

Feature selection for mean shift outlier model via the conic-fused Lasso

Pakize Taylan¹ and Fatma Yerlikaya-Özkurt²

¹Dicle University, Diyarbakır, Turkey

²Atılım University, Ankara, Turkey

Abstract

Outlier detection and variable selection are among main objectives of statistical analysis. In our study, we address the outlier problem for classification by using the Mean Shift Outlier Model (MSOM). Since the MSOM has more coefficients than the linear regression model, the complexity of the MSOM is high. Therefore, we consider feature selection for MSOM by using fused Lasso which is useful when the number of features is much larger than the sample size. Fused Lasso is penalizing the L1- norm of both the coefficients and their successive differences and it allows sparsity for both of them while Lasso only allows the coefficients. In this study, we take into account Iterated Ridge approximation for fused Lasso problem and it is solved by using continuous optimization technique which is permitting the use of interior point methods. The newly developed method is called C-Fused Lasso (C-FL) and is applied to real world data set to show the performance of C-FL.

Keywords

Outlier, Lasso, Mean shift outlier model, Classification problem, Convex optimization.

References

- [1] Ben-Tal A. and Nemirovski, A. (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications*, MPS-SIAM Series on Optimization.
- [2] J. Renegar, A. (2001). *Mathematical View of Interior-Point Methods in Convex Programming*, SIAM, Philadelphia.
- [3] Rao, C.R., Toutenburg, H. and Fieger, A. (1999). *Linear models: Least squares and alternatives*, Springer.
- [4] Tibshirani, R. Saunders, M., Rosset, S. and Zhu, J. (2005). Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc. B*, 67, 91–108.