模式识别第一次作业

孙浩淼 201928013229100

Email: sunhm15@gmail.com

1 问题一

a) 证明过程如下: 根据统计独立的性质

$$p(x_i, x_j) = p(x_i)p(x_j)$$

和概率的性质

$$\int p(x_i)dx_i = 1, \quad \int p(x_i)x_i dx_i = x_i$$

可以推导出,

$$\begin{aligned}
\sigma_{ij} &= \varepsilon[(x_i - \mu_i)(x_j - \mu_j)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_i, x_j)(x_i - \mu_i)(x_j - \mu_j)dx_i dx_j \\
&= \int_{-\infty}^{\infty} p(x_i)(x_i - \mu_i)dx_i \int_{-\infty}^{\infty} p(x_j)(x_j - \mu_j)dx_j \\
&= 0
\end{aligned}$$

b) 对于二维正态分布而言,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right)$$

有,

$$p(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

若 x_1, x_2 统计独立，即

$$\sigma_{21} = \sigma_{12} = \varepsilon[(x_1 - \mu_1)(x_2 - \mu_2)] = 0$$

有,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \quad |\Sigma| = \sigma_1^2\sigma_2^2$$

$p(x_1, x_2)$ 可以重新表示为，

$$
\begin{aligned}
p(x_1,x_2) = &\quad \frac{1}{2\pi\sigma_1\sigma_2} exp\left[-\frac{1}{2}\{(\frac{x_1-\mu_1}{2})^2 + (\frac{x_2-\mu_2}{2})^2\}\right] \\
= &\quad \frac{1}{\sqrt{2\pi}\sigma_1} exp\left[-\frac{1}{2}(\frac{x_1-\mu_1}{2})^2\right] \cdot \frac{1}{\sqrt{2\pi}\sigma_2} exp\left[-\frac{1}{2}(\frac{x_2-\mu_2}{2})^2\right] \\
= &\quad p(x_1)p(x_2)
\end{aligned}
$$

2 问题二

a)

$$P[z_{ik} = 1|p(\omega_i)] = P(\omega_i)$$

$$P[z_{ik} = 0|p(\omega_i)] = 1 - P(\omega_i)$$

上述两个表达式可以合并为

$$P(z_{ik}|P(\omega_i)) = [P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}$$

因此

$$P(z_{i1}, z_{i2}, ..., z_{in}|P(\omega_i)) = \prod_{k=1}^{n} P(z_{ik}|P(\omega_i)) = \prod_{k=1}^{n} [P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}$$

b) 对数似然函数可以表示为

$$
\begin{aligned}
l(P(\omega_i)) &= log(P(z_{i1}, z_{i2}, ..., z_{in}|P(\omega_i))) \\
&= log\left[\prod_{k=1}^{n} [P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}\right] \\
&= \sum_{k=1}^{n} z_{ik}[log[P(\omega_i)] + (1-z_{ik})log(1-P(\omega_i))]
\end{aligned}
$$

最大似然解需要满足

$$\nabla_{P(\omega_i)} l(P(\omega_i)) = \frac{\sum\limits_{k=1}^{n} z_{ik}}{P(\omega_i)} - \frac{\sum\limits_{k=1}^{n}(1-z_{ik})}{1-P(\omega_i)} = 0$$

解得，

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$$

3 问题三

$p_2(x)$ 和 $p_1(x)$ 之间的距离可以表示为，

$$D_{KL}(p_2, p_1) = \int p_2(x) ln \frac{p_2(x)}{p_1(x)} dx$$

将 $p_1(x)$ 的表达式代入散度公式可以得到，

$$D_{KL}(p_2, p_1) = \int p_2(x) ln(p_2(x)) dx + \frac{1}{2} \int p_2(x)[dln(2\pi) + ln|\Sigma| + (x-\mu)^t \Sigma^{-1}(x-\mu)]$$

对 μ 求导，有，

$$\frac{\partial}{\partial \mu} D_{KL}(p_2, p_1) = -\int \Sigma^{-1}(x-\mu) p_2(x) dx = 0$$

因为 Σ 为非奇异矩阵，所以可以消去，即

$$\int (x-\mu) p_2(x) dx = 0$$

利用 $\int p_2(x)dx = 1$，可以得到

$$\mu = \int x p_2(x) dx = \varepsilon_2[x]$$

同理，对 Σ 求导，有，

$$\frac{\partial}{\partial \Sigma} D_{KL}(p_2, p_1) = -\int p_2(x)[-\Sigma^{-1} + \Sigma^{-1}(x-\mu)(x-\mu)^T \Sigma^{-1}] dx = 0$$

因此，

$$\Sigma = \varepsilon[(x-\mu)(x-\mu)^T]$$

4 问题四

a) 对于均值，

$$\mu_{n+1} = \frac{1}{n+1} \sum_{k=1}^{n+1} x_i = \frac{1}{n+1} [\sum_{k=1}^{n} x_i + x_{n+1}] = \frac{1}{n+1}(n\mu_n + x_{n+1}) = \mu_n + \frac{1}{n+1}(x_{n+1} - \mu_n)$$

对于协方差矩阵，有，

$$
\begin{aligned}
C_{n+1} =& \frac{1}{n}\sum_{k=1}^{n+1}(x_k-\mu_{n+1})(x_k-\mu_{n+1})^T \\
=& \frac{1}{n}\left[\sum_{k=1}^{n}(x_k-\mu_{n+1})(x_k-\mu_{n+1})^T+(x_{n+1}-\mu_{n+1})(x_{n+1}-\mu_{n+1})^T\right] \\
=& \frac{1}{n}\left[\sum_{k=1}^{n}(x_k-\mu_n)(x_k-\mu_n)^T-\frac{1}{n+1}(x_{n+1}-\mu_n)\sum_{k=1}^{n}(x_k-\mu_n)^T\right. \\
& \left.-\frac{1}{n+1}\left(\sum_{k=1}^{n}(x_k-\mu_n)\right)(x_{n+1}-\mu_n)^T+\frac{1}{(n+1)^2}\sum_{k=1}^{n}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^T\right] \\
& +\frac{1}{n}\left(x_{n+1}-(\mu_n+\frac{1}{n+1}(x_{n+1}-\mu_n))\right)\left(x_{n+1}-(\mu_n+\frac{1}{n+1}(x_{n+1}-\mu_n))\right)^T \\
=& \frac{1}{n}\left[(n-1)C_n+\frac{n}{(n+1)^2}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^T\right]+\frac{1}{n}\left(\frac{n^2}{(n+1)^2}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^T\right) \\
=& \frac{n-1}{n}C_n+\frac{1}{n+1}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^T
\end{aligned}
$$

b) 递归公式的计算复杂度分别为 $O(d)$ 和 $O(d^2)$，但如果从头开始计算的话，复杂度分别为 $O(dn)$ 和 $O(d^2n)$

5 问题五

a) 根据 E 步的公式有，

$$
\begin{aligned}
Q(\theta;\theta_0) =& \varepsilon_{x_{32}}[ln\ p(x_g,x_b;\theta)|\theta_0,D_g] \\
=& \int\left[\sum_{i=1}^{2}p(x_i|\theta)+p(x_3|\theta)\right]p(x_{32}|\theta_0;x_{31}=2)dx_{32} \\
=& \sum_{i=1}^{2}lnp(x_i|\theta)+\int lnp(x|\theta)\frac{p(x|\theta_0)}{\int p(x'=(2,x'_{32})^T|\theta_0)dx'_{32}}dx_{32} \\
=& -4\theta_1^{-1}-2ln(\theta_1\theta_2)+2e\int lnp(x|\theta)p(x|\theta_0)dx_{32} \\
=& -4\theta_1^{-1}-2ln(\theta_1\theta_2)+\frac{1}{4}\int lnp(x|\theta)dx_{32} \\
=& -4\theta_1^{-1}-2ln(\theta_1\theta_2)+\frac{1}{4}\int_0^{\min\{\theta_2,4\}} lnp(x|\theta)dx_{32}
\end{aligned}
$$

根据 θ 可以分三个情况：

$$
Q(\theta;\theta_0)=\begin{cases}-4\theta_1^{-1}-2ln(\theta_1\theta_2)-\frac{\theta_2}{4}(2\theta_1^{-1}+ln(\theta_1\theta_2)) & 3\le\theta_2<4 \\ -6\theta_1^{-1}-3ln(\theta_1\theta_2) & \theta_2\ge4\end{cases}
$$

b) 对于 θ_1 而言，

$$\frac{\partial}{\partial\theta_1}Q=(2-\frac{\theta_2}{4})(\frac{2}{\theta_1^2}-\frac{1}{\theta_1})=0$$

解得，

$$\arg\max_{\theta_1}Q=2$$

同理，对于 θ_2 而言，

$$\frac{\partial}{\partial\theta_2}Q=\begin{cases}-\frac{2}{\theta_2}-\frac{1}{4}(2\theta_1^{-1}+ln(\theta_1\theta_2))-\frac{1}{4} & 3\leq\theta_2<4\\ -\frac{3}{\theta_2} & \theta_2\geq 4\end{cases}$$

$\frac{\partial}{\partial\theta_2}Q<0$，因此，

$$\arg\max_{\theta_2}Q=3$$

因此，

$$\arg\max_{\theta}Q=(2,3)^T,\quad \max Q=-\frac{11}{4}-\frac{11}{4}ln6=-7.677$$

6 问题六

$\hat{a}_{ij},\hat{b}_{ij}$ 的迭代公式显示如下，

$$\gamma_{ij}(t)=\frac{\alpha_i(t-1)a_{ij}b_{ij}\beta_i(t)}{P(V^T|\theta)}$$

$$\hat{a}_{ij}=\frac{\sum\limits_{t=1}^{T}\gamma_{ij}(t)}{\sum\limits_{t=1}^{T}\sum\limits_{k}\gamma_{ik}(t)}$$

$$\hat{b}_{jk}=\frac{\sum\limits_{v(t)=k}\sum\limits_{l}\gamma_{jl}(t)}{\sum\limits_{t=1}^{T}\sum\limits_{l}\gamma_{jl}(t)}$$

其中 $\alpha_i(t)$和$P(V^T|\theta)$ 可以由 HMM 前向迭代公式同时求出，复杂度为 $O(c^2T)$，同理，β_j 可以由后向迭代公式得出，复杂度为 $O(c^2T)$。若已知 $\alpha_{ij},\beta_{ij},P(V^T|\theta)$, 可以经过 $O(c^2T)$ 次操作计算出 γ_{ij}。因此，γ_{ij} 的计算复杂度可以表示为，

$$T(\gamma_{ij})=T(\alpha_{ij},P(V_T|\theta))+T(\beta_{ij})+O(c^2T)=O(c^2T)$$

接下来，根据求出的 $\gamma_{ij},\hat{a}_{ij},\hat{b}_{ij}$ 的计算复杂度均为 $O(c^2T)$。故总的迭代过程的时间复杂度为 $O(c^2T)$。