International Journal of Advanced Statistics and Probability, 3 (1) (2015) 93-99
www.sciencepubco.com/index.php/IJASP
©Science Publishing Corporation
doi: 10.14419/ijasp.v3i1.4364
Research Paper

Performance analysis on least absolute shrinkage selection operator, elastic net and correlation adjusted elastic net regression methods

Pascalis Kadaro Matthew, Abubakar Yahaya*

Department of Mathematics, Ahmadu Bello University, Zaria – Nigeria
**Corresponding author E-mail: abubakaryahaya@abu.edu.ng*

Copyright © 2015 Pascalis Kadaro Matthew, Abubakar Yahaya. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Some few decades ago, penalized regression techniques for linear regression have been developed specifically to reduce the flaws inherent in the prediction accuracy of the classical ordinary least squares (OLS) regression technique. In this paper, we used a diabetes data set obtained from previous literature to compare three of these well-known techniques, namely: Least Absolute Shrinkage Selection Operator (LASSO), Elastic Net and Correlation Adjusted Elastic Net (CAEN). After thorough analysis, it was observed that CAEN generated a less complex model.

Keywords: *Convex Optimization; Cross Validation; Multicollinearity; Penalized Regression.*

1. Introduction

Variable selection techniques are, undoubtedly, among the most important tools used in eliminating the problem of multicollinearity and/or reducing problems related to modeling bias when a large number of predictor variables are introduced in multiple linear regression (MLR) analyses. MLR is one of the most commonly used data mining techniques that can provide insightful information in cases where the classical assumptions associated with MLR are met. It is a versatile tool that is found to be applicable to diverse areas of human endeavor. Much has been published regarding this concept. For instance, Kutner et al. [21] as well as Adams [1] provided a thorough account of MLR, and these literatures will be found to be indispensable for most interested readers.
A key step in developing an appropriate MLR model can be achieved by selecting a model-building method that guarantees a set of best model criteria. For instance, Efroymson [9] introduced stepwise regression, which is among the most commonly used tools for model building. Stepwise regression was intended to be an automated procedure that selects the most statistically significant variables from a finite pool of independent variables. It is noteworthy at this point to know that, there are three separate stepwise regression procedures, namely: forward selection, backward selection and mixed selection. According to Kutner et al. [21], Neter et al. [24] as well as Draper and Smith [9]; Mixed selection – which is a mixture of the forward and backward procedures; is the most statistically defendable stepwise regression procedure. As noted by Kutner et al. [21]; model validation is the final step in regression model-building process. Furthermore, it was highlighted therein also that, there are three main methods associated with model validation, as follows:

1) Collection of new data to validate the current model and its predictability.
2) Comparison of current results with other theoretical values, empirical and simulation results.
3) Use of a cross-validation sample to validate and assess the predictive power of the current model.

In this article, the cross-validation approach was used to assess the validity and predictive power of the regression models under review, i.e., a certain amount of the data (about twenty percent) were removed from the model-building process, which are then used in the constructed models to estimate their computed values. As noted by Kutner et al. [21], a general rule of thumb in regression model building is to use 80 percent of the data set for the development of the training model, while the remaining 20 percent can be utilized for validating the model. Validation records can be

selected at random from the entire data set, or in the case of time-series data, the validation set can be the most current 20 percent. See Kutner et al. [21] for details.

Breiman and Friedman [5] observed that, when considering multiple regression models, it is of great importance for the predictors to share strength among different models. Turlach et al. [30] also observed that, it was of particular interest, when there were a large number of covariates, to find a common set of variables that can be used for all models under investigation. In the context of mean regression, Turlach et al. [30] considered the problem of selecting a subset of 770 wavelengths that are suitable as predictors for 14 different but correlated infra-red spectrometry measurements, where they proposed a novel regularization method to perform simultaneous variable selection. This regularization is as a result of the fact that, classical regression approaches require the number of samples to exceed the number of variables, hence cannot be applicable in case of genome wide association (GWA) data. Additionally, least-squares estimates of regression coefficients may be highly unstable, especially in cases of correlated predictor variables, which lead to low prediction accuracy. In genomic settings {such as the prediction of cancer patient survival from tumor gene expression (Beer et al., [2]; Shedden et al., [25]; Sørlie et al., [27]; van de Vijver et al., [32] and Wigle et al., [34]}, where collinear predictors, say p, typically outnumber available sample of size n (i.e. p > n); OLS regression is subject to overfitting and instability of coefficients and as well stepwise variable selection methods do not scale well as observed in the research conducted by Fan and Li [13]. In these type of settings, regression has been successfully adapted to high-dimensional situations by penalization methods (see for instance, Hesterberg [17]), and penalized regression methods have been shown to outperform univariate and other classical multivariate regression methods (Bøvelstad et al. [4]).

Several simulation studies {Adams [1] as well as Hurvich and Tsai [19]} observed that, least-squares estimates could be poor, OLS prediction errors tend to be underestimated and that the usual 95% confidence intervals, quite often, includes the true value of the parameter only in roughly 50% of cases. However, when predictor variables are strongly correlated, the prediction errors were shown to become too large. It is a well-known fact that, OLS often does poorly in both prediction and interpretation, especially when some of the predictor variables are collinear, in view of this, penalization techniques have been proposed to improve on the prediction flaws inherent in OLS. The penalized least squares (PLS) method, which is equivalent to penalized maximum likelihood is found to deal about the issue of multicollinearity by putting constraints on the values of the estimated parameters. A wonderful consequence is that the entries of the variance-covariance matrices are reduced significantly. Hoerl and Kennard [18] introduced the ridge regression which estimates the regression coefficients through an l_2-norm penalized least-squares criterion. Friedman et al. [14] observed that ridge regression shrinks the coefficients of correlated predictor variables, thereby allowing them to borrow strength from each other. However, this behaviour is a little costly, in the sense that, for example, in the case of the k identical predictor variables. They get identical coefficients with size 1/k, which any single one would get if fit alone. The ridge penalty is ideal if there are many predictor variables, and all have non-zero coefficients (from a Bayesian perspective. This can be achieved only if these are drawn from a Gaussian prior distribution). According to Breiman and Friedman [5] best subset selection produces a sparse model, but it is extremely variable because of its inherent discreteness.

Tibshirani [29] proposed the LASSO estimator which estimates the regression coefficients through an l_1-norm penalized least-squares criterion. This is equivalent to minimizing the RSS plus an l_1 penalty on the regression coefficients. Due to the nature of the l_1 penalty, LASSO performs continuous shrinkage and variable selection simultaneously. In addition LASSO possesses the properties of both the l_2 (ridge) penalization and best-subset selection. It was argued that, the automatic feature selection property makes the LASSO a better choice than the l_2 penalization in high dimensional problems, especially when there are lots of redundant noise features (Friedman et al. [15]).

Although the l_2 regularization has been widely used in various learning problems such as smoothing splines (Wahba [33]), the support vector machine and neural networks where the l_2 regularization is called weight decay (Hastie et.al [16]). An l_1 method called basis pursuit was also used in signal processing (Chen et al. [7]). There are many theoretical works to prove the superiority of the l_1 penalization in sparse settings. The LASSO estimator has two desirable properties. Firstly, the nature of regularization used in the LASSO leads to sparse solutions. Secondly, it is also computationally feasible as it was seen in the works of Efron et al. [12] and Friedman et al. [15]. The sparse solutions obtained by using LASSO automatically leads to model selection. In the finite dimensional case, many authors have studied the model-consistency properties of the LASSO and investigated conditions under which the LASSO can recover the true sparsity pattern as in the case of Zhao and Yu [35] as well as Zou and Hastie [36].

Zou and Hastie [36] proposed the Elastic Net penalty which is based on combined penalties of LASSO and ridge regression. The penalty parameter α determines how much weight should be given to either the LASSO or the ridge regression. The Elastic Net with α set to 0 is equivalent to ridge regression; however when α is close to 1, Elastic Net performs much like the LASSO, but removes any degeneracies and odd behaviour caused by high correlations. Bühlmann and van de Geer [6] have shown that analysis with the Elastic Net can result in lower mean squared errors (MSE) than the LASSO and ridge regression when predictor variables are correlated. (Tutz and Ulbricht [31]) also showed that, the Elastic Net produces a higher number of correctly identified influential variables than the LASSO, and has a much lower false-positive rate than ridge regression. Li and Lin [23] introduced the Bayesian Elastic Net. Tan [28] introduced what is known as CAEN regression, which is an extension of Elastic Net regression. Tan [28] also

introduced correlation adjusted regression, which is an extension of ridge regression. There are many penalized regression methods introduced in recent years, but this paper focuses only on LASSO, Elastic Net and CAEN methods. In this article, our main aim is to compare the statistical properties of the LASSO, Elastic Net and CAEN methods using a diabetes dataset. The paper is organized as follows: in the next section, we provide the materials and methodology used; under which some general discussion on the penalized regression methods under consideration is presented. In section 3, we provide some results and discussions on the analyses conducted. Finally in section 4, we offer some concluding remarks.

2. Materials and methods

The data used in this research comes from a study conducted by (Efron et al. [12]); whereby 442 diabetic patients were measured on 10 baseline variables to get a prediction model that measure a disease progression one year after baseline. The 10 baseline variables include Age, Sex, Body Mass Index (BMI), Blood Pressure (BP), and six other blood serum measurements. The data analysis was performed using R package glmnet which utilizes the capabilities of fast cyclical coordinated descent (CCD) algorithm (Friedman et al., 2010).

2.1. Penalized regression

Consider a standard MLR model given by:

$$y = X\beta + \varepsilon \tag{1}$$

Let $y = (y_1, y_2, \dots, y_n)^T$ be the response vector and $X = [X_1 \,|, \dots, |\, X_p]$ be the model matrix, where $X_i = (x_{1i}, \dots, x_{ni})^T$, and $i = 1, \dots, p$ are predictor variables. $\beta = (\beta_0, \dots, \beta_p)$ is a column vector that contains the regression coefficients and ε is a vector of error terms that are assumed to be normally distributed with mean, 0 and variance, σ_ε^2 {i.e. $\varepsilon \sim N(0, \sigma_\varepsilon^2)$}. For models where n > p, the values of the unknown parameters β can be uniquely estimated by minimizing the RSS,

$$RSS = \left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right) \tag{2}$$

After a location and scale transformation, we can assume the response is centered and the predictors are standardized. In general, the PLS is an optimization problem aimed at minimizing Sum of Squares due to Error (SSE) subject to some penalty on the values of the unknown parameters. In a nutshell, one can write PLS as:

Minimize $RSS = \left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right)$

Subject to $Pen(\beta) \leq t$, (3)

Where $Pen(\beta)$ – a specific penalty, is a function of $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p\right)^T$, while t is a tuning parameter.
The constrained optimization problem formulated earlier can be solved using the equivalent Lagrangian formulation which can be achieved by minimizing:

$$\left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right) + \lambda\, Pen(\beta) \tag{4}$$

Where λ is a tuning parameter that controls the strength of shrinkage. For example, when λ = 0, that is when no any penalty is applied we end up with the classical OLS regression. However, when λ gets larger, more weight is given to the penalty term.

2.1.1. LASSO regression

The LASSO penalty regulates the linear regression coefficients through an L_1 PLS procedure: $[i.eP(\lambda, \beta) = \lambda\|\beta\|_{l_1}]$
The objectives is to minimize

$$\hat{\beta}_{LASSO} = \arg\min_{\beta \in R^P} (\mathbb{Y} - \mathbb{X}\beta)^T (\mathbb{Y} - \mathbb{X}\beta) + \lambda \sum_{i=1}^{p} |\beta_i| \tag{5}$$

The resulting regression problem is non-linear in y and results in a convex optimization problem. The regularization parameter λ controls the amount of shrinkage and needs to be tuned or chosen based on some prior result.

2.1.2. Elastic net regression

The Elastic Net (EN) method (Zou and Hastie, 2005) is based on a compromise between the LASSO and ridge regression penalties

$$\hat{\beta}_{EN} = \arg\min_{\beta \in R^P} (\mathbb{Y} - \mathbb{X}\beta)^T \ (\mathbb{Y} - \mathbb{X}\beta) + \sum_{i=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_i^2 + \alpha|\beta_i|\right] \quad (6)$$

Where $0 \leq \alpha \leq 1$ is a penalty weight. The EN with $\alpha = 1$ is identical to the LASSO, whereas it turns out to be ridge regression when $\alpha = 0$ (Friedman et al., 2010). Setting α Close to 1 makes the EN to behave similar to the LASSO, but eliminates problematic behavior caused by high correlations. When α increases from 0 to 1, for a given λ the sparsity of the minimization (i.e., the number of coefficients equal to zero) increases monotonically from 0 to the sparsity of the LASSO estimation.

2.1.3. CAEN regression

Tan (2012) introduced the CAEN regression which is a combination of L_1 penalized regression and Correlation Adjusted Regression. It is also an extension of elastic net regression.
Hence, minimizing,

$$LASSO^* = (\mathbb{Y}^* - \mathbb{X}^*\beta^*)^T(\mathbb{Y}^* - \mathbb{X}^*\beta^*) + \gamma \sum_{i=1}^{p} \left|\beta^*_i\right| \quad (7)$$

Is equivalent to minimizing?

$$CAEN = (\mathbb{Y} - \mathbb{X}\beta)^T(\mathbb{Y} - \mathbb{X}\beta) + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \beta^T W \beta \quad (8)$$

Where $W = D^T D$ and

$$D = \begin{pmatrix} 1 & -r_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & 1 & -r_{2,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -r_{p-1,p} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

Due to quadratic regularization, the solution paths of CAEN are more stable than the solution paths of LASSO. CAEN can also be regarded as a stabilized version of the LASSO.

2.2. Fitting and analyzing models

The whole path of results (in λ) for the LASSO, Elastic Net and CAEN models were calculated using the path wise cyclical coordinate descent (CCD) algorithms – computationally effective techniques for finding out these convex optimization examples– in glmnet in R. We used 10-fold cross validation (CV) within glmnet to entirely search for the optimal λ. A regularized profile plot of the coefficient paths for the three methods was also shown. Predictive accuracy was also assessed using the mean squared error (MSE).

3. Results and discussions

In previous sections, we described the three penalized linear regression methods considered in this article; while in this section, we conduct an analysis using numerical data obtained from previous literature to investigate their individual performances. Penalization techniques can make the predictive error of the model better by lowering the variability in the measures of a regression coefficient by shrinking the estimates toward zero. These three methods will shrink some coefficient estimates to exactly 0, thus supplying a scheme on predictor selection. Regularization plots are plots of the regression coefficients versus the regularization penalty. The analysis was performed using R statistical package with the aid of glmnet library.

3.1. Results on elastic net regression

The following Figure (Fig. 1) gives the relationship between ln λ and MSE. The integers at the top show the number of non-zero estimators for the model. The left line gives the smallest MSE with eight variables in the model, and the right line gives the smallest Standard Error (SE) with only seven variables in the model. The Elastic Net was calculated based on an optimal value of $\alpha = 0.16$.

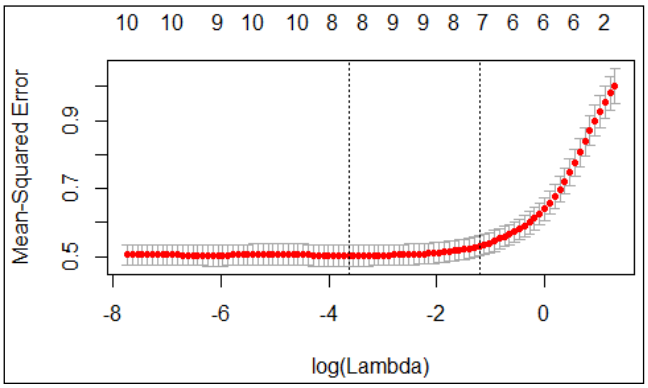


Fig. 1: MSE Plot and the Number of Variables in the Model as A Function of Log (λ) for the 10-Fold Cross Validation.

3.2. Results on LASSO regression

Figure 2 illustrates the relationship between ln λ and MSE. The integers at the top show the number of non-zero estimators for the model. The left line gives the smallest MSE with seven variables in the model, and the right line gives the smallest Standard Error (SE) with only four variables in the model. The LASSO regression model was obtained based on an optimal value of λ = 0.0129.

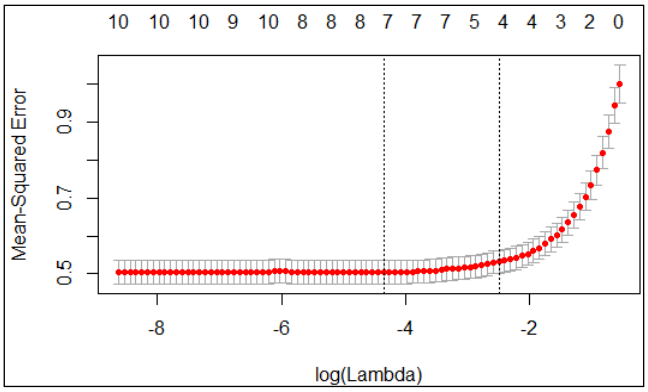


Fig. 2: MSE Plot and the Number of Variables in the Model as A Function of Log (λ) for the 10-Fold Cross Validation.

3.3. Results on CAEN regression

Figure 3 shows the graph of the relationship between ln λ and MSE. The integers at the top show the number of non-zero estimators for the model. The left line gives the smallest MSE with seven variables in the model, and the right line gives the smallest Standard Error (SE) with only four variables in the model. The CAEN regression model was obtained based on an optimal value of λ1 = 0.01417 and λ2 = 0.02.

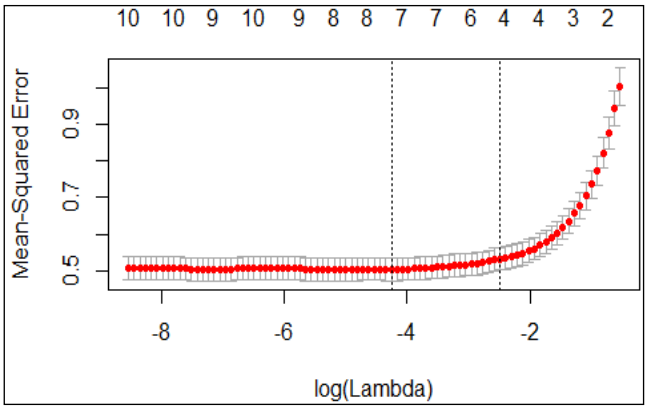


Fig. 3: MSE Plot and the Number of Variables in the Model as A Function of Log (λ) for the 10-Fold Cross Validation.

4. Main results

The following table shows the comparison between the models generated by the three penalized regression methods, namely: LASSO, Elastic Net and CAEN.

Table 1: Comparison for OLS, LASSO, Elastic Net and CAEN Regression Methods

Variable	OLS	LASSO	Elastic Net	CAEN
INT	0.0000	0.0000	0.0000	0.0000
AGE	-0.0062	0.0000	0.0000	0.0000
SEX	-0.1481	-0.1211	-0.1342	-0.8404
BMI	0.3211	0.3225	0.3189	2.2779
BP	0.2004	0.1830	0.1907	1.2857
TC	-0.4893	-0.0630	-0.1010	-0.4265
LDL	0.2945	0.0000	0.0000	0.0000
HDL	0.0624	-0.1379	-0.1078	-0.9701
TCH	0.1094	0.0000	0.0513	0.0000
LTG	0.4641	0.3173	0.3151	2.2322
GLU	0.0418	0.0333	0.0423	0.2282
MSE	0.5050	0.5040	0.5034	0.5037
Df	10	7	8	7

It can easily be observed from Table 1 that the OLS regression model includes all the 10 variables in the diabetes data set. The LASSO regression model included seven variables with an MSE value of 0.5040. Elastic Net regression model included eight variables with an MSE value of 0.5034; while the CAEN regression model produced an MSE value of 0.5037 with seven variables.
The resultant models for the compared regression models are as follows:
OLS Regression

$$y = -0.0062Age - 0.1481Sex + 0.3211Bmi + 0.2004Bp - 0.4893Tc + 0.2945Ldl + 0.0624Hdl + 0.1094Tch + 0.4641Ltg + 0.0418Glu$$

LASSO Regression

$$y = -0.1211Sex + 0.3225Bmi + 0.1830Bp - 0.0630Tc - 0.1379Hdl + 0.3173Ltg + 0.0333Glu$$

Elastic net Regression

$$y = -0.1342Sex + 0.3189Bmi + 0.1907Bp - 0.1010Tc - 0.1078Hdl + 0.0513Tch + 0.3151Ltg + 0.0423Glu$$

CAEN Regression

$$y = -0.8404Sex + 2.2779Bmi + 1.2857Bp - 0.4265Tc - 0.9701Hdl + 2.2322Ltg + 0.2282Glu$$

5. Conclusion

In this article, we studied and examined the performance of LASSO, Elastic Net and CAEN regression methods against the classical OLS regression. We presented characteristics of these penalized methods through MSE plots. We also compared the predictive performance of these methods, using their numerical results. The CAEN generates a less complex model at the same value of MSE with the Elastic Net that has eight variables remaining.

References

[1] Adams, J., "A computer experiment to evaluate regression strategies", Proceedings of the Statistical Computing Section, American Statistical Association, (1990), pp: 55 - 62.

[2] Beer, D. G., Kardia, S. L., Huang, C. C, Giordano, T. J., Levin, A. M., "Gene-expression profiles predict survival of patients with lung adenocarcinoma", Nat. Med., 8, (2002), pp: 816 – 824. http://dx.doi.org/10.1038/nm733.

[3] Bornn, L., Gottardo, R., and Doucet, A., "Grouping priors and the Bayesian elasticnet", Technical Report 254, Department of Statistics. University of British Columbia, (2010).

[4] Bøvelstad, H. M., Nygard, S., Storvold, H. L., Aldrin, M., Borgan, O., Frigessi, A., Lingjarde, O. C., "Predicting survival from microarray data a comparative study", Bioinformatics, 23, (2007), pp: 2080 – 2087. http://dx.doi.org/10.1093/bioinformatics/btm305.

[5] Breiman, L., Friedman, J., "Predicting multiple responses in multiple linear regression (with discussion)", Journal of the Royal Statistical Society: Series B59, (1997), pp: 3 – 54. http://dx.doi.org/10.1111/1467-9868.00054.

[6] Bühlmann, P., van de Geer, S., Statistics for High-Dimensional Data: Methods, Theory and Applications, Springer-Verlag, NewYork, (2011), pp: 97 – 115.

[7] Chen, H. Y., Yu, S. L., Chen, C. H., "A five gene signature and clinical outcome in non–small cell lung cancer", N. Engl. J. Med., 356, (2007), pp: 11 – 20. http://dx.doi.org/10.1056/NEJMoa060096.

[8] Cho, S., Kim, K., Lee, J. K., "Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis", Ann. Hum. Genet. 74, (2010), pp: 416 – 428. http://dx.doi.org/10.1111/j.1469-1809.2010.00597.x.

[9] Draper, N. R., Smith, H., Applied Regression Analysis, 2nd Ed. John Wiley and Sons, Inc. New York, (1981), pp: 75 – 95.

[10] Efromyson, M. A., Multiple Regression Analysis. Mathematical Methods for Digital Computers, John Wiley and Sons, Inc. NewYork, (1960), pp: 65 - 79.

[11] Efron, B., Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction, Cambrige University Press, Cambrige, UK, (2010), pp: 46 – 67. http://dx.doi.org/10.1017/CBO9780511761362.005.

[12] Efron, B., Turnbull, B. B., Narasimhan, B., Locfdr:Computes Local False Discovery Rates. R-packageVersion1.1-7, (2011), Available online: http://CRAN.R- project.org/package=locfdr.

[13] Fan, J., Li, J., "A selective overview of variable selection in high dimensional feature space", Stat.Sin, 20, (2010), pp: 101 – 148.

[14] Friedman, J., Hastie, T., Tibshirani, R., "Regularization paths for generalized linear models via coordinate descent", Journal of Statistical Software, 33, (2010), pp: 1 – 22.

[15] Friedman, J., Hastie, T., Hoefling, H., Tibshirani, R. "Pathwise Coordinate Optimization", Annals of Applied Statistics, 2, (2007), pp: 302 – 332. http://dx.doi.org/10.1214/07-AOAS131.

[16] Hastie, T. R., Tibshirani, R., Friedman, J., Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition, Springer-Verlag, NewYork, (2009), pp: 37 - 71.

[17] Hesterberg, T., Choi, N. H., Meier, L., Fraley, C., "Least angle and L1 penalized regression: a review", Statistical .Survey, 2, (2008), pp: 61 – 93. http://dx.doi.org/10.1214/08-SS035.

[18] Hoerl, A. E., Kennard, R., "Ridge regression: biased estimation for non-orthogonal problems", Technometrics, 12, (1970), pp: 55 – 67. http://dx.doi.org/10.1080/00401706.1970.10488634.

[19] Hurvich, C., Tsai, C., "The impact of model selection on inference in linear regression", American Statistician, 44, (1990), pp: 214 - 217. http://dx.doi.org/10.2307/2685338.

[20] Kooperberg, C., LeBlanc, M., Obenchain, V., "Risk prediction using genome-wide association studies", Genet. Epidemiol., 34, (2010), pp: 643 – 652. http://dx.doi.org/10.1002/gepi.20509.

[21] Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W., Applied linear statistical models (5th edition), McGraw-Hill/Irwin, New York, (2005), pp: 67 - 83.

[22] Kyung, M., Gill, J., Ghosh, M., Casella, G., "Penalized regression, standard errors, and Bayesian Lassos", Bay.Anal., 5, (2010), pp: 369 – 412. http://dx.doi.org/10.1214/10-BA607.

[23] Li, Q., Lin, N., "The Bayesian elasticnet", Bay.Anal, 5, (2010), pp: 151 – 170. http://dx.doi.org/10.1214/10-BA506.

[24] Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., Applied Linear Regression Models. 3rd Ed. McGraw-Hill/Irwin, Chicago, IL, (1996), pp: 49 - 87.

[25] Shedden, K., Taylor, J. M., Enkemann, S. A., "Gene expression-based survival prediction in lung adenocarcinoma: a Multi-site, blinded validation study", Nat. Med., 14, (2008), pp: 822 – 827. http://dx.doi.org/10.1038/nm.1790.

[26] Shieh, G., "Suppression situations in multiple linear regressions", Educational and Psychological Measurement, (2006), pp: 435 - 447. http://dx.doi.org/10.1177/0013164405278584.

[27] Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen M. B., Van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., "Gene expression patterns of breast carcinomas distinguish tumor Subclasses with clinical implications", Proc. Natl Acad. Sci. USA, 98, (2001), 10869–10877. http://dx.doi.org/10.1073/pnas.191367098.

[28] Tan, Q., Correlation Adjusted Penalization in Regression Analysis. PhD Thesis, (2012), Department of Statistics, University of Manitoba.

[29] Tibshirani, R., "Regression shrinkage and selection via the lasso". Journal of Royal Statistical Society, B58, (1996), pp: 267 – 288.

[30] Turlach, B., Venables, W., Wright, S., "Simultaneous variable selection", Technometrics, 47, (2005), pp: 349 – 363. http://dx.doi.org/10.1198/004017005000000139.

[31] Tutz, G., Ulbricht, J., "Penalized regression with correlation-based penalty", Statistical Computing, 19, (2009), pp: 239 – 253. http://dx.doi.org/10.1007/s11222-008-9088-5.

[32] Van de Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., "A gene-expression signature as a predictor of survival in breast cancer", N. Engl. J. Med., 347, (2002), pp: 1999 – 2009. http://dx.doi.org/10.1056/NEJMoa021967.

[33] Wigle, D. A., Jurisica, I., Radulovich, N., Pintilie, M., Rossant, J., Liu, N., Lu, C., Woodgett, J., "Molecular profiling of non-small cell lung cancer and correlation with disease-free survival", Cancer Res., 62, (2002), pp: 3005 – 3008.

[34] Wahba, G., "Splines models for observational data", SIAM CBMS-NFS regional conference in applied mathematics, V.59, (1990).

[35] Zhao, P., Yu, B., "On model selection consistency of Lasso", Journal of Machine Learning Research, 7, (2006), pp: 2541 - 2563.

[36] Zou, H., Hastie, T., "Regularization and variable selection via the elastic net", Journal of Royal Statistical Society, B67, (2005), pp: 301 - 320. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x