Expectation Maximization Algorithms and PPCA

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

***Abstract*—The paper basically shows the Expectation Maximization and the Probabilistic Principle Component Analysis algorithms which are generative algorithms**

***Index Terms*—Generative Algorithms, PPCA, Likelihood, EM, Likelihood estimation.**

I. Introduction

The expectation maximization algorithm enables parameter estimation in probabilistic models with incomplete data iteratively.PPCA is widely used technique for dimension reduction.PPCA has many drawbacks like high sensitivity to outliers,no proper dealing of missing data.PPCA addresses limitations of PCA.Maximum-likelihood estimates can be computed for elements associated with principal components.The EM algorithm can be used to computing the principle sub space iteratively.

A. Generative Algorithms

In problems like regression ,what we do is try to model y with respect to the given data x i.e. we focus on p(y/x) (Based on input space x).Here ,the approach is to model p(x/y) and p(y).After modeling p(x/y) and p(y),bayes rule can be applied in order to find the posterior probability p(y/x).

II. Expectation Maximization

suppose we are given a training set $x(1), \ldots, x(m)$ these points do not come with any labels and we want to model the data by specifying a joint distribution $p(x(i), z(i)) = p(x(i)|z(i))p(z(i))$ where $z(i)$ s are latent random variables. For different value of $z(i)$, log likelihood $logP(x, z; \theta)$ will have multiple local maxima and .

In this model each x (i) was generated by randomly choosing $z(i)$ from $1, \ldots, k$, and then $x(i)$ was drawn from one of k Gaussian depending on $z(i)$. This is called the mixture of Gaussian model. First step is guessing a probability distribution over completions of missing data given the current model (known as the E-step) in other word guessing value of $z(i)$ and then re-estimating the model parameters using these completions. During the E-step, expectation maximization chooses a function gt that lower bounds the likelihood function everywhere, and for which $g\theta^t = logP(x; \theta^t)$. During the M-step, the expectation maximization algorithm moves to a new parameter set θ^{t+1} that maximizes g_t. EM causes the likelihood to converge monotonically.One reasonable convergence test would be to check if the increase in () between successive iterations is smaller than some tolerance parameter, and to declare convergence if EM is improving () too slowly. There are other numerical optimization techniques, such as gradient descent or Newton-Raphson, could be used instead of expectation maximization but expectation maximization has the advantage of being simple, robust and easy to implement.

III. Probabilistic Principle Component Analysis

PPCA model or factor analysis is given by below equation which is linear relation.

$$t = Wx + \mu + \epsilon \tag{1}$$

Where,
t is d-dimensional observed vector,
W is a $d \times q$ matrix relates the two sets of variables,
x is q dimensional latent variable (unobserved),
μ is a nonzero mean of this model,
ϵ is error or noise.
Latent variable is independent and Gaussian with unit variance x $\sim$ N(0, I).

Error variable is Gaussian that is $\varepsilon \sim$ N(0, ψ).

t is also a Gaussian distribution that is $t \sim N(W^T$+$\psi)$.

Error covariance is diagonal matrix whose elements can be estimated from data. From one value of latent variable, the observed variables t_i are conditionally independent.Thus these latent variables are explaining the correlations between observation variables. ϵ_i represents variability unique to a particular t_i.The conditional probability of t given latent variable x is given by

$$\mathbf{t}|\mathbf{x} \sim \mathcal{N}(Wx + \mu, \sigma^2\mathbf{I}) \tag{2}$$

Estimation of is given by the mean of the data, where S is the sample covariance matrix of the observed data.

Where,

$$S = \frac{1}{N}\sum_{n=1}^{N}(t_n - \mu)(t_n - \mu)^T \tag{3}$$

Estimates for W and σ^2 may be obtained by iterative maximization of log-likelihood. It can be done by either using EM-algorithm or using maximum likelihood estimators.

References

[1] Chuong B Do Serafim Batzoglou What is the expectation maximization algorithm? ,2008 Nature Publishing Group
[2] 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.
[3] Stanford Online Machine Learning Hand outs.