Integrating multiple random sketches for sufficient dimension reduction in large-p-small-n problems

Su-Yun Huang

Academia Sinica

Abstract

Sufficient dimension reduction (SDR) is continuing an active research field nowadays. When estimating the central subspace (CS), inverse regression based SDR methods involve solving a generalized eigenvalue problem, which can be problematic under the large-p-small-n situation. In recent years, there are emerging new techniques in numerical linear algebra, called randomized algorithms or random sketching, for high dimensional and large scale problems. To overcome the large-p-small-n problem in SDR, we combine the idea of statistical inference with random sketching to propose a new SDR method, named integrated random-partition SDR (iRP-SDR). Our method consists of the following steps. (1) Randomly partition the covariates into subsets to construct an envelope subspace with low dimension. (2) Obtain a sketch of the CS by applying conventional SDR method within the constructed envelope subspace. (3) Repeat the above two steps for multiple times and integrate these multiple sketches to form the final estimate of the CS. The advantageous performance of iRP-SDR is demonstrated via simulation studies and the EEG data analysis.

(Joint work with Hung Hung)