| | | |
|---|---|---|
| **Quiz 8** | **Name:** | ____________________ |
| **Week 13, Dec/2/2020** | **On your left:** | ____________________ |
| **CS 280: Fall 2020** | **On your right:** | ____________________ |
| **Instructor: Xuming He, Lan Xu** | | |

**Instructions:**
Please answer the questions below. Show all your work. This is an open-book test. NO discussion/collaboration is allowed.

**Problem 1.** (10 points) **EM and GMM**. Suppose we have a data set $\{x_1, x_2, ...x_n\}$, how can we fit the data set with Gaussian Mixture Model by Expectation Maximization method? What is the difference between this method and k means clustering? (You need to write down the detail of derivation of the first question here, and it takes account for most of scores of Probelm 1, You can find information from Lecture16 Page 8-25.)

**Problem 2.** (10 points) **ELBO** The EM algorithm is to find maximum likelihood solution for probablistic model having latent variables. First, give the lower bound of the objective function.(You should describle the definition of the variables you use). Then, explain and visualize E-step and M-step and why the EM algorithm finally converge.

**Solution:**

1. (3') We denote all of the observed variables by $\mathbf{X}$, all the hidden variables by $\mathbf{Z}$ and the model parameters by $\theta$. We introduce a distribution $q(\mathbf{Z})$ defined over $\mathbf{Z}$.

   Our goal is to maximize the objective function $p(\mathbf{X}|\theta)$. For any $\mathbf{Z}$, there is

$$\begin{aligned}\log p(\mathbf{X}|\theta) &= \int_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{X},\mathbf{Z}|\theta)}{q(\mathbf{Z})} - \int_{\mathbf{Z}} q(\mathbf{Z}) \log \frac{p(\mathbf{Z}|\mathbf{X},\theta)}{q(\mathbf{Z})} \\ &= \mathcal{L}(q,\theta) + KL(q||p)\end{aligned}$$

   as the $KL(q||p) \geq 0$ iff $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X},\theta)$, the $\mathcal{L}(q,\theta)$ is the lower bound on $\log p(\mathbf{X}|\theta)$.

2. Now we maximized the lower bound $\mathcal{L}(q,\theta)$.

   E-step: (2' for explanation, 1' for illustration) Suppose the current value of the parameter is $\theta^{old}$, we fix it and maximize lower bound $\mathcal{L}(\mathbf{X},\theta^{old})$ with respect to $q(\mathbf{Z})$. As the value of $\log p(\mathbf{X}|\theta^{old})$ does not depend on $q(\mathbf{Z})$, so $\mathcal{L}(\mathbf{X},\theta^{old})$ will be maximized when $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X},\theta^{old})$. After the E step, the lower bound becomes:

$$\begin{aligned}\mathcal{L}(q,\theta) = &\int_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\theta^{old}) \log p(\mathbf{X},\mathbf{Z}|\theta) - \int_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\theta^{old}) \log p(\mathbf{Z}|\mathbf{X},\theta^{old}) \\ =&Q(\theta,\theta^{old}) + \text{constant}\end{aligned}$$

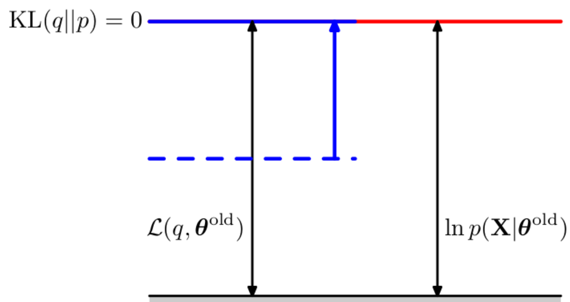


Figure 1: Illustration of E-step

   M-step: (2' for explanation, 1' for illustration) We fix the distribution $q(\mathbf{Z})$ and maximize $\mathcal{L}(q,\theta)$ with respect to $\theta$. This equals to maximized the expectation of complete data log likelihood $Q(\theta,\theta^{old})$. Assume $\theta^{new} = \arg\max_\theta \mathcal{L}(q,\theta)$, as $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X},\theta^{old}) \neq p(\mathbf{Z}|\mathbf{X},\theta^{new})$, there is $KL(q||p) > 0$.

   As in each E-step and M-step, EM algorithm increases the lower bound of the log likelihood function, the algorithm will finally converge.(1')

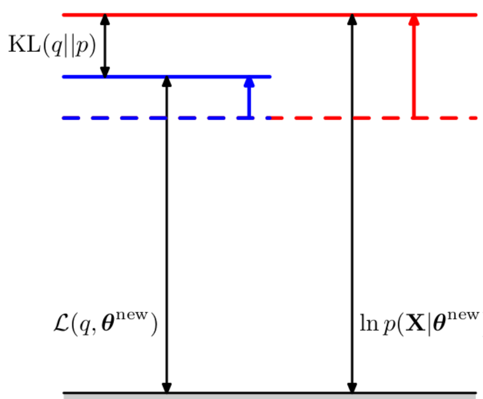


Figure 2: Illustration of M-step