CSCI 567 - Home Work - 4

Sai Swarna Latha Vempaty - 5824899937

1.1)

Likelihood $\mathcal{L}(\theta|x) = \begin{cases} \frac{1}{\theta^N} & \text{if } 0 < x_i \leq \theta \; \forall \; i = 1, 2, ..., N \\ 0 & otherwise \end{cases}$

MLE(θ) = $\hat{\theta}$ = max$\{x_1, x_2, ...x_N\}$

$\frac{1}{\theta^N}$ is decreasing function, lower the θ, maximum the likelihood.But, If θ=min$\{x_1, x_2, ...x_N\}$,second condition kicks in, and Likelihood becomes 0.So, the minimum value of x_n, that maximizes likelihood of θ is max$\{x_1, x_2, ...x_N\}$.

1.2)

$$\mathrm{P}(k|x_n, \theta_1, \theta_2, \omega_1, \omega_2) = \frac{\mathrm{P}(x_n|k,\theta_1,\theta_2,\omega_1,\omega_2).\mathrm{P}(k)}{\mathrm{P}(x_n|k{=}1,\theta_1,\theta_2,\omega_1,\omega_2).\mathrm{P}(k{=}1)+\mathrm{P}(x_n|k{=}2,\theta_1,\theta_2,\omega_1,\omega_2).\mathrm{P}(k{=}2)}$$

$$= \frac{\omega_k.U(X{=}x_n|\theta_k)}{\omega_1.U(X{=}x_n|\theta_1)+\omega_2.U(X{=}x_n|\theta_2)} \qquad = \frac{\omega_k\cdot\frac{1}{\theta_k}.\mathbf{1}[0<x_n\leq\theta_k]}{\omega_1\cdot\frac{1}{\theta_1}.\mathbf{1}[0<x_n\leq\theta_1]+\omega_2\cdot\frac{1}{\theta_2}.\mathbf{1}[0<x_n\leq\theta_2]}$$

Expected complete-data log-likelihood:

Let $\theta = \theta_1, \theta_2, \omega_1, \omega_2$

$Q_q(\theta, \theta^{OLD}) = \sum_n \sum_k \mathrm{P}(k|x_n, \theta^{OLD}).log\mathrm{P}(x_n, k|\theta)$

Let $\mathrm{P}(k|x_n, \theta^{OLD}) = \mathrm{P}_{OLD}(k|x_n) = \frac{\omega_k^{OLD}\cdot\frac{1}{\theta_k^{OLD}}.\mathbf{1}[0<x_n\leq\theta_k^{OLD}]}{\omega_1^{OLD}\cdot\frac{1}{\theta_1^{OLD}}.\mathbf{1}[0<x_n\leq\theta_1^{OLD}]+\omega_2^{OLD}\cdot\frac{1}{\theta_2^{OLD}}.\mathbf{1}[0<x_n\leq\theta_2^{OLD}]}$

$Q_q(\theta, \theta^{OLD}) = \sum_n \sum_k \mathrm{P}_{OLD}(k|x_n).log(\mathrm{P}(k).\mathrm{P}(x_n|k, \theta))$

Substituting $\mathrm{P}(k) = \omega_k$, $\mathrm{P}(x_n|k, \theta) = U(X = x_n, \theta_k)$

$Q_q(\theta, \theta^{OLD}) = \sum_n \sum_k \mathrm{P}_{OLD}(k|x_n).[log(\omega_k) + logU(X = x_n, \theta_k)]$

$Q_q(\theta, \theta^{OLD}) = \sum_n \sum_k \mathrm{P}_{OLD}(k|x_n).[log(\omega_k) + log(\frac{1}{\theta_k}.\mathbf{1}[0 < x_n \leq \theta_k])]$

M-Step:

θ = arg max $Q_q(\theta, \theta^{OLD})$

In the question it is mentioned that, $\theta_1 \geq min\{x_1, x_2, ...x_N\}$,so for the values of x, less than θ_1^{OLD}, $\mathrm{P}_{OLD}(k = 1|x_n) = 0$,(As 0*(-$\infty$) = 0).

So, while optimizing $\theta_2 = max\{x_1, x_2, ...x_N\}$,where $0 < \{x_1, x_2, ..x_N\} \leq \theta_2^{OLD}$.

But, in case of $\theta_1 = max\{x_1, x_2, ...x_i\}$,, where $0 < \{x_1, x_2, ..x_i\} \leq \theta_1^{OLD}$

2.1)

$\mathrm{P}(x_b|x_a) = \frac{\mathrm{P}(x_b,x_a)}{\mathrm{P}(x_a)}$
$= \frac{\mathrm{P}(x)}{\mathrm{P}(x_a)}$
$= \sum_k \frac{\pi_k.\mathrm{P}(x|k)}{\mathrm{P}(x_a)}$
$= \sum_k \frac{\pi_k}{\mathrm{P}(x_a)}.\mathrm{P}(x_b, x_a|k)$
$= \sum_k \frac{\pi_k}{\mathrm{P}(x_a)}.\mathrm{P}(x_b|x_a, k).\mathrm{P}(x_a|k)$

by comparing question with the above equation,

$\lambda_k = \frac{\mathrm{P}(x_a|k).\pi_k}{\mathrm{P}(x_a)}$

but, $\mathrm{P}(x_a) = \sum_{j=1}^{K} \mathrm{p}(k=j).\mathrm{p}(x_a|k=j)$

$\lambda_k = \frac{\mathrm{P}(x_a|k).\pi_k}{\sum_{j=1}^{K} \pi_j.\mathrm{P}(x_a|k=j)}$

3.1)

$\gamma(z_{nk})$ when $\sigma- > 0, jth$ term for which the responsibility is highest i.e., smallest $||x_n - \mu_j||_2^2$ goes to zero slowly in the denominator. So,$\gamma(z_{nk}) = 1$ for that j term, and all others it is 0.

$=> \gamma(z_{nk}) = \begin{cases} 1 & \text{if k= arg } min_j||x_n - \mu_j||_2^2 \\ 0 & otherwise \end{cases}$

So, as limit $\sigma- > 0$, $\gamma(z_{nk}) = r_{nk} - - - - - - - -$ (1)

$\Sigma = \sigma^2.I =>$ same variance for all classes,

Therefore, $\mathcal{N}(x_n|\mu_k, \sigma^2 I) = constant.exp\{\frac{-1}{2\sigma^2}.||x_n - \mu_k||_2^2\} - - - - - - - - - - - - -$ (2)

Substituting (1) and (2) in GMM expected log likelihood equation,

GMM$=\frac{-1}{2}.\sum_n^N \sum_k^K r_{nk}.||x_n - \mu_k||_2^2 + constant$

$J = \sum_n^N \sum_k^K r_{nk}.||x_n - \mu_k||_2^2$

As, GMM is -ve of J, we can say that under limit $\sigma- > 0$, maximizing GMM is equal to minimizing to J.

4.1)

$\mathcal{L} = \sum_n^N log\mathrm{P}(x_n, y_n)$
$= \sum_n^N log\mathrm{P}(y_n).\mathrm{P}(x_n|y_n)$
$= \sum_c^C \sum_{n:y_n=c}^N log\mathrm{P}(y_n).\mathrm{P}(x_n|y_n)$

re-arranging the equation, and introducing $\gamma_{nc} = \begin{cases} 1 & \text{if } x_n \in c \\ 0 & otherwise \end{cases}$

$= \sum_{c=1}^C \sum_{n=1}^N \gamma_{nc}.log(\mathrm{P}(y_n).\mathrm{P}(x_n|y_n))$
$= \sum_{c=1}^C \sum_{n=1}^N \gamma_{nc}[log(\mathrm{P}(Y=c)) + \sum_{d=1}^D log(\mathrm{P}(X_d = x_{nd}|Y=c))]$
$= \sum_{c=1}^C \sum_{n=1}^N \gamma_{nc}\ log(\pi_c) + \sum_{c=1}^C \sum_{n=1}^N \sum_{d=1}^D \gamma_{nc}\ log(\frac{1}{\sqrt{2.\pi.\sigma_{cd}^2}}.exp\{\frac{-(x_{nd}-\mu_{cd})^2}{2.\sigma_{cd}^2}\})]$

$\mathcal{L} = \sum_{c=1}^{C}\sum_{n=1}^{N}\gamma_{nc}\ log(\pi_c) + \sum_{c=1}^{C}\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma_{nc}[-\frac{1}{2}.log(2.\pi.\sigma_{cd}^2) - \frac{1}{2}\frac{(x_{nd}-\mu_{cd})^2}{\sigma_{cd}^2}]$

4.2)

$(\mu_{cd}^*, {\sigma_{cd}^2}^*, \pi_c^*) = \text{arg max}\{\sum_{c=1}^{C}\sum_{n=1}^{N}\gamma_{nc}\ log(\pi_c) + \sum_{c=1}^{C}\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma_{nc}[-\frac{1}{2}.log(2.\pi.\sigma_{cd}^2) - \frac{1}{2}\frac{(x_{nd}-\mu_{cd})^2}{\sigma_{cd}^2}]\}$

maximizing $\mathcal{L}$ w.r.t μ_{cd}

$\frac{\partial\mathcal{L}}{\partial\mu_{cd}} = \sum_n \gamma_{nc}.\frac{(x_{nd}-\mu_{cd})}{\sigma_{cd}^2}.(-\mu_{cd}) = 0$

$\mu_{cd}^* = \frac{\sum_n \gamma_{nc}.x_{nd}}{\sum_n \gamma_{nc}}$

i.e., μ_{cd}^* =(Sum of all x_nd in class c)/(no.of points in class c)

maximizing $\mathcal{L}$ w.r.t σ_{cd}

$\frac{\partial\mathcal{L}}{\partial\sigma_{cd}} = \sum_n \gamma_{nc}[\frac{-2}{\sigma_{cd}} + \frac{2*(x_{nd}-\mu_{cd})^2}{\sigma_{cd}^2}] = 0$

${\sigma_{cd}^2}^* = \frac{\sum_n \gamma_{nc}.(x_{nd}-\mu_{cd})^2}{\sum_n \gamma_{nc}}$

maximizing $\mathcal{L}$ w.r.t π_c,

term related to π_c is $\sum_{c=1}^{C}\sum_{n=1}^{N}\gamma_{nc}.log(\pi_c)$

using Lagrangian, and $\sum_c \pi_c = 1$, Likelihood becomes,

$\sum_{c=1}^{C}\sum_{n=1}^{N}\gamma_{nc}.log(\pi_c) + \lambda(\sum_c \pi_c - 1)$

$\frac{\partial\mathcal{L}}{\partial\pi_c} = \sum_{c=1}^{C}\sum_{n=1}^{N}\gamma_{nc}.\frac{1}{\pi_c} + \lambda = 0$

$\pi_c = \frac{-1}{\lambda}.\sum_{n=1}^{N}\gamma_{nc}$

substituting in $\sum_c \pi_c = 1$, we get, $\lambda = -N$

$=> \pi_c^* = \frac{\sum_{n=1}^{N}\gamma_{nc}}{N}$