S09-4 (02/17 16:30-16:55)

Supervised Stratified Subsampling for Big Data Reduction

Ming-Chung Chang
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Abstract

Extraordinary amounts of data are generated across disciplines owing to advanced technology. Such data richness, however, may yield difficulties in statistical model fitting and prediction either in terms of time cost or numerical stability. On the other side, stratified sampling has been used to control the homogeneity of data in the literature. In this talk, I will introduce a new subsampling approach, referred to as supervised stratified subsampling, that integrates nonparametric regression and stratified sampling for big data reduction. Theoretical and numerical results are provided to show the benefits of the proposed method.

Keywords:
Large-scale data; Partitioning estimate; Bagging.