



Supervised Sparse and Functional Principal Component Analysis

Gen Li

UNC-CH, Chapel Hill, USA – ligen@live.unc.edu

Haipeng Shen*

UNC-CH, Chapel Hill, USA; HKU, Hong Kong, PRC – shenhaipeng@gmail.com

Jianhua Z. Huang

Texas A&M University, Collage Station, USA – jianhua@stat.tamu.edu

Principal component analysis (PCA) is an important tool for dimension reduction in multivariate analysis. Regularized PCA methods, such as sparse PCA and functional PCA, have been developed to incorporate special features in many real applications. Sometimes additional variables (referred to as supervision) are measured on the same set of samples, which can potentially drive low-rank structures of the primary data of interest. Classical PCA methods cannot make use of such supervision data. In this paper, we propose a supervised sparse and functional principal component (SupSFPC) framework that can incorporate supervision information to recover underlying structures that are more interpretable. The framework unifies and generalizes several existing methods and flexibly adapts to the practical scenarios at hand. The SupSFPC model is formulated in a hierarchical fashion using latent variables. We develop an efficient modified expectation-maximization algorithm for parameter estimation. We also implement fast data-driven procedures for tuning parameter selection. Our comprehensive simulation and real data examples demonstrate the advantages of SupSFPC.

Keywords: Regularized PCA; Supervised dimension reduction; Penalized likelihood; Low rank approximation; Latent variable; SupSFPC.