Multiple-Instance Logistic Regression with LASSO Penalty

張升懋
成功大學統計學系

Abstract

Ordinary multiple-instance data consists of many independent bags and each bag is composed of several instances. The statuses of bags and instances are binary. Moreover, the statuses of instances are not observed, whereas the statuses of bags are observed. In this work, we consider the data generated from a stable manufacturing process which can be described by a multiple-instance logistic regression. The goal of the data analysis is to identify important covariates that affect the statuses of bags as well as to predict bag and instance statuses. For estimation purpose, an expectation -maximization algorithm is proposed, and the proposed model is extended to identify important covariates by adding LASSO penalty. In addition to essential technical details, we demonstrate the usefulness of the proposed method by simulations and real examples.