Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{\text{Online}}$: 2312-9883
QJAE, Volume 24, Issue 4 (2022) ISSN$_{\text{Print}}$: 1816-9171

Manuscript ID:

Support Vector Machine Variable Selection with Lasso and Group lasso

Prof. Dr. Tahir R. Dikheel* , Zahraa K. Aswad

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq

Abstract:

The support vector machine (SVM) is a binary classification approach that is both accurate and flexible. It has had significant success, but if too many variables are added, its performance might decrease. The lasso method penalizes least squares regression by adding the absolute values of the coefficients (ℓ1-norm). The structure of this penalty encourages sparse solutions (with many variables coefficients equal to 0). The major goal of group lasso is to construct the lasso, the group formula, in order to find the common elements of groups. The simulation shows that group lasso method outperforms the lasso.

Keywords: Support vector machine; Variable selection; lasso; group lasso.

الخلاصة

آلة متجه الدعم (SVM) هي طريقة تصنيف ثنائية تتسم بالدقة والمرونة. لقد حققت نجاحًا كبيرًا ، ولكن إذا تمت إضافة العديد من المتغيرات ، فقد ينخفض أداؤها. طريقة lasso تعاقب على انحدار المربعات الصغرى بإضافة القيم المطلقة للمعاملات (-ℓ1معيار). يشجع هيكل هذه العقوبة على حلول متفرقة (مع العديد من معاملات المتغيرات التي تساوي 0). الهدف الرئيسي للمجموعة lasso هو بناء lasso بصيغة المجموعة ، من أجل إيجاد العناصر المشتركة للمجموعات. توضح المحاكاة أن طريقة Lasso group تتفوق على lasso .

Introduction:

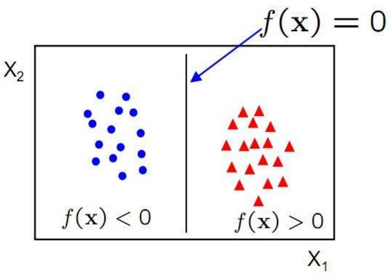
In light of the technological development and dealing with the problems raised electronically led to a more accurate summary of the required information, prompting researchers to focus on the electronic mechanism in their academic studies and research. One of the topics focused on the electronic mechanism is the technique of statistical classification of specific phenomenon data, which relates to the classification of vocabulary to their indigenous communities according to a set of statistical methods. Among these methods is the support vector method and the discriminatory analysis function method, which focuses on the concept of correctly classifying new views with the lowest possible classification line. SVM is a strong binary classification technique developed by (Vapnik, 1996) excellent precision and adaptability. It has found success in a variety of applications. However, one significant disadvantage of the conventional SVM is that its performance might suffer as a result. If the decision rule has a large number of redundant variables (Friedman et al., 2001). Variable selection is crucial in the construction of a support vector machine. This approach gives shrinkage for appropriate estimating parameters, good production, and identification of the key variables. The supply of interpretable models distinguishes statistical techniques for variable selection. Variable selection approaches, such as stepwise and best subset selection, may be unstable. (Tibshirani, 1996) presented the least absolute shrinkage and selection operator to solve this problem (lasso). Yuan and Lin,(2006) proposed the group lasso the main objective of the method is to find the common elements of groups. (Huo et.al.2020) Sparse Group Lasso (least absolute shrinkage and selection operator) and Support Vector Machine (SGL-SVM) are used for tumor classification.

Support vector machine (SVM):

The supporting vector technology is divided by its use into two parts: support vector machine for regression and support vector for classification . This method is considered one of the most important methods of learning the machine, which was proposed by the researcher (Vapnik 1992)summary of this method is to build an algorithm learned by a supervisor or wave (supervised) and the basic idea in the work of this technique depends on the theory of statistical learning (Ge, S et.al 2007). The origin of the discovery of the supporting vector technology is to find the best solution to the problem of pattern discrimination by choosing the selection of the dividing level of data, that this technique is centered around the main goal of finding the dividing level and ideal of the studied data to be classified and separated into two categories (Ivanciuc 2007). Where the supporting vector technology has a great potential in completing linear and no linear classification issues by relying on the written and classified work not linear (Ahmad.et.al 2002), Some classification issues have only a simple break used for separation, so the concept of the work is based on a no-linear one, which can be determined by using concept kernels. SVM is one of the classical machine learning techniques that can still help solve big data classification problems. Especially, it can help the multidomain applications in a big data environment. Suthaharan, S. (2016)

training data (xi, yi) for i = 1. . . N, with xi ∈ R^d and yi ∈ {−1, 1}.
A linear classifier has the form:

$$f(x) = w^T . x + b \qquad (1)$$



• in 2D the discriminant is a line

• W is the normal to the line, and b the bias

- W is known as the weight vector

Learning the SVM can be formulated as an optimization:

$$\max_{|\mathrm{w}} \frac{2}{\| \mathrm{w} \|} \text{ subject to } (\mathrm{w}^{\top}\mathrm{x}_i + b) \begin{matrix} \geq 1 & \text{if } y_i = +1 \\ \leq -1 & \text{if } y_i = -1 \end{matrix} \qquad (2)$$

for i = 1. . . N

Or equivalently:

$$\min_{\mathrm{w}} \ \|\mathrm{w}\|^2 \text{ subject to } \mathrm{y}_\mathrm{i}\, (\mathrm{w}^{\top}\mathrm{x}_i + b) \ \geq 1 \quad \text{for i = 1. . . N} \quad (3)$$

Regularization methods:

Regularization techniques may also conduct the variable selection , thus regularization approaches can be described as the method used to solve the problem of model complexity. The generalization's performance is closely related to the model's complexity, with a high complexity model having a large variance and a low bias. Because low complexity models have low variance and high bias, regularization methods are frequently used to regulate the complexity of the model by punishing higher complexity models. Donoho and Johnstone were the first to use regularization methods to variable selection. (1994). Regularization methods can be constructed by adding a penalty term to conventional loss functions, such as the O.L.S loss function. The variable selection is applied in regularization methods as part of the parameter estimation process. The Lasso (Tibshirani, 1996), group Lasso (Yuan and Lin, 2006), are examples of regularization techniques .

The Least Absolute Shrinkage and Selection Operator (lasso):

Tibshirani (1996) suggested lasso method to estimate the regression coefficients depend on L_1 -norm penalized least squares criterion. "The lasso algorithm is shrinkage coefficient and variable selection simultaneously, which it minimizes the mean squared error MSE. Lasso performs shrinkage some the coefficients and forces others to be zero, which it provides the interpretable results". Lasso can be written as:

$$\beta_{lasso} = arg\ min \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda \sum_{i=1}^{p}|\beta_i| \qquad (4)$$

$\lambda \geq 0$ Controls the strength of penalty.

There are some drawback on using lasso:

1- If $\boldsymbol{p} > \boldsymbol{n}$, lasso select $\boldsymbol{n}$ variables.

2- Ignoring the grouping information of correlated predictor variables, and select one variable of the group.

3- If $\boldsymbol{n} > \boldsymbol{p}$, with highly correlated predictor variables, ridge outperforms lasso. See (Tibshirani, 1996), (Zou and Hastie, 2005) for more details

Group lasso:

The group Lasso is a generalization or expansion for Lasso estimator has been suggested by Yuan and Lin,(2006) to solve the following problem

$$\hat{\beta} = \text{argmin } (y - X\beta)'(y - X\beta) + \sum_{g=1}^{G} \lambda_g|\beta_g| \qquad (5)$$

Where $y = (y_1, \dots y_n)', \lambda_g > 0$ is the regularization parameters, G is sizes of the groups and $|\beta_g|$ is the L_1 penalty of β_g .The Lasso group performs well when the structure of the group of variables is known (Huang and Zhang, 2010). The attractive feature of this method is to get rid of a group of unimportant variables by making their coefficients equal to zero, this leads to automatic variable selection and estimation of parameters simultaneously Kim et al. (2006). The group Lasso finds various solutions on the level of groups (Yuan and Lin, 2006).

Simulation:

The simulation study will be conducted to show our behavior .The proposed model, group Lasso using the R package. And compare it with the various existing model Lasso. Our comparison is based on the criterion of average sum of errors (MSE) and the criterion of classification error (MIS). Also, we used the mean used to measure the performance of prediction accuracy for different model. Where samples were generated with a volume of (n=100,150,200), For the purpose of generating data according to the following form.

$$y_i = \text{sign}(b + x \,.\, w + \text{error}) \qquad (6)$$

Where (b) is a constant bias amount equal to (3). and that (x) is generated from a multivariate normal distribution. And (w) is the weight vector , where (k = 7) and $\rho = 0.25$.As for the random error term, it was established according to the standard normal distribution. The results described in the theoretical side were obtained and compared between them based on the MSE and MIS. Replication of the experiment 1000 times to obtain the results is stable. The number of influencing variables (p = 81,100,256). The number of samples (n) were classified into two groups (g1,g2). We used the MSE criterion and the MIS criterion to choose the best method, with the least valuable method for the MSE and MIS criteria being the best.

Table (1): explains the results of MSE when $\rho = 0.25$ and K=7.

n	**P**	MSE	
		lasso	**Group lasso**
100	81	0.813	0.609
	100	1.892	0.673
	256	2.889	0.396
150	81	3.083	1.001
	100	1.950	0.929
	256	1.050	0.520
200	81	2.091	1.198
	100	1.974	1.032
	256	1.534	0.457

Table (2): explains the results of MIS when $\rho = 0.25$ and K=7.

n	**P**	MIS	
		lasso	**Group lasso**
100	81	0.307	0.298
	100	0.493	0.425
	256	0.273	0.273
150	81	0.389	0.415
	100	0.433	0.401
	256	0.389	0.302

200	81	0.479	0.439
	100	0.467	0.428
	256	0.298	0.298

Table (1) and table (2) show that the results when ρ=0.25 , n=100,150,200, and k=7 with its seven weights (0.5,1.5,1,2.4,3,3.5,4). We note that the proposed method gives a best results compering with lasso method depend on the values of (MSE and MIS), especially that the method group lasso has the smallest values for MSE and MIS .

Conclusion:

In this paper, lasso and group lasso are used. The results showed that the group method is more stable than lasso method in comparison.

Reference:

Ahmad, A. R., Khalid, M., & Yusof, R. (2002). Machine learning using support vector machines. *Centre for Artificial Intelligence and Robotics*.

Ge, S., Gao, Y., & Wang, R. (2007, August). Least significant bit steganography detection with machine learning techniques. In *Proceedings of the 2007 international workshop on Domain driven data mining* (pp. 24-32).

Huang, J., & Zhang, T. (2010). The benefit of group sparsity. *The Annals of Statistics*, *38*(4), 1978-2004.

Huo, Y., Xin, L., Kang, C., Wang, M., Ma, Q., & Yu, B. (2020). SGL-SVM: a novel method for tumor classification via support vector machine with sparse group Lasso. *Journal of Theoretical Biology*, *486*, 110098.

Ivanciuc, O. (2007). Applications of support vector machines in chemistry. *Reviews in computational chemistry*, *23*, 291.

Kim, Y., Kim, J., & Kim, Y. (2006). Blockwise sparse regression. *Statistica Sinica*, 375-390.

Suthaharan, S. (2016). Support vector machine. In *Machine learning models and algorithms for big data classification* (pp. 207-235). Springer, Boston, MA.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49-67.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.