Reading: Sec 11.1 - 11.4

1. Mixture Normal
   (a) Write down the density function of mixture normal
   (b) What are the constraints of the parameters?

2. EM algorithm
   (a) Why do we need EM algorithm? What is it designed for?
   (b) Consider the EM Algorithm with Single Observation. Let X be the observed variable, with density function $p_\theta(X)$ (here θ is a parameter). Let Z be the latent variable, and assigned an ARTIFICIAL distribution $Q(Z)$ to it (with density $q(Z)$)
      i. Let $p_\theta(X, Z)$ be the true joint density function of (X, Z) and $p_\theta(Z \mid X)$ be the true conditional density function of Z given X. Write $p_\theta(X)$ and $p_\theta(Z \mid X)$ in terms of $p_\theta(X, Z)$.
      ii. For continuous Z, show that the log likelihood function is
      $$\log p_\theta(X) = \int q(Z) \log \frac{p_\theta(X, Z)}{q(Z)} dZ + D_{\mathrm{kl}}\left(Q(Z) \| P_\theta(Z \mid X)\right)$$
      iii. So, with the identity above, we actually have
      $$\max_\theta \log p_\theta(X) = \max_\theta \left\{ \int q(Z) \log \frac{p_\theta(X, Z)}{q(Z)} dZ + D_{\mathrm{kl}}\left(Q(Z) \| P_\theta(Z \mid X)\right) \right\} \tag{1}$$
      iv. Since $q(Z)$ can be ARTIFICIALLY selected, what will happen if we let $q(Z) = p_\theta(Z \mid X)$?
      v. This means that we can just focus on the first term. Now the E-step is
      $$q_{\text{new}}(Z) = p_{\theta_{\text{old}}}(Z \mid X)$$
      Can you write down the M-step based on our observation above?
   (c) Now we consider i.i.d. observations $(X_1, Z_1), \ldots, (X_n, Z_n)$.
      i. Write down the target in the form of equation (1)
      ii. Write down the expression for E step and M step.

3. EM for Mixture Normal
   (a) This is the general setting: Let Z be the latent variable which is responsible for generating X. For mixture normal, Z follows the multinomial distribution, $P_\theta(Z = \ell) = \pi_\ell$, where $\ell = 1, \ldots, K$. Given $Z = \ell, X$ is generated from the ℓ th normal distribution $N(\mu_\ell, \Sigma_\ell)$, which means that $P_\theta(X \mid Z = \ell) = f_{(\mu_\ell, \Sigma_\ell)}(X)$.
   But do notice that this is how we "deliberately" generate X. What if we just observed $(X_1, \cdots, X_n)$? (Consider Figure 11.11 on page 353 WITHOUT colors, how will you separate different color groups? That's one of the things what we need to do.)
   (b) E-Step for Mixture Normal
      i. Write down the joint density function $p_\theta(X, Z = \ell)$ and $p_\theta(X, Z)$
      ii. Write down the density function of X
      iii. Find $p_\theta(Z \mid X)$ and $p_\theta(Z_i = j \mid X_i)$
      iv. Recall the E-Step of EM as
      $$q_{\text{new}}(Z_i = j) = p_{\theta_{\text{old}}}(Z_i = j \mid X_i), \quad j = 1, \ldots, K$$
      Let $q_{ij} = p_\theta(Z_i = j \mid X_i)$, and $\theta^{(t)} = \left\{ \left(\pi_\ell^{(t)}, \mu_\ell^{(t)}, \Sigma_\ell^{(t)} \right), \ell = 1, \ldots, K \right\}$ as the solution of $t-$th iteration step. Write down how we should update q_{ij} (denoted as $q_{ij}^{(t)}$).

(c) M-Step for Mixture Normal. Now $q_{ij}^{(t)}$ is treated as a constant. The M-Step needs to update (π_j, μ_j, Σ_j) for $j = 1, \ldots, K$

i. Show that

$$\max_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{K} \left\{ q_{ij}^{(t)} \log \frac{\pi_j f_{(\mu_j, \Sigma_j)}(X_i)}{q_{ij}^{(t)}} \right\} = \max_{\mu_j, \Sigma_j} \sum_{i=1}^{n} \sum_{j=1}^{K} \left\{ q_{ij}^{(t)} \log f_{(\mu_j, \Sigma_j)}(X_i) \right\} + \max_{\pi_j} \sum_{i=1}^{n} \sum_{j=1}^{K} q_{ij}^{(t)} \log \pi_j$$

ii. Based on the above observations, we have

$$\max_{\mu_j, \Sigma_j} \sum_{i=1}^{n} \sum_{j=1}^{K} \left\{ q_{ij}^{(t)} \log f_{(\mu_j, \Sigma_j)}(X_i) \right\} = \max_{\mu_1, \Sigma_1} \sum_{i=1}^{n} q_{i1}^{(t)} \log f_{(\mu_1, \Sigma_1)}(X_i) + \ldots + \max_{\mu_K, \Sigma_K} \sum_{i=1}^{n} q_{iK}^{(t)} \log f_{(\mu_K, \Sigma_K)}(X_i)$$

So, for the update of (μ_j, Σ_j), we only need to consider

$$\max_{\mu_j, \Sigma_j} \sum_{i=1}^{n} q_{ij}^{(t)} \log f_{(\mu_j, \Sigma_j)}(X_i)$$

iii. If $q_{ij}^{(t)} = 1$ for $i = 1, \ldots, n$, compute

$$\mu_j^{(t+1)}, \quad \Sigma_j^{(t+1)}$$

Hint: using MLE.

iv. For $0 \leq q_{ij}^{(t)} \leq 1$, Write down the log-likelihood equations and Find the solution to your log-likelihood equations.

v. For the update of π_j, use the Lagrange multiplier method to find the solution. Hint: consider

$$\max_{\pi_j} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{K} q_{ij}^{(t)} \log \pi_j + \lambda \left(\sum_{j=1}^{K} \pi_j - 1 \right) \right\}$$