Homework 4

Yunian Pan

November 10, 2018

1 Problem 1: EM

1.1 E step:

$$\begin{aligned} Q_i(z_i) &= p(z_i|x_i;\theta) \\ &= \frac{p(z_i, x_i|\theta)}{\sum_{z_i} p(z_i, x_i|\theta)} \\ &= \frac{\pi_{z_i} \prod_{j=1}^{M} \mu_{z_i}(j)^{x_i(j)}}{\sum_{z_i}^{k} \pi_{z_i} \prod_{j=1}^{M} \mu_{z_i}(j)^{x_i(j)}} \end{aligned}$$

1.2 M step:

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i;\theta)}{Q_i(z_i)} \\ &= \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log p(x_i, z_i;\theta) - \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log Q_i(z_i) \\ &= Q(\theta) - const \end{aligned}$$

$$\begin{aligned} \mathcal{Q}(\theta) &= \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log p(x_i, z_i;\theta) \\ &= \sum_{i=1}^{N} \sum_{z_i} \tau_{i,z_i} (\log \pi_{z_i} + \log \prod_{j=1}^{M} \mu_{z_i}(j)^{x_i(j)}) \\ &= \sum_{i=1}^{N} \sum_{z_i} \tau_{i,z_i} (\log \pi_{z_i} + x_i(q) \log \mu_{z_i}(q)) \qquad (x_i(q) = 1) \end{aligned}$$

Where we write $Q_i(z_i)$ as τ_{i,z_i}.

There are 2 constraints: $\sum_j^M \mu_k(j) = 1$, $\sum_i^K \pi_i = 1$, with an observation $\sum_j^M x_i(j) = 1\ \forall x_i \in \{x_1, \ldots, x_N\}$.

Use lagrange multiplier to find the optimum value of π_{z_i} and μ_{z_i} as below:

$$L(\mu, \pi, \lambda_1, \lambda_2) = \mathcal{Q}(\pi, \mu) - \lambda_1(\sum_j^M \mu_k(j) - 1) - \lambda_2(\sum_i^K \pi_i - 1)$$

$$\frac{\partial L}{\partial \pi_k} = \sum_{i=1}^N \tau_{i,k}\frac{1}{\pi_k} - \lambda_2 = 0$$

$$sum \ \Rightarrow \ 1 = \sum_{k=1}^K \pi_k = \frac{1}{\lambda_2}\sum_{k=1}^K\sum_{i=1}^N \tau_{i,k}$$

$$\Rightarrow \lambda_2 = \sum_{k=1}^K\sum_{i=1}^N \tau_{i,k}$$

$$plug\ in \Rightarrow \pi_k = \frac{\sum_{i=1}^N \tau_{i,k}}{\sum_{k=1}^K\sum_{i=1}^N \tau_{i,k}}$$

$$\frac{\partial L}{\partial \mu_k(q)} = \sum_{i=1}^N \tau_{i,k} x_i(q)\frac{1}{\mu_k(q)} - \lambda_1 = 0$$

$$\Rightarrow \mu_k(q) = \sum_{i=1}^N \tau_{i,k} x_i(q)\frac{1}{\lambda_1}$$

$$sum \ \Rightarrow \ 1 = \sum_{q=1}^M \mu_k(q) = \sum_{i=1}^N \tau_{i,k}\sum_{q=1}^M x_i(q)\frac{1}{\lambda_1}$$

$$\Rightarrow \lambda_1 = \sum_{i=1}^N \tau_{i,k}$$

$$plug\ in \Rightarrow \mu_k(q) = \sum_{i=1}^N \frac{\tau_{i,k} x_i(q)}{\sum_{i=1}^N \tau_{i,k}}$$

To conclude, the probability for class $z = k$ is $\pi_k = \frac{\sum_{i=1}^N \tau_{i,k}}{\sum_{k=1}^K\sum_{i=1}^N \tau_{i,k}}$, the probability of x_i which belongs to the class $z = k$ taking on the q^{th} value is $\mu_k(q) = \frac{\sum_{i=1}^N \tau_{i,k} x_i(q)}{\sum_{i=1}^N \tau_{i,k}}$.

2 Problem 2: clustering

Proof:

$$\begin{aligned}
\phi(W_1) &= \sum_{x_i \in W_1} \|x_i - \mu_{i,1}\|^2 \\
\phi(W_2) &= \sum_{x_i \in W_2} \|x_i - \mu_{i,2}\|^2 \\
\phi(W_1 \cup W_2) &= \sum_{x_i \in (W_1 \cup W_2)} \|x_i - \mu_{i,3}\|^2
\end{aligned}$$

Where $\mu_{i,k}$ is one of the centers for dataset W_k, $W_3 = W_1 + W_2$, consider the iterations of K-means method, $\mu_{i,1} = \min_{\mu_i} \sum_{x_i \in W_1} \|x_i - \mu_i\|^2$, $\mu_{i,2} = \min_{\mu_i} \sum_{x_i \in W_2} \|x_i - \mu_i\|^2$, Therefore, we have:

$$\begin{aligned}
\sum_{x_i \in W_1} \|x_i - \mu_{i,1}\|^2 + \sum_{x_i \in W_2} \|x_i - \mu_{i,2}\|^2 &\leqslant \sum_{x_i \in W_1} \|x_i - \mu_{i,3}\|^2 + \sum_{x_i \in W_2} \|x_i - \mu_{i,3}\|^2 \\
&= \sum_{x_i \in (W_1 \cup W_2)} \|x_i - \mu_{i,3}\|^2
\end{aligned}$$

$$\Leftrightarrow \phi(W_1) + \phi(W_2) \leqslant \phi(W_1 \cup W_2)$$

3 Problem 3: MLE and MAP

3.1 a) MLE

$$\begin{aligned}
l(\mu) &= \log \prod_{i=1}^{N} \mathcal{N}(x_i | \mu, \sigma^2) \\
&= \sum_{i=1}^{N} (-\frac{1}{2\sigma^2}(x_i - \mu)^2 + C_1) \\
\frac{\partial l}{\partial \mu} &= \sum_{i=1}^{N} \frac{1}{\sigma^2}(x_i - \mu) = 0 \\
\Rightarrow \mu &= \frac{\sum_{i=1}^{N} x_i}{N}
\end{aligned}$$

3.2 b) MAP

$$
\begin{aligned}
l(\mu) &= \log \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2) + \log \mathcal{N}(\mu|\nu, \beta^2) \\
&= \sum_{i=1}^{N} (-\frac{1}{2\sigma^2}(x_i - \mu)^2 + C_1) + (-\frac{1}{2\beta^2}(\mu - \nu)^2 + C_2) \\
\frac{\partial l}{\partial \mu} &= \sum_{i=1}^{N} \frac{1}{\sigma^2}(x_i - \mu) - \frac{1}{\beta^2}(\mu - \nu) = 0 \\
\Rightarrow \mu &= \frac{\frac{\sum_{i=1}^{N} x_i}{\sigma^2} + \frac{1}{\beta^2}}{\frac{N}{\sigma^2} + \frac{1}{\beta^2}} \\
\Rightarrow \mu &= \frac{\sum_{i=1}^{N} x_i \beta^2 + \sigma^2}{N\beta^2 + \sigma^2}
\end{aligned}
$$

3.3 c)

When $N \to +\infty$, $\frac{\sum_{i=1}^{N} x_i \beta^2 + \sigma^2}{N\beta^2 + \sigma^2} \to \frac{\sum_{i=1}^{N} x_i \beta^2}{N\beta^2} = \frac{\sum_{i=1}^{N} x_i}{N}$ Therefore, there's no difference between the μ's we derived from a) and b), to sum up, when samples N goes to infinity, the likelihood we observe will not be affected by the hidden prior probability of the parameters.