ECE C143A Homework 3

Lawrence Liu

May 12, 2022

Problem 1

(a)

We have

$$\begin{aligned} L &= \Pi_{i=0} P(\mathbf{x}_i, t_i) \\ &= \Pi_{i=0} P(t_i) P(\mathbf{x}_i | t_i) \end{aligned}$$

let $P(t_i = j) = P(C_j)$, therefore we have

$$\begin{aligned} L &= \Pi_{i\in C_1} P(C_1) P(\mathbf{x}_i | C_1) \dots \Pi_{i\in C_k} P(C_k) P(\mathbf{x}_i | C_k) \\ \log(L) &= \sum_{i=1}^{k} N_i \log(P(c_i)) + \sum_{j\in C_1} \log(P(\mathbf{x}_j | C_1)) + \dots + \sum_{j\in C_k} \log(P(\mathbf{x}_j | C_k)) \end{aligned}$$

We have

$$P(\mathbf{x}_j | C_k) = (2\pi \det(\Sigma_k))^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_j - \mu_k))$$

$$\log(P(\mathbf{x}_j | C_k)) = -\frac{1}{2} \log(2\pi \det(\Sigma_k)) - \frac{1}{2}(\mathbf{x}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_j - \mu_k)$$

In order to find the maximum likelihood estimator μ_k we find the values of μ_k such that $\frac{\partial \log L}{\partial \mu_k} = 0$, Therefore we have

$$\frac{\partial \log L}{\partial \mu_k} = \sum_{j \in C_k} -\frac{1}{2} \frac{\partial}{\partial \mu_k} (\mathbf{x}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_j - \mu_k) \quad = -\frac{1}{2} \sum_{j \in C_k} (\Sigma_k + \Sigma_k^T)(\mathbf{x}_k - \mu_k)$$
$$= -\frac{1}{2}(\Sigma_k + \Sigma_k^T) \sum_{j \in C_k} \mathbf{x}_k - \mu_k$$

therefore we have that in order for $\frac{\partial \log L}{\partial \mu_k} = 0$,

$$\sum_{j \in C_k} \mathbf{x}_k - \mu_k = 0$$

$$\boxed{\mu_k = \frac{1}{N_k} \sum_{j \in C_k} \mathbf{x}_k}$$

Likewise, the maximum likelihood estimator of Σ_k is the values such that $\frac{\partial \log L}{\partial \Sigma_k} = 0$, this is equivalent to finding the values of Σ_k^{-1} such that $\frac{\partial \log L}{\partial \Sigma_k^{-1}} = 0$

$$\frac{\partial \log L}{\partial \Sigma_k^{-1}} = 0$$
$$\frac{1}{2} \frac{\partial N_k \log(2\pi \det(\Sigma_k)) + \sum_{i \in k} (\mathbf{x}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_j - \mu_k)}{\partial \Sigma_k^{-1}} = 0$$

We have

$$\frac{\partial \log(2\pi \det(\Sigma_k))}{\partial \Sigma_k^{-1}} = \frac{\partial \log(\det(\Sigma_k))}{\partial \Sigma_k^{-1}}$$
$$= \frac{1}{\det(\Sigma_k)} \frac{\partial \det(\Sigma_k)}{\partial \Sigma_k^{-1}}$$
$$= \frac{1}{\det(\Sigma_k)} (-\det(\Sigma_k)) \Sigma_k^T$$
$$= -\Sigma_k^T$$

Furthermore we have

$$\frac{\partial (\mathbf{x}_j - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_j - \mu_k)}{\partial \Sigma_k^{-1}} = (\mathbf{x}_j - \mu_k)^T (\mathbf{x}_j - \mu_k)$$

Therefore we have

$$
\begin{aligned}
0 = \frac{\partial \log L}{\partial \Sigma_k^{-1}} &= -N_k \Sigma_k^T + \sum_{i \in C_k} (\mathbf{x}_j - \mu_k)^T (\mathbf{x}_j - \mu_k) \\
N_k \Sigma_k^T &= \sum_{i \in C_k} (\mathbf{x}_j - \mu_k)^T (\mathbf{x}_j - \mu_k) \\
\Sigma_k &= \boxed{\frac{1}{N_k} \sum_{i \in C_k} (\mathbf{x}_j - \mu_k)(\mathbf{x}_j - \mu_k)^T}
\end{aligned}
$$

And to find the maximum likelihood estimator for $P(C_i)$ we must find the value of $P(C_i)$ for any $i \in k$ such that $\sum_{j \in k} N_j \log(p(C_j))$ is maximized, since $\sum_{i \in K} p(C_i) = 1$, we have $\sum_{j \in k} N_j \log(p(C_j)) = \sum_{j \in k} N_j \log(p(C_j)) + \lambda(\sum_{i \in K} p(C_j) - 1)$ Therefore taking the derivative of each with respect to $P(C_i)$ for any $i \in k$ we get

$$
\begin{aligned}
\frac{\partial}{\partial P(C_1)} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \frac{N_j}{C_j} + \lambda = 0 \\
&\vdots \\
\frac{\partial}{\partial P(C_k)} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \frac{N_k}{C_k} + \lambda = 0 \\
\frac{\partial}{\partial \lambda} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \sum_{i \in K} p(C_i) = 1
\end{aligned}
$$

Solving these we get that $p(C_k) = \boxed{\frac{N_k}{N}}$

(b)

We have

$$
\begin{aligned}
L &= \Pi_{i=0} P(\mathbf{x}_i, t_i) \\
&= \Pi_{i=0} P(t_i) P(\mathbf{x}_i | t_i)
\end{aligned}
$$

let $P(t_i = j) = P(C_j)$, therefore we have

$$
\begin{aligned}
L &= \Pi_{i \in C_1} P(C_1) P(\mathbf{x}_i|C_1) \ldots \Pi_{i \in C_k} P(C_k) P(\mathbf{x}_i|C_k) \\
\log(L) &= \sum_{i=1}^{k} N_i \log(P(c_i)) + \sum_{j \in C_1} \log(P(\mathbf{x}_j|C_1)) + \cdots + \sum_{j \in C_k} \log(P(\mathbf{x}_j|C_k))
\end{aligned}
$$

We have

$$
P(\mathbf{x}_j|C_k) = \Pi_{i=1}^{D} P(x_i|C_k)
$$

where

$$
P(x_i|C_k) = \text{Poisson}(\lambda_{ik}) = \frac{\lambda_{ik}^{x_i} e^{-\lambda_{ik}}}{x_i!}
$$

Therefore we have

$$
\begin{aligned}
log(P(\mathbf{x}_j|C_k)) &= \sum_{i=1}^{D} \log\left(\frac{\lambda_{ik}^{x_i} e^{-\lambda_{ik}}}{x_i!}\right) \\
&= \sum_{i=1}^{D} \left(x_i \log(\lambda_{ik}) - \lambda_{ik} - \log(x_i!)\right)
\end{aligned}
$$

the maximum likelihood estimator of λ_{ik} is the values such that $\frac{\partial \log L}{\partial \lambda_{ik}} = 0$, we have

$$
\begin{aligned}
\frac{\partial}{\partial \lambda_{ik}} \log(L) &= \frac{\partial}{\partial \lambda_{ik}} \sum_{i=1}^{k} N_i \log(P(c_i)) + \frac{\partial}{\partial \lambda_{ik}} \sum_{j \in C_1} \log(P(\mathbf{x}_j|C_1)) + \cdots + \frac{\partial}{\partial \lambda_{ik}} \sum_{j \in C_k} \log(P(\mathbf{x}_j|C_k)) \\
&= \frac{\partial}{\partial \lambda_{ik}} \sum_{j \in C_k} \log(P(\mathbf{x}_j|C_k))
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \lambda_{ik}} \log(P(\mathbf{x}_j|C_k)) &= \frac{\partial}{\partial \lambda_{ik}} \left(x_i \log(\lambda_{ik}) - \lambda_{ik} - \log(x_i!)\right) \\
&= \frac{x_i}{\lambda_{ik}} - 1
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \lambda_{ik}} \log(L) \\
0 &= \sum_{j \in C_k} \frac{x_j i}{\lambda_{ik}} - 1 \\
N_k \lambda_{ik} &= \sum_{j \in C_k} x_j i \\
\lambda_{ik} &= \boxed{\frac{1}{N_k} \sum_{j \in C_k} x_j i}
\end{aligned}
$$

And to find the maximum likelihood estimator for $P(C_i)$ we must find the value of $P(C_i)$ for any $i \in k$ such that $\sum_{j \in k} N_j \log(p(C_j))$ is maximized, since $\sum_{i \in K} p(C_i) = 1$, we have $\sum_{j \in k} N_j \log(p(C_j)) = \sum_{j \in k} N_j \log(p(C_j)) + \lambda(\sum_{i \in K} p(C_j) - 1)$ Therefore taking the derivative of each with respect to $P(C_i)$ for any $i \in k$ we get

$$
\begin{aligned}
\frac{\partial}{\partial P(C_1)} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \frac{N_j}{C_j} + \lambda = 0 \\
&\vdots \\
\frac{\partial}{\partial P(C_k)} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \frac{N_k}{C_k} + \lambda = 0 \\
\frac{\partial}{\partial \lambda} \sum_{j \in k} N_j p(C_j) + \lambda(\sum_{i \in K} p(C_i) - 1) &= \sum_{i \in K} p(C_i) = 1
\end{aligned}
$$

Solving these we get that $p(C_k) = \boxed{\frac{N_k}{N}}$