Homework # 3

Final Report

On

Implementation of

EM Algorithm

For Training

Gaussian

Mixture Model

Submitted by:- To:-

Gurpreet Singh Dr. Changhe Yuan

CS 3813(Machine Learning)

**Contents**

* Introduction: EM algorithm & ­­Gaussian Mixture Model
* Total Log Likelihood against Number of EM Iterations (Table + Graph + Analysis)
* Initial Parameters vs. Number of EM Iterations(Table + Analysis)
* Final Observations

**Introduction: EM** **Algorithm**

An expectation–maximization (EM) algorithm is an [iterative method](http://en.wikipedia.org/wiki/Iterative_method) for finding [maximum likelihood](http://en.wikipedia.org/wiki/Maximum_likelihood) or [maximum a posteriori](http://en.wikipedia.org/wiki/Maximum_a_posteriori) (MAP) estimates of [parameters](http://en.wikipedia.org/wiki/Parameter) in [statistical models](http://en.wikipedia.org/wiki/Statistical_model), where the model depends on unobserved [latent variables](http://en.wikipedia.org/wiki/Latent_variable). The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the [log-likelihood](http://en.wikipedia.org/wiki/Likelihood_function#Log-likelihood) evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

Filtering and smoothing EM algorithms arise by repeating the following two-step procedure.

**E-Step**

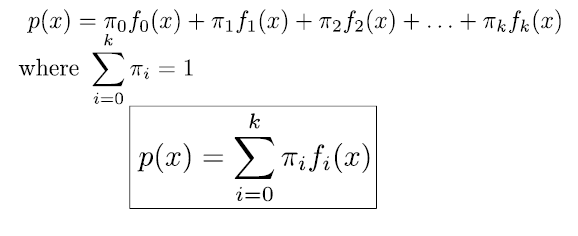
Operate a minimum-variance smoother designed with current parameter estimates to obtain updated state estimates.

**M-Step**

Use the filtered or smoothed state estimates within maximum-likelihood calculations to obtain updated parameter estimates.

**Gaussian Mixture Model**

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model.



**Experiment Results:**

1) Total Log Likelihood against Number of EM Iterations

|  |  |  |
| --- | --- | --- |
| S.no. | Total  Log Likelihood | Number of  EM Iterations |
| 1 | -4146 | 0 |
| 2 | -4034 | 2 |
| 3 | -3875 | 4 |
| 4 | -3856 | 8 |
| 5 | -3845 | 10 |
| 6 | -3828 | 13 |
| 7 | -3825 | 16 |
| 8 | -3824.74 | 20 |
| 9 | -3824.23 | 29 |
| 10 | -3824.14 | 35 |
| 11 | -3824.037 | 42 |
| 12 | -3824.010 | 63 |

**Table 1: Total log likelihood against the Number of EM Iterations**

**Graph 1🡪Table1: Total Log Likelihood against Number of EM Iterations**

**Analysis:**- One can see in Graph 1, how Total Log Likelihood increases abruptly in the beginning and then almost becomes constant. In the constancy region (TLL > -3824) of Total Log Likelihood, we find best set of Parameters for Gaussian Mixture Model. As we can see computing after -3824 is worthless, as Total Log Likelihood increases too slow after that.

2) Co-Variance (Initial Parameter) vs. Number of EM Iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no. | Sigma  0 | Sigma  1 | Sigma  3 | Number of EM Iterations |
| 1 | 2 | 4 | 3 | 71 |
| 2 | 7 | 6 | 5 | 53 |
| 3 | 7 | 6 | 15 | 52 |
| 4 | 7 | 5 | 15 | 50 |
| 5 | 7 | 4 | 15 | 48 |
| 6 | 7 | 3 | 15 | 47 |
| 7 | 6 | 3 | 14 | 43 |
| 8 | 5 | 3 | 14 | 38 |
| 9 | 3 | 3 | 14 | 10 |
| 10 | 3 | 3 | 5 | 10 |
| 11 | 3 | 3 | 7 | 9 |
| 12 | 1.28 | 1.968 | 1.69 | 3 |

**Table 2: Initial Parameters vs. Number of EM Iterations**

**Analysis of Table 2:**

As one can see in Table 2, how change initial Co Variance values affects the number of EM Iterations. Initial Higher Values of Co-Variance lead to higher number of EM Iterations. This is because of the difference(Initial Value – True Value) is too high.

**# No Singularities are encountered even after trying Multiple sets of Initial Values on both Data Sets Provided.**

**Final Observations:-**

1. The Initial Values of Co Variance affects the Number of EM Iterations very much. Higher Values lead to higher number of Iterations.
2. Final Set of Parameters is same (approximate is same) no matter what initial parameters you started with some exceptions.
3. Initial Parameters mainly affects the Number of EM Iterations.
4. If you initialize the parameters with final set of parameters from previous experiment, the number of Iterations will reduce to 2 or 3.