Greenshtein, Eitan; **Park, Junyong**; **Lebanon, Guy**

Regularization through variable selection and conditional MLE with application to classification in high dimensions. (English) Zbl 1149.62052

J. Stat. Plann. Inference 139, No. 2, 385-395 (2009).

Summary: It is often the case that high-dimensional data consist of only a few informative components. Standard statistical modeling and estimation in such a situation is prone to inaccuracies due to overfitting, unless regularization methods are practiced. In the context of classification, we propose a class of regularization methods through shrinkage estimators. The shrinkage is based on variable selection coupled with conditional maximum likelihood. Using Stein's unbiased estimator of the risk, we derive an estimator for the optimal shrinkage method within a certain class. A comparison of the optimal shrinkage methods in a classification context, with the optimal shrinkage method when estimating a mean vector under squared loss, is given. The latter problem is extensively studied, but it seems that the results of those studies are not completely relevant for classification. We demonstrate and examine our method on simulated data and compare it to feature annealed independence rule and Fisher's rule.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62F10 Point estimation
65C60 Computational problems in statistics (MSC2010)

Keywords:

classification; high dimensions; conditional MLE; Stein's unbiased estimator

Software:

EBayesThresh

Full Text: DOI

References:

[1] Bickel, P.; Levina, E., Some theory for Fisher's linear discriminant function, "naive bayes", and some alternatives where there are many more variables than observations, Bernoulli, 10, 6, 989-1010, (2004) · Zbl 1064.62073

[2] Brown, L.D., Fundamentals of statistical exponential families, with applications in statistical decision theory, (1986), IMS Hayward, CA · Zbl 0685.62002

[3] Campbell, N.A., Shrunken estimation in discriminant and canonical variable analysis, Appl. statist., 29, 1, 5-14, (1980) · Zbl 0454.62053

[4] Donoho, D.L.; Johnstone, I.M., Ideal spatial adaptation by wavelet shrinkage, Biometrika, 81, 425-455, (1994) · Zbl 0815.62019

[5] Donoho, D.L.; Johnstone, I.M., Adapting to unknown smoothness via wavelet shrinkage, J. amer. statist. assoc., 90, 4, 1200-1224, (1995) · Zbl 0869.62024

[6] Fan, J., Fan, Y., to appear. High dimensional classification using features annealed independence rules. Ann. Statist. · Zbl 1360.62327

[7] Foster, D.P.; George, E.L., The risk inflation criterion for multiple regression, Ann. statist., 22, 1947-1975, (1994) · Zbl 0829.62066

[8] Friedman, J., Regularized discriminant analysis, J. amer. statist. assoc., 84, 165-175, (1989)

[9] Hastie, T.; Buja, A.; Tibshirani, R., Penalized discriminant analysis, Ann. statist., 23, 73-102, (1995) · Zbl 0821.62031

[10] Johnstone, I.M.; Silverman, B.W., Empirical Bayes selection of wavelet thresholds, Ann. statist., 33, 4, 1700-1752, (2005) · Zbl 1078.62005

[11] Lehmann, E.L., Testing statistical hypothesis, (1986), Wiley New York · Zbl 0608.62020