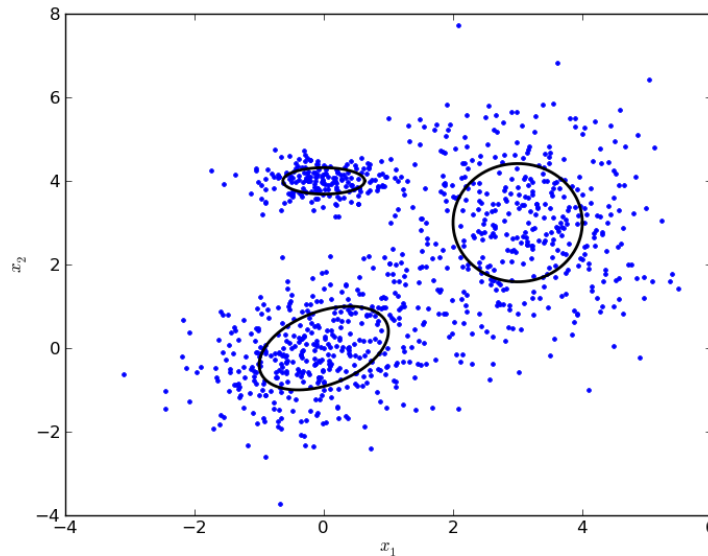


# Expectation-Maximization Algorithm for Gaussian Mixture Model

## Description of the Problem

Let's assume you have been hired by Bangladesh Navy as a data analyst for their newly acquired Ming-class submarines. They are planning to use these submarines to secure national maritime border. One of their potential targets is pirate ship. They use active acoustic sonar to find floating ships. By transmitting a sound pulse and receiving the echo on an array, the sonars can determine the direction of the echoes that return from objects hit by the sound. They can also measure the time it takes for echoes to return and calculate the distance to the object causing the echo. Unfortunately, if there are several ships nearby, the signals cause interference. Using periscope visuals, it is possible to determine the number of ships but exact location of each ship is difficult to locate. Without precise location, a torpedo hit is not guaranteed. The 2D plot of a sample sonar data is given in Figure 1. There are three ships in surface water, one inside each black circle. Each blue dot represents a single location measurement of one of the three ships. You have been instructed to build a system which takes the sonar data as well as the number of ships as input and provide the estimated location of each ship.



Since you are a new recruit, you are not yet well versed in applying machine learning model to real world problem. So, you plan to consult Dr. Google who is an expert of virtually every field. He informs you that you should use Gaussian mixture model in the above mentioned case. One possible implementation for this model is an EM (expectation-maximization) algorithm. He is also generous enough to give you a pseudocode for this which is given below.

Let a vector  $\mathbf{x}$  with dimension  $D$  can be generated from any one of the  $k$  Gaussian distribution where the probability of selection of Gaussian distribution  $i$  is  $w_i$  where  $\sum_{i=1}^k w_i = 1$  and the probability of generation of  $\mathbf{x}$  from Gaussian distribution  $i$  is given as,

$$N_i(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_i|}} e^{\left(-\frac{1}{2}(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i)\right)}$$

To learn a Gaussian mixture model using EM algorithm, we need to maximize the likelihood function with respect to the parameters (comprising the means and covariances of the components and the mixing coefficients). The steps are given below.

1. Initialize the means  $\boldsymbol{\mu}_i$ , covariances  $\boldsymbol{\Sigma}_i$  and mixing coefficients  $w_i$ , and evaluate the initial value of the log likelihood.

2. **E step:** Evaluate the conditional distribution of latent factors using the current parameter values

$$p_{ij} = p(z_j = i | \mathbf{x}_j, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \frac{p(\mathbf{x}_j | z_j = i, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) P(z_j = i | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})}{p(\mathbf{x}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})} = \frac{w_i N_i(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{i=1}^k w_i N_i(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}$$

3. **M step:** Re-estimate the parameters using the conditional distribution of latent factors

$$\begin{aligned} \boldsymbol{\mu}_i &= \frac{\sum_{j=1}^N p_{ij} \mathbf{x}_j}{\sum_{j=1}^N p_{ij}} \\ \boldsymbol{\Sigma}_i &= \frac{\sum_{j=1}^N p_{ij} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T}{\sum_{j=1}^N p_{ij}} \\ w_i &= \frac{\sum_{j=1}^N p_{ij}}{N} \end{aligned}$$

4. Evaluate the log likelihood and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to step 2.

$$\ln p(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \sum_{j=1}^N \ln p(\mathbf{x}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}) = \sum_{j=1}^N \ln \left( \sum_{i=1}^k w_i N_i(\mathbf{x}_j | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right)$$

Now you are excited to get this detailed model and algorithm. However, since this is an issue of national security you want to make sure that the suggestion from Dr. Google is indeed correct. At the outset you want to answer the following questions in your report.

1. Why should you use a Gaussian mixture model (GMM) in the above scenario?
2. How will you model your data for GMM?
3. What are the intuitive meaning of the update equations in **M step**?
4. Derive the log-likelihood function in step 4.
5. Implement the above pseudocode and estimate the location of enemy ships.

