Homework 3

Haoyu Zhen

2022 年 5 月 15 日

1 Some Lemmas in Linear Algebra

Lemma 1.

$$x^T A x = \operatorname{tr}\big(A x x^T\big).$$

Lemma 2.

$$\frac{\partial \log |A|}{\partial A} = 2A^{-1} - \operatorname{diag}(A^{-1}) \quad \textit{and} \quad \frac{\partial \operatorname{tr}(AB)}{\partial A} = (B + B^T) - \operatorname{diag}(B).$$

2 Update Σ_k in EM

Now we want to find a good enough Σ_k. Our goal is to maximize the liklihood

$$\mathcal{L} = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)\right)$$

where

$$\mathcal{N}(x_n|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k)\right).$$

Thus

$$\begin{aligned}
\frac{\partial L}{\partial A} &= \sum_{n=1}^{N} \left\{ \frac{1}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times \frac{\partial}{\partial A} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\} \\
&= \sum_{n=1}^{N} \left\{ \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times \frac{\partial}{\partial A} \left[-\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1}(x_n - \mu_k) \right] \right\}. \\
&= \sum_{n=1}^{N} \left\{ \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times \frac{\partial}{\partial A} \left[\frac{1}{2} \log \left|\Sigma_k^{-1}\right| - \frac{1}{2} \operatorname{tr}\big(\Sigma_k^{-1} N_{n,k}\big) \right] \right\}
\end{aligned}$$

where $N_{n,k} = (x_n - \mu_k)(x_n - \mu_k)^T$. $\partial L / \partial \Sigma_k^{-1} = 0$ gives

$$\sum_{n=1}^{N} \left\{ \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times \frac{1}{2} \Big[2(\Sigma_k - N_{n,k}) - \operatorname{diag}(\Sigma_k - N_{n,k}) \Big] \right\} = 0$$

which entails that

$$\sum_{n=1}^{N} \left[\frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times (\Sigma_k - N_{n,k}) \right] = 0$$

which entails

$$\Sigma_k = \left(\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \right)^{-1} \sum_{n=1}^{N} \left[\frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)} \times (x_n - \mu_k)(x_n - \mu_k)^T \right].$$

For simplicity,

$$\gamma(z_n k) \triangleq \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)}$$

Finally

$$\Sigma_k = \frac{1}{\sum_{n=1}^{N} \gamma(z_n k)} \sum_{n=1}^{N} \left[\gamma(z_n k)(x_n - \mu_k)(x_n - \mu_k)^T \right].$$

Nota Bena: if we update μ first, then we could let μ_k above be μ_k^{new}.