

6-8-2021

VIF-Regression Screening Ultrahigh Dimensional Feature Space

Hassan S. Uraibi

University of Al-Qadisiyah, hssn.sami1@gmail.com

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Uraibi, H. S. (2020). VIF-Regression Screening Ultrahigh Dimensional Feature Space. *Journal of Modern Applied Statistical Methods*, 19(1), eP2916. <https://doi.org/10.22237/jmasm/1608553020>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

VIF-Regression Screening Ultrahigh Dimensional Feature Space

Hassan S. Uraibi
Univ. of Al-Qadisiyah
Al Diwaniyah, Iraq

Iterative Sure Independent Screening (ISIS) was proposed for the problem of variable selection with ultrahigh dimensional feature space. Unfortunately, the ISIS method transforms the dimensionality of features from ultrahigh to ultra-low and may result in unreliable inference when the number of important variables particularly is greater than the screening threshold. The proposed method has transformed the ultrahigh dimensionality of features to high dimension space in order to remedy of losing some information by ISIS method. The proposed method is compared with ISIS method by using real data and simulation. The results show this method is more efficient and more reliable than ISIS method.

Keywords: VIF-Regression, ISIS, screening, feature selection, high dimensional data

Introduction

The rapid evolution of communication networks and information technology has allowed us to collect millions of data with massive features, especially in gene expression and microarray data related fields, as well as in financial sciences, machine learning, computer sciences and other multidisciplinary scientific fields (Algamal & Lee, 2015; Algamal, Lee, & Al-Fakih, 2016; Zhang, Fu, Jiang, & Yu, 2007; Zheng & Liu, 2011). This type of data forms a big challenge to researchers for providing statistical tools with the ability to deal with such data. It is well known that regression feature selection targets a subset of important features which satisfies a determined criterion (Alhamzawi & Algamal, 2018). Tools for analyzing high dimensional data include Lasso (Tibshirani, 1996), SCAD (Fan & Li, 2001), LARS (Efron, Hastie, Johnstone, & Tibshirani, 2004), Elastic-Net (Zou & Hastie, 2005), adaptive Elastic-Net (Zou & Zhang, 2009), and Adaptive Lasso (Zou, 2006).

Data collection of unprecedented size and complexity has become easier due to the daily growth of computing power and its technologies that allowed to appear ultrahigh dimensional features in which the number of predictors or features (p) extremely exceeds the number of observations (n), $p \gg n$ (Algamal, Lee, Al-Fakih, & Aziz, 2015; Fan & Lv, 2010). Unfortunately, the high dimensional methods mentioned above cannot apply directly with this type of data because they are time consuming and they lead to loss of statistical inference accuracy and algorithmic stability.

Jianqing Fan and Jinchi Lv (2008) suggested sure independence screening (SIS) which is a very effective procedure to tackle the ultrahigh dimensional feature problem. In the context of least squares regression, SIS algorithm starts with very simple procedure called screening. Screening means ranking features that have the best marginal correlation with the response, and then picking up the top features that indexed from the first rank to the feature that has been ranked at d where $d = n/\log(n)$. Lasso or SCAD can be applied in the second stage to select the important features among d of them.

Jianqing Fan and Jinchi Lv (2008) pointed out that regularity conditions may fail with SIS in some cases, so they Iterated SIS (ISIS) using subsamples procedure to process these cases. Subsampling performs false selection rate for controlling inclusion noise variables in the second stage of SIS. An important Lasso limitation is it selects at most n features where $p > n$. But, it is remarkable that the key idea of SIS is forcing the dimensionality p to be less than the sample size n . It means the screening feature transforms the dimensionality feature space from ultrahigh to ultra-low. As a result of this procedure, selection of important features cannot exceed d , where $d < n$ and $d \ll p$, consequently SIS algorithm has a problem like lasso's problem. The second stage ISIS works by iteratively performing Lasso or SCAD algorithm to recruit z features where $z < d$.

Sure independence screening method

Consider a dataset $\left\{ (y_i, \mathbf{x}_i) \right\}_{i=1}^n$, where y_i is a response variable and $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents a p -dimensional explanatory variable vector. Without loss of generality, it is assumed that the response variable is centered and the explanatory variables are standardized.

Consider the following linear regression model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (1)$$

VIF-REGRESSION FOR ULTRAHIGH DIMENSIONAL FEATURE SPACE

where \mathbf{y} is an $n \times 1$ vector of observations of the response variable, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ is an known design matrix of $n \times p$ explanatory variables, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)^T$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with mean 0 and variance σ^2 . Using OLS method, the parameter estimation of (1) is given by

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2)$$

The OLS estimator is unbiased and it has minimum variance among all linear unbiased estimators. In practice, for multiple linear regression model, this model contains irrelevant or noisy variables leading to low performance with less prediction precision. As a consequence, analyzing variables in terms of their importance has become a necessary task. The sure independence screening method (SIS) has shown excellent performance in reducing high dimensionality (Fan & Lv, 2008). To improve the prediction stability and to expedite computation time Fan and Lv (2008) proposed SIS procedure for ultrahigh-dimensional linear models. They utilized the Pearson correlation to rank the importance of each variable. An advantage of SIS is it has the sure screening property which can retain all truly important variables with probability tending one (Fan & Lv, 2008; Fan & Song, 2010).

When X and Y are normalized, the correlations between the response variable and explanatory features are equal to the regression coefficients estimate. This advantage can be very useful to rank the absolute value of these correlations from the maximum to minimum ones. Where the important feature correlated with noise feature, the absolute value of marginal correlation between noise feature and the response variable will be ranked after the lowest absolute correlation between important feature and response. To illustrate, suppose, X_1, \dots, X_p standardized features and only X_1, \dots, X_k has the best contribution with corresponding Y where $k < p$. The absolute value of the marginal correlation $\hat{R}_{j=1, \dots, p}^*$ between Y and X_1, \dots, X_p allows to sequence the estimated correlations from the maximum to minimum ones. Assume that, the feature X_1 is strongly correlated with the noise feature X_{k+i} , $i > k$ where the sequencing list of marginal correlations should be such that $|\hat{R}_1^*| > |\hat{R}_2^*| > \dots, |\hat{R}_k^*| > |\hat{R}_{k+i}^*| > \dots, > |\hat{R}_p^*|$ which begins from the largest and ends with the lowest value. Postulate that $k + i < d < n < p$; the screening feature of SIS will pick up d features which suffer from collinearity problem. Feature selection

may include the noise feature to the right model dependent on the degree of strong relation between X_1 feature and X_{k+i} .

Lin, Foster, and Ungar (2011) pointed out high dimensional penalized methods introduce bias estimates, therefore they proposed attracted approach to correct this bias by pre-sampling a small set of data to compute the variance inflation factor (VIF) of each variable. The VIF-regression is a non-greedy version of forward selection that combined with α -investing rule Zhou et al. (2006). A significant level (α) is iteratively adjusted p -value for multiple hypotheses testing (Foster & Stine, 2008). This rule is based on one of gamble kinds in which the gambler begins with initial wealth w_0 and he can earn Δw when the null hypothesis $H_0 : \beta_j = 0$ of j^{th} test is rejected. On the contrary, the gambler losses an $\alpha/(1 - \alpha)$, where $\alpha = w_j/(1 + j - f)$ and f is the time of hypothesis was rejected. Unfortunately, the performance of VIF-Regression is very slow for ultrahigh dimensional features. This paper aims to suggest an approach for improving the performance of VIF-Regression for ultrahigh dimensional features. However there are several practical questions with the with screening feature:

1. Does the screening procedure remove at least one important feature when the number of important features really exceeds d ?
2. Does the performance of ISIS exceed d when the number of samples is small and the number of important features is not?
3. Does the ISIS method select the most important feature in the presence of multicollinearity problem?

To answer these questions, a two-step approach is presented. First, reducing the ultrahigh dimension to high dimension without loss, by suggesting some cut-off points instead of screening feature threshold. The second step is to answer the remaining questions by using VIF-Regression (Lin et al., 2011) for the reduced space.

The proposed method: VIF Regression Screening (VIFRegS) Algorithm

Let X' and Y' be the standardized form of X and Y respectively, that to find the positive values of their marginal correlations $|\hat{R}_j^*|$. The VIFRegS algorithm can be described in the following steps:

Step I: Screening Features

Sequencing features based on their corresponding positive correlations from the largest to lowest value, $|\hat{R}_1^*| > |\hat{R}_2^*| > \dots, |\hat{R}_{d_1^*}| > \dots, > |\hat{R}_{d_2^*}| > \dots, > |\hat{R}_{d_3^*}| > \dots, > |\hat{R}_p^*|$,

where $d_1^* = n$, $d_2^* = n + \left(\frac{n}{\log(n)}\right)$ and $d_3^* = \frac{p}{\sqrt{p}}$. We consider d_2^* and d_3^* as two

proposals of thresholds can be used to reduce the dimensionality from ultrahigh to low, moderate and high dimension respectively. Only d_j^* features will construct the new matrix design X^* , where $X^* = \left\{ X_j : |\hat{R}_j^*| \geq |\hat{R}_{d_j^*}^*| \right\}$.

Step II: VIF-Regression

To take full advantage of VIF-Regression speed, Lin et al. (2011) pointed out that avoiding selecting the optimal parameters in the VIF regression technique will consume more time in selecting the best model. Thus, they consider using larger parameters such as ($w_0 = 0.5$) for initial wealth parameter and small value such as ($\Delta w = 0.05$), for investment parameters. However, the VIF regression as follows,

1. Centered y , and X_j^* where $j = 1, \dots, d$
2. Suppose the algorithm starts with the following initial information
 - $w_0 = 0.5$, $\Delta w = 0.05$, and subsample size m ;
 - model = $\{\emptyset\}$; $\varepsilon = y - \bar{y}$, $\hat{\sigma} = sd(y)$
 - Set $j = 1$, $w_1 = w_0$, $f = 0$
3. Repeat the following procedure d times
 - Set significant level $\alpha_j = w_j / (1 + j - f)$

- Find $\hat{\gamma}_j = \langle \varepsilon, X_j^* \rangle / \|X_j^*\|$ to correct t -value
- Randomly draw small subsample from original data $\zeta = \{1, \dots, m\} \subset \{1, \dots, n\}$
- Fit $X_j^{*\zeta}$ on $\{X_c^{*\zeta}\}$ where $c = 1, \dots, k$ and $k \neq j$, so that k is the number of features that were selected to be in the model
- Compute $R_\zeta^2 = X_j^{*\zeta} X_c^{*\zeta} \left((X_c^{*\zeta})' X_c^{*\zeta} \right)^{-1} (X_c^{*\zeta})' X_j^* / \|X_j^*\|^2$
- The corrected t -value is $t_j = \hat{\gamma}_j / \hat{\sigma} \left(\sqrt{1 - R_\zeta^2} \right)$
- If the CDF of normal distribution, $2\Phi(|t_j| > 1 - \alpha)$ then
 - a. model = model $\cup \{j\}$ // add feature to model
 - b. update $\varepsilon = y - \hat{y}_{\text{model}}$, $w_{j+1} = w_j + \Delta w$, $f = j$
- Else $w_{j+1} = w_j - \alpha / (1 - \alpha)$
- Next j until reach to d or the maximum time of computer processor is reached.

Numerical Example

Eye data were taken from Scheetz et al. (2006) to assess the performance of the VIFRegS method. The data were sampled from 120 rats to know the effectiveness of 200 genes on expression level of TRIM32 gene. The result of VIFRegS rely on d_2^* and d_3^* respectively are compared with ISIS. The non-zero genes coefficients selected by each method are tested using multiple hypothesis test at 0.05 significant level. The adjusted determinant coefficients square (R_{adj}^2) and the residual standard error of selected model (RSE) respectively are considered as comparison criterion.

The best method is that one which possesses the highest R_{adj}^2 and lowest RSE. Table 1 presents the results of regressing expression level of TRIM32 gene with genes that have been selected (SG) by three methods. The least squares method is used to estimate Gene's Coefficients (Ceof.Est) as a multiple linear regression model (MLP). The analysis of MLP permits testing each Ceof.Est individually and its standard error and p value which are denoted as CSE and p -value respectively. As shown in Table 1 ISIS method selects three genes (G87, G153 and G07) that is

totally different selection from VIFRegS $_{d_2}^*$ and VIFRegS $_{d_3}^*$ methods which are selected (G196, G181 and G136). It has been found that the CSE of G87 is very close to CSE of G196 while the CSE of G181 and G136 are more accurate than the CSE of G153 and G07.

Even though p values of G87, G153 and G07 are less than 0.05, the selection of VIFRegS $_{d_2}^*$ and VIFRegS $_{d_3}^*$ show more significant levels. Finally, genes that possess the significant Ceof.Est will be in the final model and R_{adj}^2 and RSE should be crucial to determine which model is the best one. As may be seen in Table 1, the selected genes by VIFRegS $_{d_2}^*$ and VIFRegS $_{d_3}^*$ methods explain about $R_{adj}^2 = 0.61$ from expression level of TRIM32 gene while R_{adj}^2 of ISIS explains only 0.53 of it. There is a good match between VIFRegS $_{d_2}^*$ and VIFRegS $_{d_3}^*$ results. There is an evident relationship in genes selection between our proposed method and ISIS method, although each method is totally selected different genes. It is found that $Cor(G87, G136) = 0.82$, $Cor(G153, G196) = 0.73$ and $Cor(G07, G181) = 0.73$.

Table 1. Selected gene, coefficient estimate, coefficient standard error, p -value, R_{adj}^2 and residual standard error of three methods for eye data.

Method	SG	Ceof.Est	CSE	p -value	R_{adj}^2	RSE
ISIS	G87	-0.1006	0.0327	0.0026	0.5275	0.0992
	G153	-0.1833	0.0468	0.0002		
	G07	0.0713	0.0376	0.0605		
VIFRegS $_{d_2}^*$	G196	-0.1429	0.0326	2.64E-05	0.6048	0.0909
	G181	0.1672	0.0348	4.54E-06		
	G136	0.0610	0.0306	0.0275		
VIFRegS $_{d_3}^*$	G196	-0.1429	0.0326	2.64E-05	0.6048	0.0909
	G181	0.1672	0.0347	4.54E-06		
	G136	0.0610	0.0306	0.0275		

Simulation

To investigate the performance of VIF-Regression screening compared with ISIS method, four simulation studies are considered based on collinearity structure of design matrix X . In the first, second and third collinearity structures p features are generated from a centered multivariate normal distribution with covariance $\tau^2 \Sigma_{j,k} = \rho^{|j-k|}$, where $j \neq k$, $\rho = 0$, $\rho = 0.5$, $\rho = 0.9$ respectively and $\tau^2 = 0.1$. For the fourth simulation study, features are generated similar to the previous simulation

one's except the covariance is structured to be $\Sigma_{j,k} = 0.5$ where $j \neq k$ and $\Sigma_{j,k} = 1$ where $j = k$. Only $n/\log(n) + c$ features are picked to generate the response variable Y where c is a small positive integer number such that $\{n/\log(n) + c\} > n$ using multiple linear regression model with the following relation:

$$Y = X_1 + X_2 + \dots + X_L + \delta(X_{L+1} + \dots + X_p) + e$$

where $\delta = 1.5$ is signal to noise ratio and $L = n/\log(n) + c$, here it is assumed $c=2$.

Table 2. Percentage of including true features of ISIS, VIFRegS $_{d_2}^*$ and VIFRegS $_{d_3}^*$ over 5000 simulated data set with sample size equals to $n = 20, 30, 40$ and 50 .

	ρ	Method	Sample size / PTF			
			$n = 20$	$n = 30$	$n = 40$	$n = 50$
$\rho = 0$	100	ISIS	0.10	0.09	0.80	0.85
		VIFRegS $_{d_2}^*$	0.35	0.81	0.91	0.89
		VIFRegS $_{d_3}^*$	0.29	0.76	0.81	0.71
	1000	ISIS	0.02	0.02	0.05	0.21
		VIFRegS $_{d_2}^*$	0.13	0.65	0.79	0.82
		VIFRegS $_{d_3}^*$	0.13	0.64	0.79	0.82
$\rho = 0.5$	100	ISIS	0.19	0.55	0.57	0.80
		VIFRegS $_{d_2}^*$	0.38	0.77	0.77	0.88
		VIFRegS $_{d_3}^*$	0.28	0.68	0.66	0.71
	1000	ISIS	0.27	0.36	0.42	0.60
		VIFRegS $_{d_2}^*$	0.15	0.45	0.75	0.83
		VIFRegS $_{d_3}^*$	0.14	0.44	0.74	0.83
$\rho = 0.9$	100	ISIS	0.18	0.22	0.27	0.25
		VIFRegS $_{d_2}^*$	0.21	0.39	0.58	0.61
		VIFRegS $_{d_3}^*$	0.08	0.30	0.50	0.53
	1000	ISIS	0.19	0.25	0.25	0.13
		VIFRegS $_{d_2}^*$	0.04	0.22	0.34	0.53
		VIFRegS $_{d_3}^*$	0.04	0.20	0.30	0.50
$\rho^{l-k} = 0.5$	100	ISIS	0.13	0.31	0.60	0.85
		VIFRegS $_{d_2}^*$	0.25	0.69	0.86	0.92
		VIFRegS $_{d_3}^*$	0.20	0.59	0.78	0.71
	1000	ISIS	0.00	0.04	0.34	0.42
		VIFRegS $_{d_2}^*$	0.12	0.63	0.82	0.80
		VIFRegS $_{d_3}^*$	0.12	0.62	0.82	0.79

VIF-REGRESSION FOR ULTRAHIGH DIMENSIONAL FEATURE SPACE

Each simulation study contains 20 models which are characterized by (p, n, ρ) with $p = 100, p = 1000, n = 10, 20, 30, 40, 50, \rho = 0, 0.5, 0.9$ and $\rho^{|j-k|} = 0.5$ where $j \neq k$. For each model, 5,000 samples are generated to know the performance of ISIS, VIF-Regression screening. The ISIS is applied to select $n/\log(n)$ feature, while VIF-Regression screening with d_1^* and d_2^* cut-off points which denoted as $\text{VIFRegS}_{d_2^*}$ and $\text{VIFRegS}_{d_3^*}$ methods are applied to select d_1^* and d_2^* feature respectively. Then test their accuracy in including the true model. The percentage of models being are selected exhibited in the Table 2. The best method is one that has the highest percentage of including the true features (PITF) with L features over 5,000 samples.

Displayed in Table 2 are the result of simulation study for ISIS, $\text{VIFRegS}_{d_2^*}$ and $\text{VIFRegS}_{d_3^*}$ methods that were performed based on simulation design aforementioned. The result shows the ISIS method failed to get high percentage of including true features in most cases. For instance it gets PITF = 0.10 and 0.09 where the sample size equals to $n = 20$ and 30 observations respectively, and irrespective the dimensionally vector $p = 100$ and 1000 or the correlation structure, $\rho = 0, 0.5$ and 0.90. When $n = 40$ and 50 the ISIS method selection was impacted by the correlation structure between features. The results are summarized in Table 2, the PITF of $\text{VIFRegS}_{d_2^*}$ method with correlation structures ρ is higher than others. On the other hand PITF of $\text{VIFRegS}_{d_3^*}$ is better than ISIS method.

Conclusion

The purpose of this study is to draw attention to the certain weakness in ISIS performing with small samples. A new, reliable alternative method was proposed that possess as an ability to overcome this problem. Two thresholds d_2^* and d_3^* were proposed to reduce the dimensionality from ultrahigh to high dimension. The $\text{VIFRegS}_{d_2^*}$ method has shown a very high stability selection and outperforms on ISIS and $\text{VIFRegS}_{d_3^*}$ methods respectively.

An example with real data includes unexpected events that occurs when two important features are highly correlated, the $\text{VIFRegS}_{d_2^*}$ and $\text{VIFRegS}_{d_3^*}$ methods selected the most important one. By returning to the result of real data that has been reported in Table 1 the model that selected by $\text{VIFRegS}_{d_2^*}$ is similar to $\text{VIFRegS}_{d_3^*}$ model and both are better than ISIS model even though some feature are correlated. In summary, the $\text{VIFRegS}_{d_2^*}$ method is more efficient than ISIS, and it is more stable

than VIFReg $S_{d_3}^*$ method. However, the VIFReg $S_{d_2}^*$ method can be readily used in practice for ultrahigh feature space and small sample size.

References

- Algamal, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, 42(23), 9326-9332. <https://doi.org/10.1016/j.eswa.2015.08.016>
- Algamal, Z. Y., Lee, M. H., & Al-Fakih, A. M. (2016). High-dimensional quantitative structure-activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression. *Journal of Chemometrics*, 30(2), 50-57. <https://doi.org/10.1002/cem.2766>
- Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. (2015). High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO. *Journal of Chemometrics*, 29(10), 547-556. <https://doi.org/10.1002/cem.2741>
- Alhamzawi, R., & Algamal, Z. Y. (2018). Bayesian bridge quantile regression. *Communications in Statistics - Simulation and Computation*, 48(3), 944-956. <https://doi.org/10.1080/03610918.2017.1402042>
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407-499. <https://doi.org/10.1214/009053604000000067>
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360. <https://doi.org/10.1198/016214501753382273>
- Fan, J., & Lv, J. (2008). Sure independence screening for ultra-high-dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849-911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- Fan, J., & Lv, J. (2010). A selective overview of variable selection. *Statistica Sinica*, 20, 101-148.
- Fan, J., & Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6), 3567-3604. <https://doi.org/10.1214/10-aos798>
- Foster, D. P., & Stine, R. A. (2008). α -investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society: Series B*

VIF-REGRESSION FOR ULTRAHIGH DIMENSIONAL FEATURE SPACE

(*Statistical Methodology*), 70(2), 429-444. <https://doi.org/10.1111/j.1467-9868.2007.00643.x>

Lin, D., Foster, D. P., & Ungar, L. H. (2011). VIF regression: a fast regression algorithm for large data. *Journal of the American Statistical Association*, 106(493), 232-247. <https://doi.org/10.1198/jasa.2011.tm10113>

Scheetz, T. E., Kim, K-Y. A., Swiderski, R. E., Philp, A. R., Braun, T. A., Knudtson, K. L., et al. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences*, 103(39), 14429-144346. <https://doi.org/10.1073/pnas.0602562103>

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

Zhang, C., Fu, H., Jiang, Y., & Yu, T. (2007). High-dimensional pseudo-logistic regression and classification with applications to gene expression data. *Computational Statistics & Data Analysis*, 52(1), 452-470. <https://doi.org/10.1016/j.csda.2006.12.033>

Zheng, S., & Liu, W. (2011). An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Computers in Biology and Medicine*, 41(11), 1033-1040. <https://doi.org/10.1016/j.combiomed.2011.08.011>

Zhou, J., Foster, D. P., Stine, R. A. and Ungar, L. H. (2006). Streamwise feature selection. *Journal of Machine Learning Research*, 7, 1861-1885.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418-1429. <https://doi.org/10.1198/016214506000000735>

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Methodological)*, 67(2), 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37(4), 1733-1751. <https://doi.org/10.1214/08-aos625>