Comparison between Probabilistic PCA with EM Algorithm and Robust PCA

Group-3, Group Name-Pantomaths
Shreyas Patel[1401025], Amee Bhuva[1401009], Akhil Vavadia[1401095], Manini Patel[164404], Malav Vora[1401062]
School of Engineering and Applied Science, Ahmedabad University

***Abstract*—Principal Component Analysis is a technique through which we can find principal components of given data which help us in data analysis. But here we are going to analyze two such models of PCA, which help us in finding principal components from given corrupted or missing data such that squared reconstruction error should be minimized. This two models are known as Robust PCA and Probabilistic PCA. More specifically we are going to look at PPCA with Expectation Maximization algorithm.**

***Keywords*—Principal Component Analysis, Probabilistic model, EM algorithm, Latent variable, Factor analysis, Maximum-likelihood, RobustPCA.**

I. Introduction

As we all known that given observed data vectors matrix, one can find principal components and this principal components projection minimizes the squared reconstruction error. But observed data vectors may have corrupted data entries and missing entries as well, so modeling with classical PCA may lead to bad analysis, so probabilistic modeling of observed data vectors is required which leads to PCA models called Robust PCA with Augmented Lagrangian Multiplier algorithm and Probabilistic PCA with Expectation Maximization algorithm.

II. Comparison of RPCA and PPCA with EM algorithm

We have implemented this two algorithms in MATLAB and for comparison of this two algorithms, two parameters are taken into consideration. (1) Time for execution of algorithms (2) Squared Reconstruction Error (RMS error). Again, Following simulations are done in MATLAB. Here image with size of 256∗256 is taken. Image with 2185 missing entries, corrupted at density level of 2% is taken as input for given algorithms.

Error Percentage (RMS Error)			
Data type / Algorithm	2185 Missing	Corrupted with 0.02 density level	Corrupted(level of corruption 0.02 density) and 2185 Missing
RPCA	2.93%	8.15%	8.69%
PPCA with EM	1.35%	6.89%	8.63%

Fig. 1. Error rate comparison between PPCA with EM and RPCA

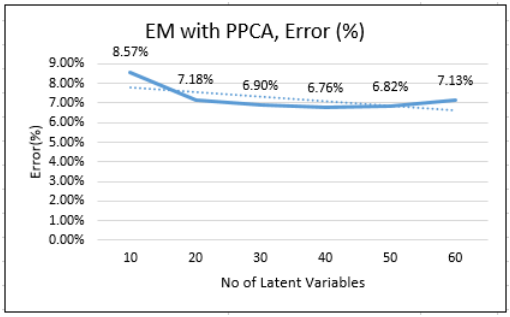
For PPCA with EM algorithm, Number of latent variables are 30 and it runs for 10 iteration. Where as for RPCA with Augmented Lagrangian Multiplier, regularization parameter(λ) is 0.02, augmented lagrangian parameter (μ) is 1, reconstruction error tolerance is 10^{-5} and it runs for 1000 iteration. As we can see PPCA with EM algorithm is more efficient compare to RPCA for given parameters.

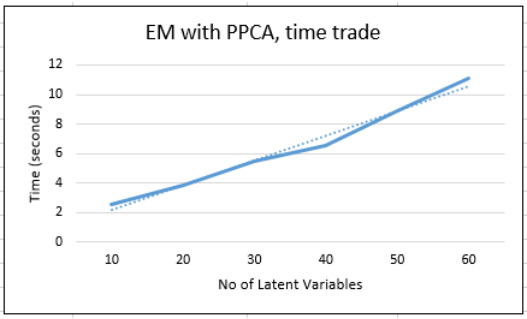
Execution Time (seconds)			
Data type / Algorithm	2185 Missing	Corrupted with 0.02 density level	Corrupted(level of corruption 0.02 density) and 2185 Missing
RPCA	76.132360 seconds	75.695431 seconds	76.434879 seconds
PPCA with EM	7.550715 seconds	3.184369 seconds	3.126828 seconds

Fig. 2. Time comparison between PPCA with EM and RPCA

III. Analysis of PPCA with EM Algorithm

This section will analyze specifically EM algorithm for PPCA with variation of latent variables. Following two graph shows that as no of latent variables increases error decreases at very slow rate but time taken increases linearly.





IV. Conclusion

PPCA with EM algorithm is more efficient then Robust PCA as it has less squared reconstruction error and time.