Machine Learning Homework 03

Qiu Yihang

May.10, 2022

GMM: $p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Hidden variable z_i^k denotes the possibility that x_i is of class k.

Define $\gamma(z_{ik}) \triangleq p(z_i^k|\mathbf{x}_i) = \frac{\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K}\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}$, $N_k \triangleq \sum_{n=1}^{N} \gamma(z_{nk})$.

The log-likelihood is $\mathcal{L} = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right)$.

In the EM, we have $\boldsymbol{\mu}_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n$.

Show that in the EM, $\boldsymbol{\Sigma}_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k^{new})(\mathbf{x}_n - \boldsymbol{\mu}_k^{new})^T$.

Proof. We have

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\Sigma}}\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\left(\sqrt{|\boldsymbol{\Sigma}|}\frac{\partial|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\partial\boldsymbol{\Sigma}} - \frac{1}{2}\frac{\partial(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\partial\boldsymbol{\Sigma}}\right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\left(-\frac{1}{2}|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\boldsymbol{\Sigma}|^{-\frac{3}{2}}\frac{\partial|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\Sigma}} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\left(\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^T\right) \\
&= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\left(-\frac{1}{2}|\boldsymbol{\Sigma}|^{-1}|\boldsymbol{\Sigma}|\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\left(\boldsymbol{\Sigma}^{-1}\right)^T\right) \\
&= -\frac{1}{2}\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\right) \\
&\qquad \text{(By the symmetry of } \boldsymbol{\Sigma} \text{ and } \boldsymbol{\Sigma}^{-1}\text{)}
\end{aligned}$$

Set the gradient to 0. We have

$$\begin{aligned}
0 = \frac{\partial\mathcal{L}}{\partial\boldsymbol{\Sigma}_k} &= \sum_{n=1}^{N}\frac{\partial}{\partial\boldsymbol{\Sigma}_k}\ln\left(\sum_{k=1}^{K}\pi_k\cdot\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right) \\
&= \sum_{n=1}^{N}\frac{1}{\sum_{k=1}^{K}\pi_k\cdot\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}\frac{\partial}{\partial\boldsymbol{\Sigma}_k}\left(\pi_k\cdot\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)\right) \\
&= -\frac{1}{2}\sum_{n=1}^{N}\frac{\pi_k\cdot\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{k=1}^{K}\pi_k\cdot\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}\left(\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right) \\
&= -\frac{1}{2}\sum_{n=1}^{N}\gamma(z_{nk})\left(\boldsymbol{\Sigma}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right)
\end{aligned}$$

i.e.

$$\boldsymbol{\Sigma}_k^{-1}\sum_{n=1}^{N}\gamma(z_{nk}) = \sum_{n=1}^{N}\gamma(z_{nk})\boldsymbol{\Sigma}_k^{-1} = \sum_{n=1}^{N}\gamma(z_{nk})\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}$$

$$= \boldsymbol{\Sigma}_k^{-1}\left(\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\right)\boldsymbol{\Sigma}_k^{-1}$$

(Since $\gamma(z_{nk})$ is a number.)

$$\iff \boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_k\sum_{n=1}^{N}\gamma(z_{nk}) = \boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_k^{-1}\left(\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\right)\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Sigma}_k$$

$$\iff \qquad N_k\boldsymbol{\Sigma}_k = \sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T$$

Therefore,

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T$$

Qed. ■