The objective function for GMM is

$$\ell(p,\mu,\Sigma) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} p_k N(x_i; \mu_k, \Sigma_k)$$

where $p_k = P(s=k)$ And s_i indicates which ~~of~~ component observation x_i belongs to.

- The E-step computes the posterior for s_i given the current parameters.

$$q(s_i) = P(s_i=k \mid x_i, \mu_k, \Sigma_k) \propto P(x_i \mid s_i=k, \mu_k, \Sigma_k)$$

$$P(s_i=k \mid \mu_k, \Sigma_k)$$

$$q(s_i=k) = \frac{p_k N(x_i, \mu_k, \Sigma_k)}{\sum_{k=1}^{K} p_k N(x_i, \mu_k, \Sigma_k)}$$

- In the M-step we optimize E

$$E = \sum_{i,k} q(s_i=k) \log P(s_i=k \mid \mu_k, \Sigma_k) \, P(x_i \mid s_i=k, \mu_k, \Sigma_k)$$

$$= \sum_{i,k} w_{ik}\left[\log p_k + \log N(x_i, \mu_k, \Sigma_k)\right] \quad \text{where } w_{ik} = \frac{q(s_i=k)}{\sum_{k=1}^{K} q(s_i=k)}$$

- optimazation is done by setting the partial derivatives of E to zero

In univariate Case:

$$\cancel{\frac{\partial E}{\partial \mu_k} = \sum_{i=1}^{n} w_{ik} \frac{(x_i - \mu_k)}{2\sigma_k^2} = 0 \Rightarrow \mu_k = \sum_{i=1}^{n} \frac{w_{ik} x_i}{w_{ik}}}$$

In univariate case:

$$\frac{\partial \epsilon}{\partial \mu_k} = \sum_{i=1}^{n} w_{ik} \frac{x_i - \mu_k}{2\sigma_k^2} = 0 \Rightarrow \mu_k = \frac{\sum_{i=1}^{n} w_{ik} x_i}{\sum_{i=1}^{k} w_{ik}}$$

$$\frac{\partial \epsilon}{\partial \sigma_k} = \sum w_{ik} \left[-\frac{1}{\sigma_k} + \frac{(x_i - \mu_k)^2}{\sigma_k^2} \right] = 0$$

$$\Rightarrow \quad \sigma_k^2 = \frac{\sum_i w_{ik} (x_i - \mu_k)^2}{\sum_{i=1}^{n} w_{ik}}$$

$$\frac{\partial \epsilon}{\partial p_k} = \sum_i w_{ik} \frac{n}{p_k}$$

$$\frac{\partial \epsilon}{\partial p_k} + \lambda = 0 \Rightarrow p_k = \frac{\sum_{i=1}^{n} w_{ik}}{n}$$

where λ is a Lagrange multiplier ensuring that the mixing proportions sum to unity.