

On the efficiency-loss free ordering-robustness of product-PCA

Hung Hung

¹Institute of Health Data Analytics and Statistics, National Taiwan University, Taiwan

Abstract

This article studies the robustness of the eigenvalue ordering, an important issue when estimating the leading eigen-subspace by principal component analysis (PCA). In Yata and Aoshima (2010), cross-data-matrix PCA (CDM-PCA) was proposed and shown to have smaller bias than PCA in estimating eigenvalues. While CDM-PCA has the potential to achieve better estimation of the leading eigen-subspace than the usual PCA, its robustness is not well recognized. In this article, we first develop a more stable variant of CDM-PCA, which we call product-PCA (PPCA), that provides a more convenient formulation for theoretical investigation. Secondly, we prove that, in the presence of outliers, PPCA is more robust than PCA in maintaining the correct ordering of leading eigenvalues. The robustness gain in PPCA comes from the random data partition, and it does not rely on a data down-weighting scheme as most robust statistical methods do. This enables us to establish the surprising finding that, when there are no outliers, PPCA and PCA share the same asymptotic distribution. That is, the robustness gain of PPCA in estimating the leading eigen-subspace has no efficiency loss in comparison with PCA. Simulation studies and a face data example are presented to show the merits of PPCA. In conclusion, PPCA has a good potential to replace the role of the usual PCA in real applications whether outliers are present or not.

Key words: cross-data-matrix PCA; dimension reduction; efficiency loss; ordering of eigenvalues; random partition; robustness.