

An Efficient Method to Estimate the Optimum Regularization Parameter in RLDA

Daniyar Bakir, Alex James Pappachen, Amin Zollanvari

- Department of Electrical and Electronical Engineering

Abstract

Motivation: The biomarker discovery process in high-throughput genomic profiles has presented the statistical learning community with a challenging problem, namely learning when the number of variables is comparable or exceeding the sample size. In these settings, many classical techniques including linear discriminant analysis (LDA) falter. Poor performance of LDA is attributed to the ill-conditioned nature of sample covariance matrix when the dimension and sample size are comparable. To alleviate this problem regularized LDA (RLDA) has been classically proposed in which the sample covariance matrix is replaced by its ridge estimate. However, the performance of RLDA depends heavily on the regularization parameter used in the ridge estimate of sample covariance matrix.

Results: We propose a range-search technique for efficient estimation of the optimum regularization parameter. Using an extensive set of simulations based on synthetic and gene expression microarray data, we demonstrate the robustness of the proposed technique to Gaussianity, an assumption used in developing the core estimator. We compare the performance of the technique in terms of accuracy and efficiency to classical techniques for estimating the regularization parameter. In terms of accuracy, the results indicate that the proposed method vastly improves on similar techniques that use classical plug-in estimator. In that respect, it is better or comparable to cross-validation based search strategies while, depending on the sample size and dimensionality, being tens to hundreds of times faster to compute.

Contact: amin.zollanvari@nu.edu.kz

Original language	English
Number of pages	7
Journal	<u>Bioinformatics</u>
State	Published - 2016

Bakir, D., James Pappachen, A., & Zollanvari, A. (2016). *An Efficient Method to Estimate the Optimum Regularization Parameter in RLDA*. *Bioinformatics*. DOI: [10.1093/bioinformatics/btw506](https://doi.org/10.1093/bioinformatics/btw506)