Probabilistic Principal Component Analysis

Group 8, Team EnvyUs
School of Engineering and Applied Science,
Ahmedabad University
Course: Algorithms and Optimization for Big Data
March 12, 2017

***Abstract*—The Probabilistic Principal Component Analysis (PPCA) is a probability-based model which demonstrates how the principal axes of a set of observed data vectors can be determined through maximum-likelihood estimation of parameters in a latent variable model, closely to factor analysis. In this report, we demonstrate applications of PPCA-EM algorithms. Demonstrated applications are Image Reconstruction and Image Denoising.**

***Index Terms*—Probabilistic Principal Component Analysis (PPCA), Expectation Maximization (EM), Latent Variables, Maximum Likelihood Estimation, Image Denoising, Expected Patch Log Likelihood**

I. Introduction

A. Probabilistic Principal Component Analysis (PPCA)

Classical PCA gives an algebraic view of data. The solution to PPCA is PCA with maximum likelihood. PPCA is a method to estimate the principal axes when any data vector has one or more missing values.

B. Expectation Maximization (EM) Algorithms

EM algorithms are general-purpose algorithms for Maximum Likelihood Estimation(MLE) in wide range of situations. Some mathematics on EM algorithms are in [2]

II. Applications Of PPCA and EM

We demonstrate two image-processing related application of PPCA-EM algorithms:

- Image Reconstruction: The input is a selected image of arbitrary dimensions, in MATLAB. The standard image is attempted to be reconstructed, and we compare the outputs of:
  - Bishop closed form PPCA
  - PPCA-EM algorithm (3-Adaptation/3-iterations)
- Image De-noising: We take the standard image 'Lena' of dimensions 256*256 in MATLAB. The standard image is normal, but we convert it to a noisy image with noise variance σ^2, Now the input image is discarded and we attempt to restore an *estimated* image from the noisy image with 2 approaches:
  - The first one is an Expected Patch Log-Likelihood (EPLL) algorithm as seen in [3]
  - The second is an Expectation-Maximization Algorithm applied *on the output* of first part (EPLL) and improves the quality of restored image. (The paper which was referred to for this algorithm was [4])

III. Process flowchart

Image reconstruction and Image de-noising applied algorithms: are shown in the figure below.

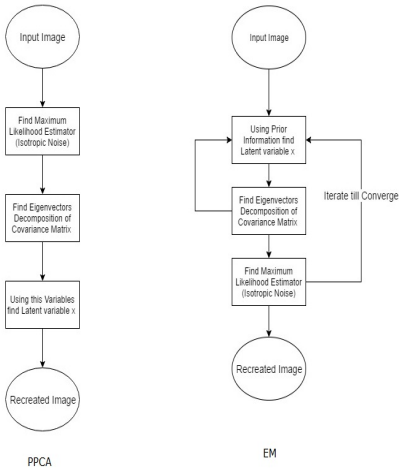

Figure 4.1: Flowchart of PPCA and EM algorithms

IV. Demonstration

1) We compare the running time, and error (matrix norm) from the input image and the reconstructed image.

	PPCA with Bishop Closed form				PPCA with EM		
Dimensions	Time(Seconds)	Norm 1 (Error)	Compression ratio	Dimensions	Time(Seconds)	Norm 1 (Error)	Compression ratio
Q = 25	20.65	8.5153	39.47%	Q = 25	20.00001	9.0668	37.50%
Q = 50	20.7787	4.7038	26.31%	Q = 50	20.007	5.4044	25.00%
Q = 100	20.831501	2.6122	11.84%	Q = 100	20.259	2.9715	10.52%
Q = 150	20.928	1.8443	3.28%	Q = 150	20.415	2.117	2.63%

Figure 4.1.1: Comparison of image reconstruction algorithms

2) The output for an adaptive EPLL algorithm

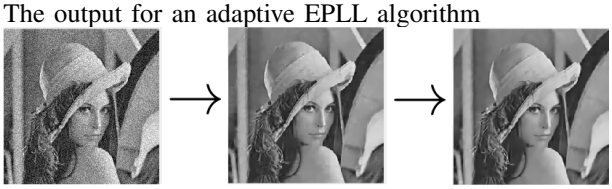

Figure 4.1.2: Output of adaptive EPLL [3]

References

[1] Tipping, Michael, and C. Bishop. *Mixtures of probabilistic principal component analyzers.*, Neural computation 11.2 (1999): 443-482.
[2] *"What are EM Algorithms?"* C. B. Do
[3] From Learning Models of Natural Image Patches to Whole Image Restoration
[4] E. Luo, S. Chan, and T. Nguyen, *Adaptive Image Denoising by MixtureAdaptation*, IEEE Trans. Image Process. 2016.