OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

isi

IPS Abstract

Robust estimation in regression and classification methods for large dimensional data

Author: Dr Chunming Zhang

Brief Description

In statistical data analysis and machine learning practice, Bregman divergence (BD) plays an important role in quantifying error measures for regression estimates, classification procedures and forecasting methods.

The quadratic loss function and the negative quasi-likelihood are two examples of widely used error measures which along with many others belong to the family of BD, but are not resistant to either outlying observations or high leverage points, more often encountered in large- and high dimensional datasets.

In this paper, we introduce a class of robust forms of BD, called "robust-BD", and explore the suitability of "penalized robust-BD estimates" of parameters in sparse large-dimensional regression models, which allow the distribution of the response variable given the covariates to be incompletely specified.

It is shown that the new estimate, combined with appropriate penalties, achieves the same oracle property of the ordinary non-robust penalized least-squares and penalized-likelihood estimates, but is less sensitive to data outliers, a very desirable property in many real-world applications.

Extensive numerical results are presented to compare the performance of the new estimates with that of the classical ones.

A real dataset is analyzed for illustration.

Abstract

In statistical data analysis and machine learning practice, Bregman divergence (BD) plays an important role in quantifying error measures for regression estimates, classification procedures and forecasting methods. The quadratic loss function and the negative quasi-likelihood are two examples of widely used error measures which along with many others belong to the family of BD, but are not resistant to either outlying observations or high leverage points, more often encountered in large- and high dimensional datasets. In this paper, we introduce a class of robust forms of BD, called "robust-BD", and explore the suitability of "penalized robust-BD estimates" of parameters in sparse large-dimensional regression models, which allow the distribution of the response variable given the covariates to be incompletely specified. It is shown that the new estimate, combined with appropriate penalties, achieves the same oracle property of the ordinary non-robust penalized least-squares and penalized-likelihood estimates, but is less sensitive to data outliers, a very desirable property in many real-world applications. Extensive numerical results are presented to compare the performance of the new estimates with that of the classical ones. A real dataset is analyzed for illustration.