

## STATISTICAL ANALYSIS AND DATA MINING

Research Article

## Multiple Response Regression for Gaussian Mixture Models with Known Labels

Wonyul Lee, Ying Du, Wei Sun, David Neil Hayes, Yufeng Liu✉

First published: 29 October 2012

<https://doi.org/10.1002/sam.11158>

About



Access

als



are

## Abstract

Multiple response regression is a useful regression technique to model multiple response variables using the same set of predictor variables. Most existing methods for multiple response regression are designed for modeling homogeneous data. In many applications, however, one may have heterogeneous data where the samples are divided into multiple groups. Our motivating example is a cancer dataset where the samples belong to multiple cancer subtypes. In this paper, we consider modeling the data coming from a mixture of several Gaussian distributions with known group labels. A naive approach is to split the data into several groups according to the labels and model each group separately. Although it is simple, this approach ignores potential common structures across different groups. We propose new penalized methods to model all groups jointly in which the common and unique structures can be identified. The proposed methods estimate the regression coefficient matrix, as well as the conditional inverse covariance matrix of response variables. Asymptotic properties of the proposed methods are explored. Through numerical examples, we demonstrate that both estimation and prediction can be improved by modeling all groups jointly using the proposed methods. An application to a glioblastoma cancer dataset reveals some interesting common and unique gene relationships across different cancer subtypes. © 2012 Wiley Periodicals, Inc. Statistical Analysis and Data Mining, 2012

About Wiley Online Library



Help &amp; Support



Opportunities



Connect with Wiley



