ISYE/CSE 6740 Homework 2

Yiming Tong

September 22, 2019

1 Q2

It's obvious that $f_v(x) = (x^T v)v$. Hence, the target function becomes

$$\begin{aligned} &\underset{||v||}{argmin} \sum_{i=1}^{n} ||x_i - (x_i^T v)v||^2 \\ =& \underset{||v||}{argmin} \sum_{i=1}^{n} (x_i^T x_i - 2(x_i^T v)^2 + (x_i^T v)^2 v^T v) \\ =& \underset{||v||}{argmin} \sum_{i=1}^{n} (x_i^T x_i - (x_i^T v)^2) \\ =& \underset{||v||}{argmin} (\Sigma - v\Sigma v^T) \end{aligned}$$

which is constraint by $v^T v = 1$, where $\Sigma = \sum_{i=1}^{n} x_i^T x_i$ is the covariance matrix of the components of the data set X. This is exactly the same optimization problem as in PCA, since the first term Σ is independent with argument v. Thus, $\underset{||v||}{argmin} \sum_{i=1}^{n} ||x_i - (x_i^T v)v||^2$ gives the principle component.

2 Q4

(a)$\mathcal{L}(\Delta_i, h_i) = log \prod_{i=1}^{m} (\frac{h_i \Delta_i}{\Sigma_i h_i \Delta_i})^{n_i}$.
(b)Added Lagrange multiplier, the target function is obained as:

$$\begin{aligned} L(h_i, \lambda) =& log \prod_{i=1}^{m} (h_i \Delta_i)^{n_i} + \lambda(1 - \sum_i \Delta_i h_i) \\ =& \Sigma_i n_i log(\Delta_i h_i) - \lambda \sum_i \Delta_i h_i + \lambda. \end{aligned}$$

Taking $\frac{\partial L}{\partial h_i}$ gives $\frac{n_i}{h_i} - \lambda \Delta_i = 0, h_i = \frac{n_i}{\lambda \Delta_i}$. Then we can determine λ by normalizing the probability: $\sum_i \Delta_i h_i = \sum_i n_i/\lambda = 1, \lambda = \sum_i n_i = N$.
In summary, the maximum log likehood esitimator $h_i = \frac{n_i}{N \Delta_i}$.
(c)

- F: More like have many parameters. The number of parameters $\sim$ number of samples.
- F: Too many bins in high dimensional cases; Full bandwidth induces higher statistical risk.
- T: The shape follows the model you choose, e.g. guassian.

3 Q5

(a) For given $z^{(k)}$, only the k^{th} term in the product exists, i.e.

$$p(z = z^{(k)}) = \pi_k,$$
$$p(x|z = z^{(k)}) = \mathcal{N}(x|\mu_k, \Sigma_k).$$

Thus,

$$\begin{aligned}(2) &= \sum_{z \in Z} p(z)p(x|z) \\ &= \sum_{k} p(z^{(k)})p(x|z^{(k)}) \\ &= \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) = (1).\end{aligned}$$

(b)

$$\begin{aligned}p(z_k^n = 1|x_n) =& \frac{p(z_n^k = 1)p(x_n|z_k^n = 1)}{p(x_n)} \\ =& \frac{\pi_k \times \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_k p(z_n^k = 1)p(x_n|z_k^n = 1)} \\ =& \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_k \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)},\end{aligned}$$

where $\mathcal{N}\left(x_i|\mu_k, \Sigma_k\right) := \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)\right)$.

(c) In M-step we maximize the following target function, which is the log-likehood function of sum of K normal distributions:

$$f(\pi_k, \Sigma_k, \mu_k) = \sum_{i=1}^{m} \sum_{k=1}^{K} \tau_k^i \left[\log \pi_k - \left(x^i - \mu_k\right)^T \Sigma_k \left(x^i - \mu_k\right) + \log \Sigma_k + c\right],$$

which is constraint by $\Sigma \pi_k = 1$. As usual we add Lagrange mutiplexer, the target function becomes:

$$L(\pi_k, \Sigma_k, \mu_k, \lambda) = \sum_{i=1}^{m} \sum_{k=1}^{K} \tau_k^i \left[\log \pi_k - \left(x^i - \mu_k\right)^T \Sigma_k \left(x^i - \mu_k\right) + \log \Sigma_k + c\right] - \lambda(1 - \sum \pi_k).$$

By setting the partial derivative of π_k, Σ_k, μ_k and λ to zero, we find out:

$$\begin{gathered}\sum_i \frac{\tau_k^i}{\pi_k} - \lambda = 0,\\ \sum_i \tau_k^i \Sigma_k (x^i - \mu_k) = 0,\\ \sum_i \tau_k^i [(x^i - \mu_k)^T (x^i - \mu_k) + \Sigma_k^{-1}] = 0,\\ \sum_k \pi_k = 0.\end{gathered}$$

By solving these equations, we could come to the updated π_k, μ_k and Σ_k:

$$\begin{aligned}\pi_k &= \frac{\sum_i \tau_k^i}{m},\\ \mu_k &= \frac{\sum_i \tau_k^i x^i}{\sum_i \tau_k^i},\\ \Sigma_k &= \frac{\sum_i \tau_k^i (x^i - \mu_k)^T (x^i - \mu_k)}{\sum_i \tau_k^i}.\end{aligned}$$

(d) By substituting $\Sigma_k = \epsilon I$ into normal distribution we get

$$\mathcal{N}(x^i, \mu_k, \Sigma_k = \epsilon I) = \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{1}{2\epsilon}||x^i - \mu_k||^2}.$$

Then the τ_k^i is given by

$$\tau_k^i = \frac{\pi_k exp(-||x^i - \mu_k||^2/2\epsilon)}{\Sigma_k \pi_k exp(-||x^i - \mu_k||^2/2\epsilon)} \to \gamma_i^k,$$

as $\epsilon \to 0$, where $\gamma_{ik} = 1$ if x^i is closest to μ_k and $\gamma_{ik} = 0$ otherwise. This is because as $\epsilon \to 0$, only the term with the smallest $||x^i - \mu_k||^2$ is significant. In this case, the log likehood function becomes:

$$f(\pi_k, \mu_k) = \sum_n \sum_k \gamma_{nk} (log(\pi_k) - \frac{1}{2\epsilon}||x^n - \mu_k||^2 + log(\frac{1}{\sqrt{2\pi\epsilon}})) \to -\sum_n \sum_k \gamma_{nk} \frac{1}{2\epsilon}||x^n - \mu_k||^2,$$

as $\epsilon \to 0$. To maximize $f(\pi_k, \mu_k)$ is equivalent to minimize $J = \sum_n \sum_k \gamma_{nk} \|x_n - \mu_k\|^2$ in this case.

(e)

$$\begin{aligned}\mu_{mixture} &= \sum_k \pi_k \mu_k\\ \Sigma_{mixture} &= \sum_k \pi_k \Sigma_k.\end{aligned}$$