Search or Article-id

(Help | Advance

All papers

**Mathematics > Statistics Theory**

# Sparse linear discriminant analysis by thresholding for high dimensional data

Jun Shao, Yazhen Wang, Xinwei Deng, Sijian Wang

*(Submitted on 18 May 2011)*

In many social, economical, biological and medical studies, one objective is to classify a subject into one of several classes based on a set of variables observed from the subject. Because the probability distribution of the variables is usually unknown, the rule of classification is constructed using a training sample. The well-known linear discriminant analysis (LDA) works well for the situation where the number of variables used for classification is much smaller than the training sample size. Because of the advance in technologies, modern statistical studies often face classification problems with the number of variables much larger than the sample size, and the LDA may perform poorly. We explore when and why the LDA has poor performance and propose a sparse LDA that is asymptotically optimal under some sparsity conditions on the unknown parameters. For illustration of application, we discuss an example of classifying human cancer into two classes of leukemia based on a set of 7,129 genes and a training sample of size 72. A simulation is also conducted to check the performance of the proposed method.

**Submission history**

From: Jun Shao [view email]
**[v1]** Wed, 18 May 2011 08:35:46 GMT (560kb)

*Which authors of this paper are endorsers?*

## Download:

- PDF
- PostScript
- Other formats

Current browse cont

**math.ST**
< prev | next >
new | recent | 1105

Change to browse b

math
stat

References & Citatio

- NASA ADS

Bookmark(what is this?)

Science WISE