Expectation Maximization Algorithms and PPCA

School Of Engineering And Applied Science, Ahmedabad University
Apexa,Parth Satodia,Pinak Divecha,Rushita Thakkar,Vatsal Patel
Group-7 (5 x 5)

***Abstract*—Probabilistic Approach to Principle Component Analysis has many advantadges over Principle Component Analysis.The paper shows the comparative study of the retrieving the image with corrupted and missing entries using the Augmented Lagrange Multiplier and the Expectation Maximization Algorithm**

***Index Terms*—PPCA, Likelihood, EM, ALM, PCP.**

I. Introduction

PPCA addresses limitations of PCA.The EM algorithm can be used to computing the principle sub space iteratively.The ALM computes the low dimensional matrix as well as the sparse component iteratively and reconstructs the image.The image used in all the programs has dimensions 242X308. Here there are two different types of problems

1) **Corrupted Value**: Some of the data entries in given data image is corrupted in other words noise is added using inbuilt MATLAB function.

2) **Missing value**:Putting NAN values to some random pixels.

A. Robust Principle Component Analysis

Robust Principle Component Analysis Uses the Principle Component Pursuit using the Augmented Lagrange Multiplier for separating both the low rank component and the sparse component. Agorithm Converges after 1000 iterations As seen

TABLE I: RPCA for missing values

Putting NaN at every..	20	30	40	50 pixels
rms error	0.1749	0.0782	0.0678	0.0606

TABLE II: RPCA Corrupted Image

Corrupt Image(%)	1	2	3	4	5
rms Error	0.0611	0.0866	0.1049	0.1242	0.1345
time	19.23	19.75	21.06	21.77	23.021
total-time	104.831				

in the table,the error computed is quite less but the time taken for retrieval is average 21s which is quite high.

II. EM Algorithm for Probabilistic PCA

In Order to maximize the likelihood for PPCA, EM algorithm is used. In this algorithm, the latent variables xn is said to be missing data which is comprised with complete data.EM algorithm comprises of two steps. We initialise as Identity matrix and take σ as 0. Algorithm converges when the difference between 2 iteration is less than 0.001.

$$M = W^T W + \sigma^2 I$$

where σ^2 is maximum-likelihood estimator.,and E-step(expectation) and M-step(Maximization) step can be written

$$\tilde{W} = SW(\sigma^2 I + M^{-1} W^T SW)^{-1}$$

$$\tilde{\sigma^2} = \frac{1}{d} tr(S - SWM^{-1}\tilde{W^T})$$

TABLE III: PPCA with EM Corrupted Image (q=100)

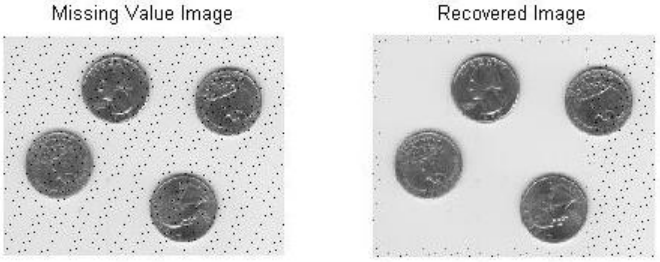
Corrupt Image (%age)	1	2	3	4	5
rms-error(Original-Corrupted)	0.0608	0.853	0.1041	0.1249	0.1368
rms Error(Original-Retrived Img)	0.0527	0.0709	0.0855	0.1013	0.1108
total-time(sec)	7.022				

As seen in above table,the error for retrieval similar images is high.But the time taken is quite less. Take a note to 2nd and 3rd row,the error for (Original - Retrived) is lesser than (Original - Input Corrputed image)

A. Conclusion

Now comparing the values of time taken and error calculated for missing entry images and corrupted entries images,the time taken for retrieval of images is quite high (Refer table 2 and table 3) by using ALM algorithm as compared to the EM algorithm.Whereas the error computed in case of EM Algorithm is a bit higher as compared to the ALM Algorithm.Now Time Complexity into consideration,we conclude that Image Retrieval using the EM Algorithm would be a better approach.So the principle components (W) obtained using the EM algorithm would be fed to the Deep Neural Network

The below figure shows the retrieval of image when some entries are corrupted using the EM algorithm.Other images are submitted in the folder.



References

[1] Micheal.E.Tipping,Christopher M.Bishop Probibilistic Principle Component Analysis.Journal of Statistical Society.Series B(Statistical Methodology),Vol.61,No 3(1999),611-612.