Optimization Techniques for SCAD Variable Selection in Medical Research

Yan Fang[1]*, YanYan Kong[2] and Yumei Jiao[3]

[1]*School of Finance, Shanghai University of International Business and Economics, China*

[2]*PET Center, Fudan University, China*

[3]*Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, China*

Received: 📅 August 15, 2018; **Published:** 📅 August 23, 2018

*Corresponding author: Yan Fang, School of Finance, Shanghai University of International Business and Economics, Shanghai, China

Abstract

High-dimensional data analysis requires variable selection to identify truly relevant variables. More often it is done implicitly via regularization, such as penalized regression. Of the many versions of penalties, SCAD has shown good properties and has been widely adopted in medical research and many more areas. This paper reviews the various optimization techniques in solving SCAD penalized regression.

Abbreviations: LQA: Local Quadratic Approximation; LLA: Local Linear Approximation; DCA: Difference Convex Algorithm; SOCP: Second Order Cone Programming; ADMM: Alternating Direction Method of Multipliers

Introduction

High-dimensional data analysis has been a common and important topic in biomedical/genomic/clinical studies. For example, the identification of genetic factors for complex diseases such as lung cancer implicates a variety of genetic variants. For high-dimensional data, there is the well-known problem of curse of dimensionality arising in modeling. Therefore, variable selection is a fundamental task for high-dimensional statistical modeling. The "old school" way of doing variable selection is to follow a subset selection procedure prior to building the model of interest. The procedure commonly adopts AIC/BIC as evaluation metric and often iterates in a stepwise fashion. Yet this is independent of the subsequent modeling task hence the effectiveness might be less desirable. A more natural way is to integrate the variable selection into the modeling itself, i.e., the penalized regression, which simultaneously performs variable selection and coefficient estimation.

Theoretically, the "best" penalty for the penalized regression is the number of non-zero variables, to push as many variables to zero as possible. Yet, it is well known that the L0 (also known as the entropy penalty) optimization [1] is infeasible. As such, the L1 (LASSO) penalty Tibshirani [2] is our "next best" candidate, which is widely adopted in statistical and machine learning community for sparse solutions. However, [3] point out that L1 suffers the problem of biasedness. They propose the Smoothly Clipped Absolute Deviation (SCAD) penalty that can produce unbiased estimates while retaining good properties of L1. Subsequently,

the SCAD penalty function has seen a wide range of applications including medical/clinical research, such as [1,4-7]. Nevertheless, the estimating procedure for SCAD penalized regression is no trivial task, because the target function

a)   is a high-dimensional non-concave function,

b)   is singular at the origin,

c)   does not have continuous second order derivatives.

Optimization Techniques for SCAD

We now review the optimization techniques in solving SCAD penalized regression over the years.

Local Quadratic Approximation (LQA) Algorithm

Due to the non-differentiability [3], in the original paper, propose to approximate the SCAD penalty by a quadratic function, which is amenable to optimization. The drawback of this approximation is that once a coefficient is pushed to zero, it will remain zero. Another problem is that the LQA estimator does not have a sparse representation Zou and Li [8].

Local Linear Approximation (LLA) Algorithm

Zou and Li [8] develop local linear approximation (LLA) for maximizing the non-concave penalized likelihood models. Not only does LLA address the issues of LQA but also it enforces the symmetric linear function in the approximation of the SCAD penalty instead. Undoubtedly, the LLA approximation also has ascent

property, which enjoys convergence guarantee of EM algorithms. However, the LLA algorithm is rather sensitive to initial estimators and can easily get stuck in local minima Mazumder Friedman and Hastie [9].

Difference Convex Algorithm (DCA)

The idea of DCA algorithm An and Tao [10] is to decompose the target function as the difference of two convex functions. Using this technique for non-convex problems, Wu and Liu [11] approximate the target function by two convex functions, and the second convex function is furthered approximated by a linear function, like LLA. However, the DCA approach does not enforce symmetry in its approximation, which is different from LLA.

Second Order Cone Programming (SOCP)

The SOCP algorithm is a method where linear function is minimized over the intersection of an affine set and the product of second-order (quadratic) cones Lobo et al., [12]. As a nonlinear convex approach, Noh et al. [13] utilize the SOCP approach to estimate the SCAD penalized regression. In addition, the authors show that the estimated relevant coefficients converge to the true functions at the univariate optimal rate.

Alternating Direction Method of Multipliers (ADMM)

As a prime-dual optimization technique, Yin and Zeng [14] apply ADMM to a family of optimization tasks, including SCAD. The method has a theoretical guarantee of convergence, and is also shown by Bertsimas, Copenhaver and Mazumder [15] to perform the best among several heuristic approaches.

Conclusion

In this paper we briefly introduced the variable selection problem in high-dimension modeling. We concluded that more naturally it is done via penalized regression. Of the many versions of penalties, SCAD stands out due to its nice properties. We reviewed the various SCAD optimization techniques proposed over the years.

References

1. Breiman L (1996) Heuristics of Instability and Stabilization in Model Selection. The annals of statistics 24(6): 2350-2383.

2. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B 58(1): 267-288.

3. 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.

4. Fan J and Li R (2002) Variable Selection for Cox's Proportional Hazards Model and Frailty Model. The annals of statistics 30(1): 74-99.

5. Breheny P, Huang J (2011) Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection. Annals of Applied Statistics 5(1): 232-253.

6. Wang Z, Ma S, Zappitelli M, Parikh C, Wang CY, et al. (2016) Penalized Count Data Regression with Application to Hospital Stay after Pediatric Cardiac Surgery. Statistical Methods in Medical Research 25(6): 2685-2703.

7. Gim J, Kim W, Kwak SH, Choi H, Park C, et al. (2017) Improving Disease Prediction by Incorporating Family Disease History in Risk Prediction Models with Large-Scale Genetic Data. Genetics 207(3): 1147-1155.

8. Zou H, Li R (2008) One-step Sparse Estimates in Nonconcave Penalized Likelihood Models[J]. The Annals of Statistics 36(4): 1509-1533.

9. Mazumder R, Friedman J, Hastie T (2011) Sparse Net: Coordinate Descent with Non-convex Penalties. Journal of American Statistical Association 106(495): 1125-1138.

10. An LT H, Tao P D (1997) Solving a Class of Linearly Constrained Indefinite Quadratic Problems by DC Algorithms. Journal of Global Optimization 11(3): 253-285.

11. Wu YC, Liu YF (2009) Variable Selection in Quantile Regression. Statistica Sinica 19(2): 801-817.

12. Lobo M, Vandenberghe L, Boyd S, Lebret H (1998) Applications of Second-Order Cone Programming[J]. Linear Algebra and its Applications 284: 193-228.

13. Noh H, Chung K, Keilegom I Van (2012) Variable Selection of Varying Coefficient Models in Quantile Regression. Electronic Journal of Statistics 6: 1220-1238.

14. Wang Y, Yin W, Zeng J (2018) Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. Journal of Scientific Computing p: 1-35.

15. Bertsimas D, Copenhaver MS, Mazumder R (2017) The Trimmed Lasso: Sparsity and Robustness.

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

https://biomedres.us/