CSE472

Machine Learning Sessional

# Assignment 4: Expectation-Maximization Algorithm for Gaussian Mixture Model

A. S. M. Ahsan-Ul-Haque

1205021

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

Ans.

The given scenario describes some ships that we are trying to locate using sonar sensors. Although the exact number of ships is known, the problem arises because sonar sensors are hit with interference and noise from several close proximity ships.

There are 2 reasons I would use a Gaussian model in the given scenario:

i. Noise and interference signals generally assume a normal distribution. A Gaussian model is appropriate for this case.
ii. Because there might be several ships in the given scenario, we need to figure out all of their Gaussian distributions (mean and covariance) which fit the data best.
    GMM fits the situation perfectly and provides the expected means (locations of the ships).

2. How will you model your data for GMM?
    Ans.
    For each data point I would try to find a distribution of the data point over the models (z). (E step)
    Using this distribution, I would calculate the weighted mean and weighted covariance of each distribution and also update the priors of the distribution (M step), until the mean and covariance converges.

3. Derive the update equations in **M step**. (To make the derivations short you can use formulas from matrix calculus)

Ans.

Since we have assumed that each of the individual models is a Gaussian, the quantity $p(\mathbf{x}_i|m,\theta)$ is simply the conditional probability of generating $\mathbf{x}_i$ given that the $m^{th}$ model is chosen:

$$\log p(\mathbf{x}_i|z_{im}=1;\boldsymbol{\theta}) = \frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_m\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)^T \boldsymbol{\Sigma}_m^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)\right\} \quad (1)$$

Taking expectations w.r.t. $Q(\mathbf{Z})$ we get:

$$\langle l_c(\boldsymbol{\theta})\rangle_{Q(\mathbf{Z})} = \sum_{i=1}^{N}\sum_{m}^{M} \langle z_{im}\rangle \log p(\mathbf{x}_i|z_{im}=1;\boldsymbol{\theta}) + \langle z_{im}\rangle \log \pi_m \quad (2)$$

The “M” step in EM takes the expected complete log-likelihood and maximizes it w.r.t. the parameters that are to be estimated; in this case, prior $\pi_m$, mean $\mu_m$, and covariance $\boldsymbol{\Sigma}_m$.

Differentiating eq. (2) w.r.t. $\mu_m$ we get:

$$\frac{\partial \langle l_c(\boldsymbol{\theta})\rangle_{Q(\mathbf{Z})}}{\partial \boldsymbol{\mu}_m} = \sum_{i=1}^{N} \langle z_{im}\rangle \frac{\partial}{\partial \boldsymbol{\mu}_m} \log p(\mathbf{x}_i|z_{im}=1;\boldsymbol{\theta}) = \mathbf{0} \quad (3)$$

We can compute $\frac{\partial}{\partial \boldsymbol{\mu}_m} \log p(\mathbf{x}_i|z_{im}=1;\boldsymbol{\theta})$ using eq. (1) as follows:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\mu}_m} \log p(\mathbf{x}_i|z_{im}=1;\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\mu}_m} \log\left\{\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_m\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)^T \boldsymbol{\Sigma}_m^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)\right\}\right\} \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}_m}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)^T \boldsymbol{\Sigma}_m^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}_m\right) \\
&= \left(\mathbf{x}_i-\boldsymbol{\mu}_m\right)^T \boldsymbol{\Sigma}_m^{-1\dagger}
\end{aligned}$$

Substituting this result into eq. (3), we get:

$$\sum_{i=1}^{N} \langle z_{im} \rangle \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T \mathbf{\Sigma}_m^{-1} = \mathbf{0}$$

giving us the update equation:

$$m = \frac{\sum_{i=1}^{N} \langle z_{im} \rangle \, \mathbf{x}_i}{\sum_{i=1}^{N} \langle z_{im} \rangle} \tag{4}$$

Differentiating eq. (2) w.r.t. $\mathbf{\Sigma}^{-}{}_{m}{}^{1}$ we get:

$$\frac{\partial \left\langle l_c(\boldsymbol{\theta}) \right\rangle_{Q(\mathbf{Z})}}{\partial \mathbf{\Sigma}_m^{-1}} = \sum_{i=1}^{N} \langle z_{im} \rangle \frac{\partial}{\partial \mathbf{\Sigma}_m^{-1}} \log p(\mathbf{x}_i | z_{im} = 1; \boldsymbol{\theta}) = \mathbf{0} \tag{5}$$

We can compute $\frac{\partial}{\partial \mathbf{\Sigma}_m^{-1}} \log p(\mathbf{x}_i | z_{im} = 1; \boldsymbol{\theta})$ using eq. (1) as follows:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{\Sigma}_m^{-1}} \log p(\mathbf{x}_i | z_{im} = 1; \boldsymbol{\theta}) &= \frac{\partial}{\partial \mathbf{\Sigma}_m^{-1}} \log \left\{ \frac{1}{(2\pi)^{d/2} \left|\mathbf{\Sigma}_m\right|^{1/2}} \exp\left\{ -\frac{1}{2} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T \mathbf{\Sigma}_m^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right) \right\} \right\} \\
&= \frac{\partial}{\partial \mathbf{\Sigma}_m^{-1}} \left\{ \frac{1}{2} \log \left|\mathbf{\Sigma}_m^{-1}\right| - \frac{1}{2} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T \mathbf{\Sigma}_m^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right) \right\} \\
&= \frac{1}{2} \mathbf{\Sigma}_m - \frac{1}{2} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right) \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T \ddagger
\end{aligned}$$

Substituting this result into eq. (5), we get:

$$\sum_{i=1}^{N} \langle z_{im} \rangle \left( \frac{1}{2} \mathbf{\Sigma}_m - \frac{1}{2} \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right) \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T \right) = \mathbf{0}$$

giving us the update equation:

$$\mathbf{\Sigma}_m = \frac{\sum_{i=1}^{N} \langle z_{im} \rangle \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right) \left(\mathbf{x}_i - \boldsymbol{\mu}_m\right)^T}{\sum_{i=1}^{N} \langle z_{im} \rangle} \tag{6}$$

We now differentiate this new expression w.r.t. each $\pi_m$ giving us:

$$\frac{\partial}{\partial \pi_m} \left\langle l_c(\boldsymbol{\theta}) \right\rangle_{Q(\mathbf{Z})} - \lambda = 0 \quad \text{for } 1 \leq m \leq M$$

Using eq. (2) we get:

for $1 \leq m \leq M$ (8)

or

$$\left.\begin{aligned}\frac{1}{\pi_m}\sum_{i=1}^{N}\langle z_{im}\rangle - \lambda = 0 \\ \sum^{N}\langle z_{im}\rangle - \lambda\pi_m = 0\end{aligned}\right\}$$

equivalently

$i$=1

Where we have used the relation

$$\frac{\partial}{\partial \mathbf{X}} \log|\mathbf{X}| = \left(\mathbf{X}^{-1}\right)^T \text{ and } \frac{\partial}{\partial \mathbf{A}}\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}\mathbf{x}^T$$

Summing eq. (8) over all M models we get:

$$\sum_{m}^{M}\sum_{i=1}^{N}\langle z_{im}\rangle - \lambda\sum_{m}^{M}\pi_m = 0$$

But since $\sum_m^M \pi_m = 1$ we get:

$$\lambda = \sum_{m}^{M}\sum_{i=1}^{N}\langle z_{im}\rangle = N \tag{9}$$

Substituting this result back into eq. (8) we get the following update equation:

$$\pi_m = \frac{\sum_{i=1}^{N}\langle z_{im}\rangle}{N} \tag{10}$$

4. Derive the log-likelihood function in step 4.

Ans.

$$\log(p(X|\theta)) = \log(\prod_{i=1}^{N} p(xi\ |\mu, \Sigma))$$

$$= \sum_{i=1}^{N} \log(p(x\ i\ |\mu, \Sigma))$$

$$= \sum_{i=1}^{N} \log(\sum_{j=1}^{M} p(xi|\mu j, \Sigma j)\ p(model = j))$$

$$= \sum_{i=1}^{N} \log(\sum_{j=1}^{M} p(xi|\mu j, \Sigma j)\theta j\ )$$

$$= \sum_{i=1}^{N} \log(\sum_{j=1}^{M} N(xi|\mu j, \Sigma j)\theta j\ )$$