Exercise 1

Consider a Gaussian mixture model in which the marginal distribution $p(\mathbf{z})$ for the latent variable is given by $p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$ (where $z_k \in \{0,1\}$ and $\sum_k z_k = 1$), and the conditional distribution $p(\mathbf{x}|\mathbf{z})$ for the observed variable is given by $p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$. Show that the marginal distribution $p(\mathbf{x})$, obtained by summing $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$ over all possible values of $\mathbf{z}$, is a Gaussian mixture of the form $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.

Exercise 2

Robotic arms are widely used for conducting robotics research. We designed two algorithms $(A1, A2)$ for catching objects thrown towards a robotic arm. Algorithm 1 catches the objects with an unknown success rate of θ ($p_1 = P(Success|A1) = P(1|A1) = \theta$) while Algorithm 2 has a 50 percent success rate ($p_2 = p(Success|A2) = p(1|A2) = 0.5$). We ran the two algorithms several number of times and recorded their results (success $x = 1$ or failure $x = 0$). The algorithms were choosen randomly. Unfortunately after n experiments, we realize that we recorded only the results (success or failure) $\boldsymbol{X} = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ without recording the identity ($A1$ or $A2$) of the algorithms $\boldsymbol{Z} = \{\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \dots, \boldsymbol{z}^{(n)}\}$ when performing experiments.

Since repeating experiments on real robot can be a costly and time consuming process, we are interested in estimating the success rate for Algorithm 1 with EM, by only using the incomplete data.

1. Write down the complete data log-likelihood if the identity of the algorithm at each trial was also recorded in the form of a discrete vector $\boldsymbol{z}$ where k-th element of $\boldsymbol{z}$ can be either 0 or 1 ($z_k \in \{0,1\}$) and $\sum_k z_k = 1$. (If the algorithm at i^{th} trial is $A1$ then $z_1^{(i)} = 1$ and $z_2^{(i)} = 0$ or simply $\boldsymbol{z}^{(i)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$)

2. In EM we have an old estimate for parameters $\boldsymbol{\theta}^{old}$ and the goal is to derive a better estimate of $\boldsymbol{\theta}$. In E-step $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = \mathbb{E}_z \left[\ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})|\boldsymbol{X}, \boldsymbol{\theta}^{old}\right]$ is calculated, where $\ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})$ is the complete data log-likelihood (which you have calculated in the previous step). Show that the Q-function for the given problem can be written as:

$$Q(\theta, \theta^{old}) = \sum_{i=1}^{n} \sum_{k=1}^{2} \gamma(z_k^{(i)}) \left\{\log \pi_k + x^{(i)} \log p_k + (1 - x^{(i)}) \log(1 - p_k)\right\}$$

3. In M-step a revised parameter estimate is calculated as $\boldsymbol{\theta}^{new} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$. Calculate the update equation for the parameter θ (probability of success for Algorithm 1).