CSCI 567 Machine Learning

Homework #4

Name : Yi Zhao

Question 1.1 Answer :

$$P(x_1, x_2, ..., x_N) = \begin{cases} \frac{1}{\theta^N}, & \forall x_1, x_2, ..., x_N \in (0, \theta] \\ 0, & \exists x_1, x_2, ..., x_N \notin (0, \theta]. \end{cases}$$

$$\theta^* = \max\{x_1, x_2, ..., x_N\}$$

because $\frac{\mathrm{d}\ \ln P(\theta|\boldsymbol{x})}{\mathrm{d}\theta} = -\frac{N}{\theta}$, $P(\theta|\boldsymbol{x})$ is a decreasing function, and the minimal number for θ make $P(\theta|\boldsymbol{x}) \neq 0$ is $\max\{x_1, x_2, ..., x_N\}$.

Question 1.2 Answer :

$$P(k|x_n, \theta_1, \theta_2, \omega_1, \omega_2) = \frac{P(k)P(x_n|k)}{P(x_n, \theta_1, \theta_2, \omega_1, \omega_2)} = \frac{\omega_k U(X = x_n|\theta_k)}{\omega_1 U(X = x_n|\theta_1) + \omega_2 U(X = x_n|\theta_2)}$$

$$P(k|x_n, \theta_1, \theta_2, \omega_1, \omega_2) = \frac{\omega_k \frac{1}{\theta_k}\mathbf{1}[0 < x_n \leq \theta_k]}{\omega_1 \frac{1}{\theta_1}\mathbf{1}[0 < x_n \leq \theta_1] + \omega_2 \frac{1}{\theta_2}\mathbf{1}[0 < x_n \leq \theta_2]}$$

$$Q_q(\boldsymbol{\theta}) = \sum_n \sum_{\boldsymbol{z}_n} q(\boldsymbol{z}_n) \mathrm{log} p(x_n, \boldsymbol{z}_n|\boldsymbol{\theta})$$

$$Q_q(\boldsymbol{\theta}) = \sum_{n=1}^{N} [P(1|x_n, \boldsymbol{\theta}) \mathrm{log} p(x_n, \boldsymbol{z}_n = 1|\boldsymbol{\theta}) + P(2|x_n, \boldsymbol{\theta}) \mathrm{log} p(x_n, \boldsymbol{z}_n = 2|\boldsymbol{\theta})]$$

$$Q_q(\boldsymbol{\theta}) = \sum_{n=1}^{N} \frac{\omega_1 U(X = x_n|\theta_1) \mathrm{log} p(x_n, \boldsymbol{z}_n = 1|\boldsymbol{\theta}) + \omega_2 U(X = x_n|\theta_2) \mathrm{log} p(x_n, \boldsymbol{z}_n = 2|\boldsymbol{\theta})}{\omega_1 U(X = x_n|\theta_1) + \omega_2 U(X = x_n|\theta_2)}$$

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{OLD}}) = \sum_{n=1}^{N} \frac{\omega_1^{\mathrm{OLD}} U(X = x_n|\theta_1^{\mathrm{OLD}}) \mathrm{log}[\omega_1 U(X = x_n|\theta_1)] + \omega_2^{\mathrm{OLD}} U(X = x_n|\theta_2^{\mathrm{OLD}}) \mathrm{log}[\omega_2 U(X = x_n|\theta_2)]}{\omega_1^{\mathrm{OLD}} U(X = x_n|\theta_1^{\mathrm{OLD}}) + \omega_2^{\mathrm{OLD}} U(X = x_n|\theta_2^{\mathrm{OLD}})}$$

In order to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{OLD}})$, then we should maximize $\sum \mathrm{log}[\omega_1 U(X = x_n|\theta_1)]$ and $\sum \mathrm{log}[\omega_2 U(X = x_n|\theta_2)]$. If $\theta_1 < \max\{x_i \mid \forall\ x_i \leq \theta_1^{\mathrm{OLD}}\}$ or $\theta_2 < \max\{x_1, x_2, ..., x_N\}$, we can find that $\exists i$, makes $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{OLD}}) = -\infty$. For $U(X = x_n|\theta_1)$ and $U(X = x_n|\theta_2)$, arccording to problem 1.1, we can get

$$\theta_1^* = \max\{x_i \mid \forall\ x_i \leq \theta_1^{\mathrm{OLD}}\}$$

$$\theta_2^* = \max\{x_1, x_2, ..., x_N\}$$

Question 2.1 Answer :

$$\text{Because } P(\boldsymbol{x}_a, \boldsymbol{x}_b) = \sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_a, \boldsymbol{x}_b|k)$$

$$\frac{P(\boldsymbol{x}_a, \boldsymbol{x}_b)}{P(\boldsymbol{x}_a)} = \sum_{k=1}^{K} \pi_k \frac{P(\boldsymbol{x}_a, \boldsymbol{x}_b|k)}{P(\boldsymbol{x}_a)}$$

$$P(\boldsymbol{x}_b|\boldsymbol{x}_a) = \sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_b|\boldsymbol{x}_a, k) \frac{P(\boldsymbol{x}_a, \boldsymbol{x}_b|k)}{P(\boldsymbol{x}_a)P(\boldsymbol{x}_b|\boldsymbol{x}_a, k)}$$

$$P(\boldsymbol{x}_b|\boldsymbol{x}_a) = \sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_b|\boldsymbol{x}_a, k) \frac{P(\boldsymbol{x}_a|k)}{P(\boldsymbol{x}_a)}$$

$$P(\boldsymbol{x}_b|\boldsymbol{x}_a) = \sum_{k=1}^{K} (\pi_k \frac{P(\boldsymbol{x}_a|k)}{P(\boldsymbol{x}_a)}) P(\boldsymbol{x}_b|\boldsymbol{x}_a, k)$$

$$P(\boldsymbol{x}_b|\boldsymbol{x}_a) = \sum_{k=1}^{K} (\frac{\pi_k P(\boldsymbol{x}_a|k)}{\sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_a|k)}) P(\boldsymbol{x}_b|\boldsymbol{x}_a, k)$$

$$\text{So, } \lambda_k = \frac{\pi_k P(\boldsymbol{x}_a|k)}{\sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_a|k)}$$

$$\text{So, } \sum_{k=1}^{K} \lambda_k = \frac{\sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_a|k)}{\sum_{k=1}^{K} \pi_k P(\boldsymbol{x}_a|k)} = 1$$

Question 3.1 Answer :

$$\lim_{\sigma \to 0} \gamma(z_{nk}) = \lim_{\sigma \to 0} \frac{1}{1 + \sum_{j \neq k} (\pi_j/\pi_k) \text{exp}[(||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2)/2\sigma^2]}$$

$$\lim_{\sigma \to 0} \gamma(z_{nk}) = \frac{1}{1 + \sum_{j \neq k} \lim_{\sigma \to 0} (\pi_j/\pi_k) \text{exp}[(||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2)/2\sigma^2]}$$

$$\text{If } k \neq \text{argmin}_{k'} ||\boldsymbol{x}_n - \boldsymbol{\mu}_{k'}||^2, \text{ then } \exists j \neq k \text{ makes } ||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2 > 0,$$

$$\text{then } \lim_{\sigma \to 0} \text{exp}[(||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2)/2\sigma^2] = +\infty, \ \lim_{\sigma \to 0} \gamma(z_{nk}) = \lim_{\delta \to +\infty} \frac{1}{\text{C} + \delta} = 0.$$

$$\text{If } k = \text{argmin}_{k'} ||\boldsymbol{x}_n - \boldsymbol{\mu}_{k'}||^2, \text{ then } \forall j \neq k \text{ makes } ||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2 < 0,$$

$$\text{then } \lim_{\sigma \to 0} \text{exp}[(||\boldsymbol{x}_n - \boldsymbol{\mu}_k||^2 - ||\boldsymbol{x}_n - \boldsymbol{\mu}_j||^2)/2\sigma^2] = 0, \ \lim_{\sigma \to 0} \gamma(z_{nk}) = \lim_{\delta \to 0} \frac{1}{1 + \delta} = 1.$$

$$\text{So, we can verify } \lim_{\sigma \to 0} \gamma(z_{nk}) = r_{nk}$$

$$\lim_{\sigma \to 0} \sum_{n}^{N} \sum_{k}^{K} \gamma(z_{nk}) \text{log} p(\boldsymbol{x}_n, z_n = k) = \lim_{\sigma \to 0} \sum_{n}^{N} \sum_{k}^{K} r_{nk} [\text{log}\pi_k - \frac{||\boldsymbol{x}_n - \boldsymbol{\mu}_k||_2^2}{2\sigma^2} - \text{log}\sqrt{2\pi}\sigma]$$

$$= \lim_{\sigma \to 0} \sum_{n}^{N} \sum_{k}^{K} r_{nk} (-\frac{||\boldsymbol{x}_n - \boldsymbol{\mu}_k||_2^2}{2\sigma^2}) = -\frac{1}{2\sigma^2} \lim_{\sigma \to 0} \sum_{n}^{N} \sum_{k}^{K} r_{nk} ||\boldsymbol{x}_n - \boldsymbol{\mu}_k||_2^2$$

$$\text{So,} \sum_{n}^{N} \sum_{k}^{K} \gamma(z_{nk}) \text{log} p(\boldsymbol{x}_n, z_n = k) = -\frac{1}{2\sigma^2} J, \text{ as } \sigma \to 0.$$

Question 4.1 Answer :

$$L = \log \prod_{n=1}^{N} \mathrm{P}(y_n)\mathrm{P}(x_n|y_n) = \log \prod_{n=1}^{N} [\mathrm{P}(y_n = c) \prod_{d=1}^{D} \mathrm{P}(X_{nd} = x_{nd}|y_n = c)]$$

$$L = \log \prod_{n=1}^{N} \mathrm{P}(y_n = c) \prod_{n=1}^{N} \prod_{d=1}^{D} \mathrm{P}(X_{nd} = x_{nd}|y_n = c)$$

$$L = \sum_{n=1}^{N} \log\pi_{y_n} + \sum_{n=1}^{N} \sum_{d=1}^{D} \log\mathrm{P}(X_{nd} = x_{nd}|y_n = c)$$

$$L = \sum_{n=1}^{N} \log\pi_{y_n} + \sum_{n=1}^{N} \sum_{d=1}^{D} (-\frac{(x_{nd} - \mu_{cd})^2}{2\sigma_{cd}^2} - \log\sqrt{2\pi}\sigma_{cd})$$

$$L = \sum_{n=1}^{N} \log\pi_{y_n} + \sum_{c} \sum_{n:y_n=c} \sum_{d=1}^{D} (-\frac{(x_{nd} - \mu_{cd})^2}{2\sigma_{cd}^2} - \log\sqrt{2\pi}\sigma_{cd})$$

$$L = \sum_{c} \log\pi_c \times (\text{number of data points labeled as } c) + \sum_{c} \sum_{n:y_n=c} \sum_{d=1}^{D} (-\frac{(x_{nd} - \mu_{cd})^2}{2\sigma_{cd}^2} - \log\sqrt{2\pi\sigma_{cd}^2})$$

Question 4.2 Answer :

Make $\#_c$ represents number of data points labeled as c in the training points

$$\pi_c^* = \frac{\#_c}{N}$$

$$\frac{\partial L}{\partial \mu_{cd}} = 0, \text{then } \mu_{cd}^* = \frac{\sum_{n:y_n=c} x_{nd}}{\#_c}$$

$$\frac{\partial L}{\partial \sigma_{cd}^2} = 0, \text{then } \sigma_{cd}^{2*} = \frac{\sum_{n:y_n=c} (x_{nd} - \mu_{cd})^2}{\#_c}$$

$$\sigma_{cd}^{2*} = \frac{1}{\#_c} \sum_{n:y_n=c} x_{nd}^2 - \frac{1}{\#_c^2} \sum_{i:y_i=c} \sum_{j:y_j=c} x_{id}x_{jd}$$