Sufficient dimension reduction via random-partitions for the large-p-small-n problem

Su-Yun Huang

Academia Sinica
E-mail: syhuang@stat.sinica.edu.tw

Abstract: Sufficient dimension reduction (SDR) continues to be an active field of research. 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, new techniques have emerged in numerical linear algebra, called randomized algorithms or random sketching, for high-dimensional and large scale problems. To overcome the large-p-small-n SDR problem, we combine the idea of statistical inference with random sketching to propose a new SDR method, called integrated random-partition SDR (iRP-SDR). Our method consists of the following three steps: (i) Randomly partition the covariates into subsets to construct an envelope subspace with low dimension. (ii) Obtain a sketch of the CS by applying a conventional SDR method within the constructed envelope subspace. (iii) Repeat the above two steps many times and integrate these multiple sketches to form the final estimate of the CS. After describing the details of these steps, the asymptotic properties of iRP-SDR are established. Unlike existing methods, iRP-SDR does not involve the determination of the structural dimension until the last stage, which makes it more adaptive to a high-dimensional setting. The advantageous performance of iRP-SDR is demonstrated via simulation studies and a practical example analyzing EEG data.