

A Statistical Framework for Contrastive Principal Component Analysis

Shao-Hsuan Wang

Graduate Institute of Statistics, National Central University

Principal component analysis (PCA) is one of the most widely used techniques for exploratory data analysis. Contrastive principal component analysis (cPCA), introduced by Abid et al. (2018), generalizes PCA by extracting features that are enriched in a target dataset relative to a background dataset while preserving as much information as possible from the target data. However, in high-dimensional settings, cPCA becomes computationally demanding due to the need to construct the contrastive covariance matrix and perform its eigenvalue decomposition to obtain the leading contrastive principal components.

Despite its growing popularity, a rigorous statistical framework for cPCA has been lacking. In this work, we establish a statistical framework for cPCA and develop its asymptotic theory under the high-dimensional spiked covariance model. Building on these theoretical results, we further propose a data-driven procedure for selecting the contrast parameter, a fundamental issue that has remained largely unresolved.

In this talk, we will present the proposed statistical framework, discuss the main theoretical results, and illustrate their performance through numerical studies, demonstrating the validity of the proposed theory.