EM implementation

Now we will cluster the two-dimensional data assuming a Gaussian mixture model using the EM algorithm.

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 k is w_k where $\sum_{k=1}^{K} w_k = 1$ and the probability of generation of $\mathbf{x}$ from Gaussian distribution k is given as,

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

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 $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ 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

$$p_{ik} = p(z_i = k|\boldsymbol{x}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w}) = \frac{p(\boldsymbol{x}_i|z_i = k, \boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w})P(z_i = k|\boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w})}{p(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w})}$$
$$= \frac{w_k N_k(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^{K} w_k N_k(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}$$

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

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{N} p_{ik}\boldsymbol{x}_i}{\sum_{i=1}^{N} p_{ik}}$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} p_{ik}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^{N} p_{ik}}$$

$$w_k = \frac{\sum_{i=1}^{N} p_{ik}}{N}$$

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

$$\ln p(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w}) = \sum_{i=1}^{N} \ln p(\boldsymbol{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma},\ \boldsymbol{w}) = \sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} w_k N_k(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)$$