CSE472 (Machine Learning Sessional)
L-4, T-2, January 2018 Term

Report on Assignment 2:
Expectation-Maximization Algorithm for Gaussian Mixture Model

Student Id: 1305003
Name: Abdullah Al Mamun

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

   Answer:
   The data points are generated following a gaussian distribution, if we consider the actual location to be the mean, and the fluctuations of the signals to be the covariance. In this case, we have to find out the mean and covariance matrix for each ship. We can cluster the datapoints using GMM and also measure the goodness of our algorithm. So, we should use GMM.

2. How will we model our data for GMM?

   Answer:
   The actual/estimated locations of the ships will be the mean parameters of the distributions. Location of each signal will be a single datapoint. Now from all the datapoints, we will separate them in three different clusters. We will use EM algorithm which initially guesses the parameters then finds the probabilities for each point to come from different distributions, and then re-estimate the parameters. The performance metric is the loglikelihood. The program will terminate when the loglikelihood converges.

3. What are the intuitive meaning of the update equations in **M step**?

   Answer: In M step, we re-estimate the parameters with the probability matrix and the feature values. We calculate mean of some values by dividing their sum by the number of elements. In this case, the values have probabilities associated with them. So, we take a weighted average by multiplying the values with their probabilities and also divide by the summation of all probabilities instead of number of datapoints.

   In case of covariance matrix, we perform $\boldsymbol{\Sigma}_i = \frac{\sum_{j=1}^{N} p_{ij}(\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T}{\sum_{j=1}^{N} p_{ij}}$

The difference of each value from its mean value is taken in a matrix. Then the matrix is multiplied with its transpose. So, we get a DxD covariance matrix where D is the dimension of the feature vector. In case of two-dimensional feature vector, the product matrix will be

$$\begin{matrix} (x_1 - \mu_1)^2 & (x_1 - \mu_1)(x_2 - \mu_2) \\ (x_1 - \mu_1)(x_2 - mu_2) & (x_2 - \mu_2)^2 \end{matrix}$$

Adding them and with associated probabilities and dividing by the sum of the probabilities we get the covariance matrix.

The last part is the process of updating weights. Summation of all the elements of the P matrix equals to N = number of datapoints. So, adding all the probabilities for a distribution and dividing it by the number of datapoints, we get the weight of that distribution.

4. Let's Derive the log-likelihood function in step 4.
Answer:

Likelihood of a single gaussian distribution,

$l = \prod_{j=1}^{N} p(x_j | \mu, \Sigma)$

Log likelihood will be $L = \log \left(\prod_{j=1}^{N} p(x_j | \mu, \Sigma) \right)$

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

In case of multiple distributions, $p(x | \mu, \Sigma)$ will be replaced by the weighted summation of probabilities of the datapoint to come from all distributions.

So, $L = \sum_{j=1}^{N} \log\left(\sum_{i=1}^{k} wi * p(xj | \mu i, \Sigma i) \right)$