False Discovery Rate Control with Concave Penalties using Stability Selection

Bhanukiran Vinzamuri and Kush R. Varshney

IBM Research

Thomas J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York, USA

Objectives

Study FDR control obtained using concave penalties with stability selection.

- Study FDR theory for concave penalties.
- Understand the FDR bound in stability selection and see how it can be improved.
- Propose new FDR bound with stability selection and concave penalties.

Introduction

The standard linear regression problem has the following form:

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

where $y \in \mathbb{R}^n$ is a response variable, $X \in \mathbb{R}^{n\times p}$ is a feature matrix, $\beta \in \mathbb{R}^p$ is a coefficient vector, and $\varepsilon \in \mathbb{R}^n$ is a noise vector which has zero mean and sub-Gaussian noise such that $\varepsilon \sim N(0, \sigma^2 I_{n\times n})$.

The following class of regularized linear regression problems is studied here:

$$\hat{\beta} = \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} L(\beta; \lambda; \gamma), \tag{2}$$

where

$$L(\beta; \lambda; \gamma) = \frac{1}{2n}\|y - X\beta\|_2^2 + \sum_{j=1}^{p} h(\beta_j; \lambda; \gamma), \tag{3}$$

$\beta = (\beta_1, \ldots, \beta_p)$, and $h(\beta_j; \lambda; \gamma)$ is a concave penalty function consisting of parameters λ and γ.

The value of this penalty evaluated for a specific regression coefficient vector $\beta \in \mathbb{R}^p$

$$\|h(\beta; \lambda; \gamma)\|_1 = \sum_{j=1}^{p} h(\beta_j; \lambda; \gamma) \tag{4}$$

$$h_{MCP}(t; \lambda; \gamma) = \min\{\lambda t - t^2/2\gamma, \lambda^2\gamma/2\}.$$

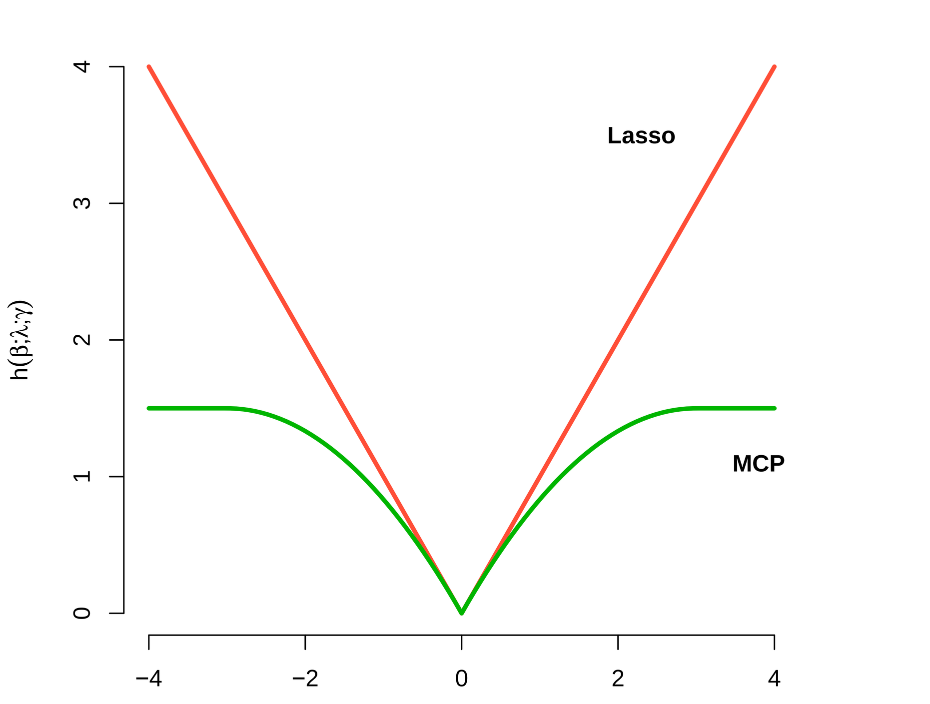

Figure 1: Comparison of Lasso vs Minimax Concave Penalty (MCP).

Motivation

Theoretical results indicate that for a noise level with standard deviation σ and universal amount of penalization $\lambda_{univ} \equiv \sigma\sqrt{\frac{2logp}{n}}$, MCP is said to have a *selection consistency* property [1, 2], which implies that the set of selected variables is identical to the set of true nonzero regression coefficients with high probability. However, estimating noise level precisely from real-world data is a non-trivial task which makes it difficult to set λ_{univ}.

Our proposed stability selection with concave penalties approach handles this problem by defining a range of permissible regularization parameters. This is easier to define and makes the framework less parameter dependent.

Notations

We review notations defined here [3]. For any regularization parameter $\lambda \in \Lambda$, the selected set $\hat{S}^\lambda$ represents the set of features active in the model at parameter λ. For every set $K \subseteq \{1, 2, \ldots, p\}$, the probability of being in the selected set $\hat{S}^\lambda$ is:

$$\hat{\Pi}_K^\lambda = \mathbb{P}^*\{K \subseteq \hat{S}^\lambda(I)\}, \tag{5}$$

where $\mathbb{P}^*$ represents the probability estimate.

For every variable $k = \{1, 2, \ldots, p\}$, the selection probabilities are given by $\hat{\Pi}_k^\lambda, \lambda \in \Lambda$. Let $\hat{S}^\Lambda = \cup_{\lambda\in\Lambda}\hat{S}^\lambda$, be the set of selected variables if varying the regularization parameter λ in the set Λ. Let V be the number of falsely selected variables where

$$V = |N \cap \hat{S}^\Lambda| \tag{6}$$

Important Result

Theorem

Assume that the distribution of $\{1_{\{k\in\hat{S}^\lambda\}}, k \in N\}$ is exchangeable for all $\lambda \in \Lambda$. The expected number V of false positives for our approach is then bounded for $\pi_{thr} \in (\frac{1}{2}, 1)$ by

$$\mathbb{E}(V) < \frac{1}{2\pi_{thr} - 1}\frac{\left(\alpha + 9/4\right)^2 |S|^2}{|N|} \quad \textit{where} \quad \alpha > 0.$$

Results

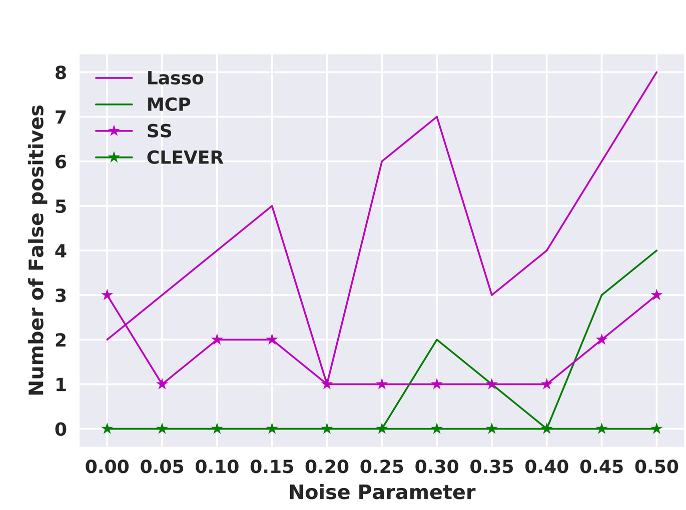

Figure 2: Comparison of false positives with varying noise levels for synthetic datasets.

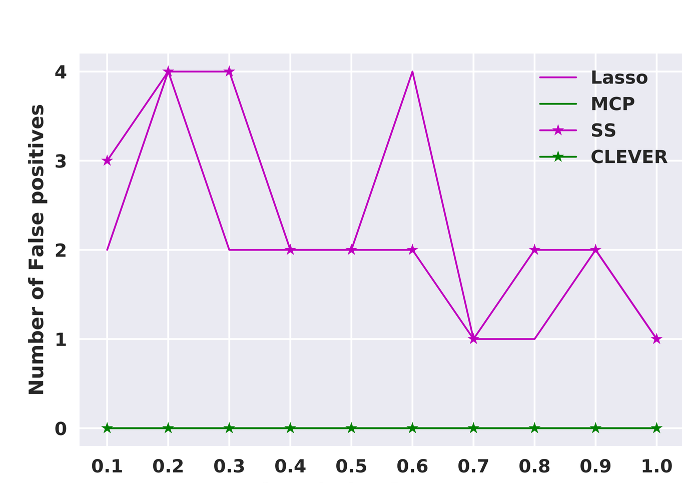

Figure 3: Comparison of false positives with varying correlation levels for synthetic datasets.

Discussion

- Results indicate that our approach (CLEVER) has lower number of false positives discovered when compared to Lasso, SS (Stability Selection) and MCP at varying levels of noise and correlation for synthetic datasets.
- Such effective FDR control helps in improving model consistency and interpretation. This analysis is very important from a practitioner's perspective, as he or she can tune the number of features to be selected at a specified false positive error rate or vice versa.

Conclusion

This approach has several advantages over other competing methods while conducting inference from the sparse and noisy data such as (i) unbiased regression, and (ii) high false positive error control. We derived theoretical guarantees for this approach which upper bounds the expected number of false positives.

Future Work

As future work, we plan to combine the knockoff method [4] with concave penalties which also has robust guarantees on false positive control.

References

[1] C. H. Zhang.
Nearly unbiased variable selection under minimax concave penalty.
The Annals of Statistics, 38(2):894–942, 2010.

[2] C. H. Zhang and T. Zhang.
A general theory of concave regularization for high-dimensional sparse estimation problems.
Statistical Science, 27(4):576–593, 2012.

[3] N. Meinshausen and P. Bühlmann.
Stability selection.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(4):417–473, 2010.

[4] Rina Foygel Barber, Emmanuel J Candès, et al.
Controlling the false discovery rate via knockoffs.
The Annals of Statistics, 43(5):2055–2085, 2015.

Acknowledgements

This work was done under the auspices of the Science for Social Good Initiative.

Contact Information

- Email: bhanu.vinzamuri@ibm.com
- Phone: +1 (914) 945 2032