Generative models using Probabilistic Principal Component Analysis

Group - 2
Bhavya Patwa - 1401063, Harsh Mehta - 1401086
Vidit Shah - 1401078, Yash Turakhia - 14010118
School of Engineering and Applied Science
Ahmedabad University

***Abstract*—The aim behind learning the Machine learning and Algorithms and Optimizations on Big Data is to replace the ruled-based expert system to probabilistic generative models. The availability of data and computational power will migrate ruled-based and manually specific model to Probabilistic data driven models. Here, we discuss a Probablistic Principal Component Analysis approach which makes use of the EM algorithm to find the maximum likelihood estimates of the model parameters. The EM algorithm is chosen regardless of having an already obtained closed-form solution for the maximum likelihood parameter values because there maybe computational advantages in spaces of high dimensionality. This EM procedure can also be extended to the factor analysis model, for which there is no closed-form solution. Finally, it allows missing data to be handled in a principled way.**

I. Introduction And Theory

A. Expectation Maximization

The goal of the algorithm is to find the parameter vector Φ that maximizes the likelihood of the observed values of X, $\mathscr{L}_c(\Phi|X)$. But in the case where this is not feasible, we associate the latent variable Z and express the model using both, to maximize the likelihood of the joint distribution of X and Z, the complete likelihood $\mathscr{L}_c(\Phi|X, Z)$. Since the Z values are not observed, we can't work directly with the $\mathscr{L}_c$; instead we work with its expectation, $\mathscr{Q}$,given X and current parameter Φ^l. This is E step of an algorithm. In the subsequent maximization(M) step, we recompute the new parameter values as Φ^{l+1}, that maximize the parameters used in the E step. The E and M steps are then repeated until a suitable convergence criterion is satisfied.

B. Probabilistic Principal Component Analysis

PCA model is used for analysing, extracting data and projecting them into lower dimensional vector space based on maximizing variance of projected data. However, this model doesnt take care of missing data. To overcome this limitation we use Probabilistic Principle Component Analysis. Other advantages of PPCA over PCA is, multiple PCA model can be combined as a probabilistic mixture. We can use PPCA to identify unknown or missing data which helps in the classification system.

C. Factor Analysis

One way to view PCA probabilistically is to relate it to latent variable models. A latent variable relates a d - dimensional observation vector t to a corresponding q - dimensional vector of latent variables x. The most common such model is factor analysis where the relationship is linear: $t = Wx + \mu + \epsilon$. Where W relates the two sets of variables, μ permits non-zero mean, ϵ is noise parameter, x is gaussian distributed random variable with zero mean and unit variance. If we specify the noise to also be gaussian then t follows a gaussian $t \rightarrow N(\mu, WW^T + \psi)$. Unfortunately, columns of W will generally not correspond to the principal subspace of the observed data. However, a link can be made using isotropic erroe model where $\psi = \sigma^2 I$. If we estimate W and σ^2 using maximum likelihood, the isotropic error model corresponds to PPCA. Using the isotropic gaussian noise model and the original factor analysis model we can obtain: $t|x \rightarrow N(Wx + \mu, \sigma^2 I)$

II. Results

Here, we compare the computational advantages of EM over the Bishop closed-form solution for maximum likelihood parameter values of PPCA.

	PPCA with Bishop Closed form		
Dimensions	Time(Seconds)	Norm 1 (Error)	Compression ratio
Q = 25	20.65	8.5153	39.47%
Q = 50	20.7787	4.7038	26.31%
Q = 100	20.831501	2.6122	11.84%
Q = 150	20.928	1.8443	3.28%

		PPCA with EM	
Dimensions	Time(Seconds)	Norm 1 (Error)	Compression ratio
Q = 25	20.00001	9.0668	37.50%
Q = 50	20.007	5.4044	25.00%
Q = 100	20.259	2.9715	10.52%
Q = 150	20.415	2.117	2.63%

Fig. 1. Feature Points Extracted

References

[1] Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 611-622.
[2] Parsons, Simon. (2005). Introduction to Machine Learning by Ethem Alpaydin, MIT Press, 0-262-01211-1, 400 pp. 432-433.
[3] Bishop. (2006). Pattern Recognition and Machine Learning, Springer Science+Business Media, 0-387-31073-8, 577-580.