

Modelling Handwritten Digit Data using Probabilistic Principal Component Analysis

Mohamed E.M. Musa, Robert P.W. Duin and Dick de Ridder
Pattern Recognition Group,
Department of Applied Physics
Delft University of Technology,
Lorentzweg 1, 2628 CJ Delft The Netherlands
e-mail: musa@ph.tn.tudelft.nl

Abstract

Principal Component Analysis (PCA) is one century old now. Nevertheless, it still undergoes research and new extensions are found. Probabilistic Principal Component Analysis (PPCA, proposed by Tipping and Bishop) is one of these recent PCA extensions. PPCA defines a probabilistic generative model for PCA. It can easily be extended to mixture models. Among recent mixture density theoretical developments is Dasgupta's algorithm for learning mixtures of Gaussians. We propose enhancing PPCA's EM training algorithm by increasing the number of submodels iteratively, together with using a version of Dasgupta's algorithm for parameter initialization. Handwritten digit classification is an extensively studied problem. Therefore, it is considered a popular benchmark for model comparison. Experimental results show remarkable improvement when using our extensions.

Keywords: Principal Component Analysis, mixture models, model selection, handwritten digit recognition.