Summary on A Selective Overview of Variable Selection in High Dimensional Feature Space

Liu Huihang

2020.1.6

Review of model selection method

High dimensional statistical problems arise from diverse fields of scientific research and technological development. Variable selection plays a pivotal role in contemporary statistical learning and scientific discoveries.

Consider the linear regression model

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

The first method we learn is Lasso Tibshirani [1996],

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \|\beta\|_1 \tag{2}$$

It is easy to understand and the speed of computation is faster than solving a $\|\cdot\|_0$ penalty function which is named as best sebset selection method

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \|\beta\|_0 \tag{3}$$

But this method takes a very long time to solve at that time. Thirty years later, we have the latest methods to solve this problem in an acceptable time Wen et al. [2017].

Later many a methods were developed to imporve the result of selection. They used the same framework to constrain the problem as following

$$\min_{\beta\in\mathbf{R}^p}\left\{\frac{1}{2n}\|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_{j=1}^{p} p_\lambda\left(|\beta_j|\right)\right\} \tag{4}$$

For example, SCAD introduced by Fan and Li [2001]

$$p'_\lambda(t) = \lambda\left\{I(t \le \lambda) + \frac{(a\lambda - t)_+}{(a-1)\lambda} I(t > \lambda)\right\} \quad \text{for some } a > 2 \tag{5}$$

where $p_\lambda(0) = 0$ and, often, $a = 3.7$ is used (suggested by a Bayesian argument).

A penalty of similar spirit is the minimax concave penalty (MCP) Zhang and Li [2011], whose derivative is given by

$$p'_\lambda(t) = \frac{(a\lambda - t)_+}{a} \tag{6}$$

A family of concave penalties that bridge the L_0 and L_1 penalties was studied by Lv et al. [2009] And Zheng et al. [2014] focus on the hard thresholding penalty

$$p_{\mathrm{H},\lambda}(t) = \frac{1}{2}\left\{\lambda^2 - (\lambda - t)_+^2\right\}, \quad t \geqslant 0 \tag{7}$$

Generally speaking, two classes of penalty functions have been proposed in the literature: convex ones and concave ones. The L_1 penalty tends to yield a larger model than the true one for optimizing predictions, and many of the selected variables may be insignificant, showing that the resulting method may not be ideal for variable selection. The relatively large model size also reduces the interpretability of the selected model. When concave penalties is used, it is generally difficult to study the properties of the global optimizer for concave regularization methods. Fan and Lv [2013] characterize theoretically the global optimizer of the regularization method with the combined L1 and concave penalty, in the setting of the high-dimensional linear model by studying the resulting regularization problem

$$\min_{\beta \in \mathbf{R}^p} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_0 \|\beta\|_1 + \|p_\lambda(\beta)\|_1 \right\} \tag{8}$$

where $\lambda_0 = c\{(logp)/n\}^{1/2}$ for some positive constant c.

Since then, the classical high dimensonal selection problems have been studied well and not many scholars have been studying general high-dimensional problems in the world.

Scholars are paying more attention to some faster calculation methods and deeper statistical problems such as statistical inference.

References

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

Yingying Fan and Jinchi Lv. Asymptotic properties for combined L 1 and concave regularization. *Biometrika*, 101(1):57–70, 2013.

Jinchi Lv, Yingying Fan, and Others. A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics*, 37(6A):3498–3528, 2009.

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

Canhong Wen, Aijun Zhang, Shijie Quan, and Xueqin Wang. BeSS: An R Package for Best Subset Selection in Linear, Logistic and CoxPH Models. *arXiv preprint arXiv:1709.06254*, 2017.

Yiyun Zhang and Runze Li. Iterative conditional maximization algorithm for nonconcave penalized likelihood. In *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P Hettmansperger*, pages 336–351. World Scientific, 2011.

Zemin Zheng, Yingying Fan, and Jinchi Lv. High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):627–649, 2014.