Statistical Analysis and Data Mining: The ASA Data Science Journal Explore this journal >

View issue TOC
Volume 10, Issue 6
December 2017
Pages 436–447

ORIGINAL ARTICLE

Cellwise robust regularized discriminant analysis

Stéphanie Aerts, Ines Wilms

First published:
10 November 2017 Full publication history

DOI:
10.1002/sam.11365 View/save citation

Cited by (CrossRef):
0 articles Check for updates Citation tools

Funding Information

Funding Information FWO (Research Foundation Flanders), 12M8217N.

Abstract

Quadratic and linear discriminant analysis (QDA and LDA) are the most often applied classification rules under normality. In QDA, a separate covariance matrix is estimated for each group. If there are more variables than observations in the groups, the usual estimates are singular and cannot be used anymore. Assuming homoscedasticity, as in LDA, reduces the number of parameters to estimate. This rather strong assumption is however rarely verified in practice. Regularized discriminant techniques that are computable in high dimension and cover the path between the 2 extremes QDA and LDA have been proposed in the literature. However, these procedures rely on sample covariance matrices. As such, they become inappropriate in the presence of cellwise outliers, a type of outliers that is very likely to occur in high-dimensional datasets. In this paper, we propose cellwise robust counterparts of these regularized discriminant techniques by inserting cellwise robust covariance matrices. Our methodology results in a family of discriminant methods that (1) are robust against outlying cells, (2) cover the gap between LDA and QDA, and (3) are computable in high dimension. The good performance of the new methods is illustrated through simulated and real data examples. As a by-product, visual tools are provided for the detection of outliers.