Integrated Randomized Algorithms for Big Data SVD and PCA

陳素雲
中央研究院統計科學研究所

Abstract

Random sketches and randomized algorithms have emerged as an important tool for high dimensional and big data analysis in recent years. In this talk we will first introduce an existing randomized algorithm for big data SVD and PCA. To increase the precision of results from the randomized algorithm, we propose an integration algorithm based on multiple randomizations. This integration algorithm consists of iterative steps of lifting points from matrix manifold to tangent, next taking mean of these lifted points, and then retracting the mean point back to the matrix manifold. Theoretical properties as well as numerical examples will be presented.