

# STATISTICAL ANALYSIS AND DATA MINING

Research Article

## Data reduction in classification: A simulated annealing based projection method

Tian Siva Tian ✉, Rand R. Wilcox, Gareth M. James

First published: 03 August 2010

<https://doi.org/10.1002/sam.10087>

Cited by: 3



About



Access

als



ire

### Abstract

This paper is concerned with classifying high - dimensional data into one of two categories. In various settings, such as when dealing with fMRI and microarray data, the number of variables is very large, which makes well - known classification techniques impractical. The number of variables might be reduced via principal component analysis or some robust analog, but these methods are usually unsatisfactory for the purpose of classification because they are unsupervised learning methods and not designed to minimize classification errors. In this paper, we propose a classification guided dimensionality reduction approach incorporating a stochastic search algorithm in order to look for a 'good' subspace in the context of classification. Two different versions of the simulated annealing algorithm are implemented to produce sparse and dense models, respectively. Using data from both simulation and real - world studies, situations are found where the misclassification rate can be reduced by the proposed approach. Copyright © 2010 Wiley Periodicals, Inc. Statistical Analysis and Data Mining 3: 319 - 331, 2010

Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley



