آرین شکر - ۴۰۰۲۳۴۹۴ تکلیف سوم شناسایی آماری الگو

Exercise 1: we know that $p(X,Z|\theta) = \prod_{n=1}^{N}\prod_{k=1}^{K}(\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k))^{z_{nk}}$

Also, using chain rule we have: $p(x,z|\theta) = P(z|x,\theta)\cdot p(x)$

$\Rightarrow p(z|X,\theta) = \dfrac{P(x,Z|\theta)}{P(x)}$. However the $P(X,Z|\theta)$ that we calculated above is a joint distribution. What we want is $P(X_n, Z_k=1|\theta)$.

Because the samples were drawn i.i.d, the joint distribution can be decomposed into non-conditional marginal multiples. Therefore:

$$P(X_n, Z_k=1|\theta) = (\pi_k \mathcal{N}(X_n|\mu_k,\Sigma_k))^{z_k} = \pi_k \mathcal{N}(X_n|\mu_k,\Sigma_k)$$

$$\rightarrow P(Z_k|x_n,\theta) = P(Z_k=1|x_n,\theta) = \frac{\pi_k \mathcal{N}(X_n|\mu_k,\Sigma_k)}{P(x)}$$

And because x is drawn from a GMM it's pdf is: $\sum_k \pi_k \mathcal{N}(x|\mu_k,\Sigma_k)$

$$\rightarrow P(Z_k|x_n,\theta) = \frac{\pi_k \mathcal{N}(X_n|\mu_k,\Sigma_k)}{\sum_{l=1}^{K}\pi_l \mathcal{N}(x_n|\mu_l,\Sigma_l)}$$

Exercise 2:

a) Because $\gamma_k(x_n)$ is calculated in the E-step we regard it as a constant in the M-step. Therefore we want to maximize $L(\theta)$ w.r.t μ_k for all $k \in K$ with a constant responsibility.

Because only one of the K terms are meaningful for taking the derivative w.r.t. to each k, the function is then simplified to:

$$\hat{L}(\theta) = \sum_{n=1}^{N} \gamma_k(x_n)\left(\ln \pi_k + \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right) \quad \text{for some specific } k \in K.$$

Additionally the $\ln \pi_k$ term is irrelevant to this derivative and finally:

$$\hat{L}(\theta) \equiv \sum_{n=1}^{N} \gamma_k(x_n) \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k) \qquad (\mathrm{I})$$

$$\rightarrow \frac{\partial \hat{L}(\theta)}{\partial \mu_k} = \partial\left(\sum_{n=1}^{N} \gamma_k(x_n) \ln\left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}\right)\right) / \partial\mu_k$$

$$= \partial\left(\sum_{n=1}^{N} \gamma_k(x_n)\left[\ln \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} - \underbrace{\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}_{x^TAx \text{ form}}\right]\right) / \partial\mu_k$$

$$= \sum_{n=1}^{N} \gamma_k(x_n)\, \partial\left[-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right] / \partial\mu_k = \sum_{n=1}^{N} \gamma_k(x_n)\, \Sigma_k^{-1}(x_n-\mu_k) = 0$$

Because $\gamma_k(x_n)$ is a constant we can multiply both sides by Σ_k from right and get:

$$\sum_{n=1}^{N} \gamma_k(x_n)(x_n-\mu_k) = 0 \Rightarrow -\sum_{n=1}^{N} \gamma_k(x_n)\mu_k + \sum_{n=1}^{N} \gamma_k(x_n)x_n = 0$$

$$\rightarrow \mu_k \sum_{n=1}^{N} \gamma_k(x_n) = \sum_{n=1}^{N} \gamma_k(x_n)x_n \rightarrow \boxed{\mu_k = \frac{\sum_{n=1}^{N} \gamma_k(x_n)x_n}{\sum_{n=1}^{N} \gamma_k(x_n)}}$$

b) starting from (I) from part a), i.e.,

$$\hat{l}(\theta) = \sum_{n=1}^{N} \gamma_k(x_n) \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

we have: $$\frac{\partial \hat{l}(\theta)}{\partial \Sigma_k} = \partial\left(\sum_{n=1}^{N} \gamma_k(x_n) \ln\left[\frac{1}{(2\pi)^{d/2}|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}\right]\right) \Big/ \partial \Sigma_k$$

$$= \partial\left(\sum_{n=1}^{N} \gamma_k(x_n)\left[-\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right]\right) \Big/ \partial \Sigma_k$$

$$\frac{\partial \ln|\Sigma_k|}{\partial \Sigma_k} = \frac{\partial |\Sigma_k| / \partial \Sigma_k}{|\Sigma_k|} = \frac{|\Sigma_k| \Sigma_k^{-1}}{|\Sigma_k|} = \Sigma_k^{-1}$$

$$\partial (x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k) \big/ \partial \Sigma_k = -\Sigma_k^{-T}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-T}$$

$$\xrightarrow{\Sigma,\ \Sigma^{-1} \text{ symmetrical}} -\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1}$$

$$\Rightarrow \frac{\partial \hat{l}(\theta)}{\partial \Sigma_k} = \sum_{n=1}^{N} \gamma_k(x_n)\left[-\frac{1}{2}\Sigma_k^{-1} + \frac{1}{2}\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1}\right] = 0$$

multiplying from left and right by $\Sigma_k \Rightarrow$

$$\frac{1}{2}\sum_{n=1}^{N} \gamma_k(x_n)\Big[\overbrace{(x_n-\mu_k)(x_n-\mu_k)^T}^{A} - \underbrace{\Sigma_k \cancel{\Sigma_k^{-1}} \Sigma_k}_{\Sigma_k}\Big] = 0 \Rightarrow$$

$$\sum_{n=1}^{N} \gamma_k(x_n) A - \sum_{n=1}^{N} \gamma_k(x_n) \Sigma_k \rightarrow \Sigma_k \sum_{n=1}^{N} \gamma_k(x_n) = \sum_{n=1}^{N} \gamma_k(x_n) A$$

$$\Rightarrow \boxed{\Sigma_k = \frac{\sum_{n=1}^{N} \gamma_k(x_n)(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} \gamma_k(x_n)}}$$