Report on
Assignment 4:
Expectation-Maximization Algorithm
For Gaussian Mixture Model

Name:Ishita Haque

Student id-1205069

Machine Learning Lab

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

   Ans. In the above scenario, there are several ships nearby, as a result sound signals cause interference. We do not know from which ship actually was the sonar data generated. Data points are not generated from a single model, so the distribution cannot be expressed by a single function. So we need a mixture model.

   Also, sonar data takes a normal distribution where mean corresponds to the estimated location of ship and response from surrounding objects is due to the variance of the distribution. For this reason we chose a Gaussian mixture model.

2. How will you model your data for GMM?

   Ans. Each sonar data with dimension D has been considered as x vector in our GMM model. This data can be generated from any one of the K Gaussian distribution. In this case K=number of ships=number of Gaussian distributions. The probability of selection of x from Gaussian distribution j is given by $N(x_i; \mu_m, \Sigma_m)$, where mean and variance of mth Gaussian distribution is expressed as μ_m and Σ_m.

   Each data point is generated according to the following algorithm:

   1: for i = 1 to N do
   2: m ←index of one of the M models randomly selected according to the prior probability vector _
   3: Randomly generate x_i according to the distribution $N(x_i; \mu_m, \Sigma_m)$
   4: end for

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

   Ans. The steps of derivation of update equations in M step are given below:

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

2.1 The M step

The "M" step in EM takes the expected complete log-likelihood as defined in eq. (2) and maximizes it w.r.t. the parameters that are to be estimated; in this case π_m, $\boldsymbol{\mu}_m$, and $\boldsymbol{\Sigma}_m$.

Differentiating eq. (2) w.r.t. $\boldsymbol{\mu}_m$ we get:

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

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

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

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

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

giving us the update equation:

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

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

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

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

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

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

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

giving us the update equation:

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

In order to maximize the expected log-likelihood in eq. (2) w.r.t. π_m, we have to keep in mind that the maximization has the constraint that $\sum_m^M \pi_m = 1$. In order to enforce this constraint we use the Lagrange multiplier λ, and augment eq. (2) as follows:

$$L'(\boldsymbol{\theta}) = \langle l_c(\boldsymbol{\theta}) \rangle_{Q(\mathbf{Z})} - \lambda \left(\sum_m^M \pi_m - 1 \right) \tag{7}$$

We now differentiate this new expression w.r.t. each π_m giving us:

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

Using eq. (2) we get:

$$\left.\begin{aligned} \frac{1}{\pi_m} \sum_{i=1}^N \langle z_{im} \rangle - \lambda = 0 \\ \text{or equivalently } \sum_{i=1}^N \langle z_{im} \rangle - \lambda \pi_m = 0 \end{aligned}\right\} \qquad \text{for } 1 \leq m \leq M \tag{8}$$

‡Where we have used the relation $\frac{\partial}{\partial \mathbf{X}} \log |\mathbf{X}| = (\mathbf{X}^{-1})^T$ 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:

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

which preserves the constraint that $\sum_m^M \pi_m = 1$.

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

Ans: Log-likelihood function in step 4 is derived below-

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

$$= \prod_{i=1}^{N}\sum_{j=1}^{K} p(x_i, z_i=j|\mu,\Sigma,\theta)$$

$$= \prod_{i=1}^{N}\sum_{j=1}^{K} p(x_i|z_i=j,\mu,\Sigma,\theta)\, p(z_i=j|\mu,\Sigma,\theta)$$

$$= \prod_{i=1}^{N}\sum_{j=1}^{K} p(z_i=j|\mu,\Sigma,\theta)\, p(x_i|z_i=j,\mu,\Sigma,\theta)$$

$$= \prod_{i=1}^{N}\sum_{j=1}^{K} p(z_i=j|\theta)\, p(x_i|z_i=j,\mu,\Sigma)$$

$$= \prod_{i=1}^{N}\sum_{j=1}^{K} \theta_j N(x_i|\mu_j,\Sigma_j)$$

Taking ln in both sides -

$$\ln p(X|\mu,\Sigma,\theta) = \ln\left(\prod_{i=1}^{N}\sum_{j=1}^{K}\theta_j N(x_i|\mu_j,\Sigma_j)\right)$$

$$= \sum_{i=1}^{N}\ln\left(\sum_{j=1}^{K}\theta_j N(x_i|\mu_j,\Sigma_j)\right)$$