Probabilistic Principal Component Analysis and Expectation Maximization - A Review

Hardil Mehta (1401018), Chintan Saraviya (1401026),
Kashish Shah (1401048), Rupande Shastri (1401102), Kishan Raval (1401117)

***Abstract*—The PCA method is widely used for dimensionality reduction in various applications, but it does not have a probabilistic model. While PPCA uses eigenvalue decomposition of covariance matrix to extract the principal axis with maximum likelihood, E-M iteratively computes it using posterior probability. Despite being iterative, E-M is computationally more efficient than PPCA. Combining the two algorithms results in better efficiency.**

I. Introduction and Theory

A. Principal Component Analysis - Limitations

Principal Component Analysis (PCA) is a statistical method that reduces dimensionality while retaining most of the variance in the data. However, it has its own set of limitations:

- There is no probabilistic model to parametrize the data. This means less information to work with at the time of predicting data trends.
- The data is assumed to be along the directions of highest variance in the covariance based algorithm. The low variance directions are said to be noise. This assumption can cause loss of important data during PCA.
- For large matrices, the computational complexity is difficult to manage.

B. Latent Variable Model

Latent variables are those that can be inferred from a given set of observed variables. They are used in this probabilistic model as they give more restrained information about the dependencies of the observed variables. For a given d-dimensional vector of observed data **t**, the relationship between **t** and a corresponding q-dimensional vector of latent variables **x** is given as **t** = **Wx** + μ + ϵ.

The parameter vector μ allows data with zero mean. The error ϵ is modeled as Gaussian, $\epsilon\ \mathcal{N}(0, \Psi)$ where error covariance Ψ is estimated from observed data. The d x q matrix **W** relates **t** to **x**. The computation of **W** is the problem at hand.

C. Probabilistic Principal Component Analysis

In PPCA, we convert the latent variable model into a probabilistic model and obtain a mean μ and covariance matrix **S** of the data via Maximum Likelihood estimation or log-likelihood method. The SVD of this covariance matrix **S** is computed to derive **C**. The other part is assumed as Gaussian distributed noise. **W** is given as, $W = C + \sigma^2 I$. This gives a closed-form solution, similar to PCA.

Potential applications of PPCA include classification, novelty detection and trend prediction.

D. Expectation Maximization

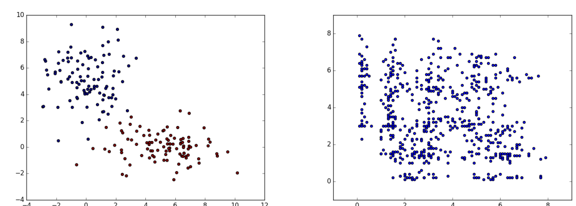
Expectation Maximization uses the latent variable model in the form $t = \mathcal{N}(\mu, W = CC^T + R)$. The E-M algorithm proceeds to compute the parameter **W** by the following steps:

- **E-step:** Compute the posterior probability $\mathbf{x}|\mathbf{t}\ \mathcal{N}(\beta(\mathbf{t} - \mu), \mathbf{I} - \beta\mathbf{C})$
  where $\beta = \mathbf{C}^T\mathbf{W}^{-1}$
- **M-step:** Using the posterior, choose **C** and **R** to maximize joint likelihood of **t,x**

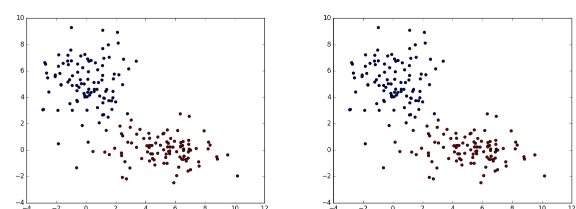
Since the E-M algorithm does not require computation of covariance matrix, it has a complexity much less than that of PPCA and PCA. The complexity of PPCA is of the order n^3), while E-M gives the same result in n^2 time.

II. Implementation and Results

The figure below shows the results of PPCA (right) for given observed data (left).



And those of E-M after 6 iterations are shown below.



III. Inference and Conclusion

While the PPCA algorithm has a better probabilistic model, E-M is computationally less complex and more efficient. Using the probability model of PPCA and the iterative computational steps of E-M, we have an algorithm with better efficiency and a probabilistic model that can be used with various probabilistic operators.

References

[1] Michael E. Tipping and Christopher M. Bishop, *Probabilistic Principal Component Analysis* Journal of the Royal Statistical Society, Series B, 61, Part 3, pp. 611−622.