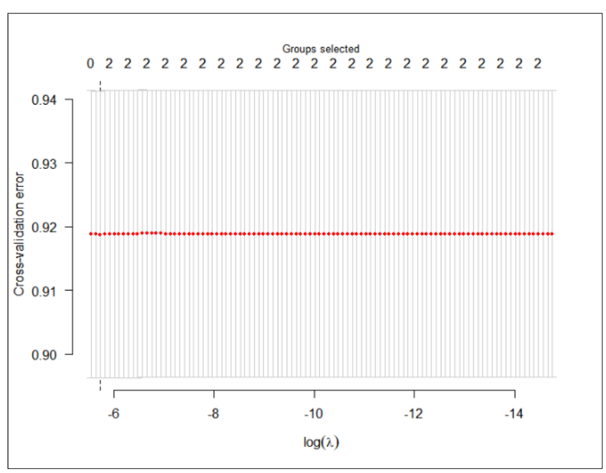
Table 6 indicates that the Sparse Group-LASSO was able to correctly predict that a student will pass the course, in another words, it was able to correctly classified that 1001 out of 1200 students will belong to the passed class, 102 out of 1200 students would failed the course. Meanwhile, we realized the estimated incorrectly predicted that 97 of the students will pass the course.

Table 7: *Result of the Group-Bridge*

Variables	**Coefficients (λ = 0.0032)**
(Intercept)	1.60402667
Subject Units	-
Grade Points	-0.02233682
Gender	0.08195955
Absence Days	-
Programs	-
Parental Education Background	0.02531375

Nonzero coefficients = 3, nonzero groups = 2.

From Table 7, we noticed that the Group-Bridge estimator performed different to the former estimators. For instance, the other two estimators are found to select the subject units and grade points, although, we realized that the Sparse Group-LASSO performed better in handling the sparsity among the predictors. However, we discovered that the Group-Bridge estimator selected three predictors, including the Grade points, Gender, and the parental educational background as the determining features to the performance of the student in terms of their final grades.



NATIONAL BOARD FOR TECHNICAL EDUCATION
NIGERIAN JOURNAL FOR TECHNICAL EDUCATION
Volume 24 Nos. 1 2025
ISSN No. 2992 - 3522

Fig. 3: *Plot of the Log. of lambda (λ) against Cross-validation errors(5-fold) for Group-Bridge*

Figure 3 depicts the plot of the log lambda against the cross validation error. We can deduce from the plot that the estimator was able to classify the predictors into two groups (nonzero group).

Table 8: Confusion matrix (Group-Bridge Estimator)

	Actual Grade		
Predicted Grade		Failed	Passed
	Failed	0	0
	Passed	199	1001

From Table 8 we discovered that the Group-Bridge estimator predicted the number of students that passed in the final grades correctly TP = 1001, but performed poorly in predicting the number of students that failed in the test data TN = 0.0.

Table 9: Model Assessment criteria

Estimators	Partitions	SEN	SPEC	MER	ACC	PPV	NPV	PREC
Group-LASSO	Train	0.51875	1.0000	0.75938	0.9175	1.0000	0.90945	0.9031
	Test	0.50754	1.0000	0.75377	0.9183	1.0000	0.91083	0.9084
Sparse Group-LASSO	**Train**	**1.0000**	**1.0000**	**0.0000**	**1.0000**	**1.0000**	**1.0000**	**0.8286**
	Test	**0.5126**	**1.0000**	**0.6128**	**0.9950**	**1.0000**	**0.9117**	**0.5561**
Group-Bridge	Train	0.0000	1.0000	0.1715	0.8286	-	0.8286	1.0000
	Test	0.0000	1.0000	0.4439	0.5561	-	0.8343	1.0000

From Table 9, which is the assessment criteria on the Sensitivity (SEN), Specificity (SPEC), Misclassification Error Rate (MER), Precision (PREC), Negative Predictive value (NPV) and Positive Predictive value (PPV), we realized that the Sparse Group-LASSO estimator outperformed the remaining two estimators, in the case of fitting the train and test set. The study discovered that the Sparse Group-LASSO is more sensitive in classifying the classes of the response variable (i.e., the final grades) with a Sensitivity value (SEN = 1.000) followed by the Group-LASSO. Also, in terms of the prediction accuracy, the Sparse Group-LASSO also outperformed the other two estimators with an Accuracy value (ACC = 0.9950) followed by the Group-LASSO, and in the case of the Negative Predictive value, the Sparse Group-LASSO also outperformed the rest with (NPV = 0.9117). Meanwhile, we considering the misclassification error rate, we noticed that the Group-Bridge estimator outperformed the other two with a misclassification error rate (MER = 0.4439), followed by the Sparse Group-LASSO estimator, and finally in terms of Precision, the result indicate that the Group-Bridge estimator outperformed the other two estimators as well, with a perfect value (PREC = 1.000).

5.0 Conclusion

This study selected the core predictive variables using Bridge Estimation Method for Students' Academic Class of Grades. It selected 4000 students randomly from among the students of the Federal Polytechnic, Ede, Osun State. Meanwhile, only eight (8) variables, including gender of the students, Number of absence days, Relation of the student, programs of the students, parent's education background, grade of the students and two continuous variables, including Course units, Grade points were considered. Based on the findings, the study revealed that the core relevant variables to the performance of the undergraduate students are the subject units and grade points. Nonetheless, according to the Group-Bridge estimator, gender and parental background of the students can also contribute to the performance of these students. Meanwhile, the Sparse Group-LASSO performed well in determining the performance of the students considering some of the students' socio-demographic factors. This is as a result of the superiority of the estimator in the presence of sparsity among the predictors.

References

Akomolafe A. A. and Amalai G. N. (2014): Efficiency of Discriminant Analysis in identifying performance of students at Risk in Nigeria Private Universities. Pacific Journal of Science and Technology 16(1), 201-205

Akomolafe, A. A., Olayiwola O. M. and Bello O. A. (2008). Using multivariate analysis to identify students at Risk in Nigeria Universities. Journal & Educational Research and policies 3, (4), 87-91

Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2004). Prediction by supervised principal components. Stanford University Department of Statistics Tech Report, Sept 2004.

Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). Technometrics 35 109-148.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96:1348{1360.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. Ann. Statist. 32 928{961.

Fu, W. J. (1998). Penalized regressions: the bridge versus the Lasso. J. Computational and Graphical Statistics 7 397-416.

Hoerl, A.E., and Kennard, R.W. (1970a), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, 12, 55–67.

Hoerl, A.E., and Kennard, R.W. (1970b), "Ridge Regression: Applications to Nonorthogonal Problems," Technometrics, 12, 69–82.

Huang, J., Wang, D. L. and Zhang, C.-H. (2005). A Two-Way Semi linear Model for Normalization and Analysis of cDNA Microarray Data. J. Amer. Statist. Assoc. 100 814-829.

Olayiwola O. M., Adebayo A. A., Okunlade A. A. and Akomolafe A. A. (2008). The effect of students' pre-admission performance on the post admission performance. International Journal of Educational Research and Administration 5(1), 44 - 47

Portnoy, S. (1984). Asymptotic behavior of M estimators of p regression parameters when $\frac{p^2}{n}$ is large: I. Consistency. Ann. Statist. 12 1298-1309.

Portnoy, S. (1985). Asymptotic behavior of M estimators of p regression parameters when $\frac{p^2}{n}$is large: II. Normal approximation. Ann. Statist. 13 1403-1417.i

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Ser. B, 58:267-288.