CSE 472

Report on

Expectation-Maximization Algorithm for Gaussian Mixture Model

Student ID:1305076

Date of Submission: 18 May, 2018

Question 1: Why should you use a Gaussian mixture model (GMM) in the above scenario?

Answer

Reasons for using GMM in this scenario:

- We know that the data points have been generated from a finite number (three) of Gaussian distributions forming clusters with different means and standard deviations.

- We do not have labeled data (unsupervised clustering) and one point can be assigned to more than one clusters.

- Here we have a hidden indicator variable Z_{ij} , where Z_{ij} is 1 if datum x_j was generated by the ith component and 0 otherwise.

Question 2: How will you model your data for GMM?

Answer

If we knew which component generated each data point, then it would be easy to recover the component parameters. On the other hand, if we knew the parameters of each component, then we could assign each data point to a component probabilistically. In this case we know neither the assignments nor the parameters.

In GMM, we try to get around this problem by pretending that we know the parameters of the model and then to infer the probability that each data point belongs to each component. After that, we refit the components to the data, where each component is fitted to the entire data set with each point weighted by the probability that it belongs to that component. The process iterates until convergence.

Question 3: What are the intuitive meaning of the update equations in the M-step?

Answer

The intuition is to increase the log-likelihood of the data at every iteration.

In the M-step (maximization step) we use the expected values P_{ij} of the hidden indicator variables Z_{ij} (Z_{ij} is 1 if datum x_j was generated by the ith component and 0 otherwise). The M-step finds the new values of the parameters that maximize the log likelihood of the data, given the expected values of the hidden indicator variables.

Question 4: Derive the log likelihood function in step 4.

Answer

Lets assume we have N data points $x_1, x_2, \ldots x_N$ and k distributions where μ_i, Σ_i are the mean and covariance respectively of the ith distribution.

The likelihood function is

$$L = P(x)$$

$$= \prod_{j=1}^{N} P(x_j)$$ [As the data points are I.I.D (Independent and Identically Distributed) random variables]

Taking the log,

$$\text{Log } L = \sum_{j=1}^{N} \log P(x_j)$$

Maximizing the log likelihood is equivalent to maximizing the likelihood.

If there was only one distribution $\Rightarrow$ $\text{Log } L = \sum_{j=1}^{N} N(x_j|\mu, \Sigma)$

But in our case,

$$\text{Log } L = \sum_{j=1}^{N} \log\left\{ \sum_{i=1}^{3} N_i(x_j|\mu_i, \Sigma_i)\, \omega_i \right\}$$

Here, $N_i(x_j|\mu_i, \Sigma_i)$ $\Rightarrow$ probability of coming from ith distribution.

ω_i $\Rightarrow$ probability of ith distribution.