第二次作业

姓名：李千鹏 学号：2021E8014682047 院系：人工智能学院

问题 1.

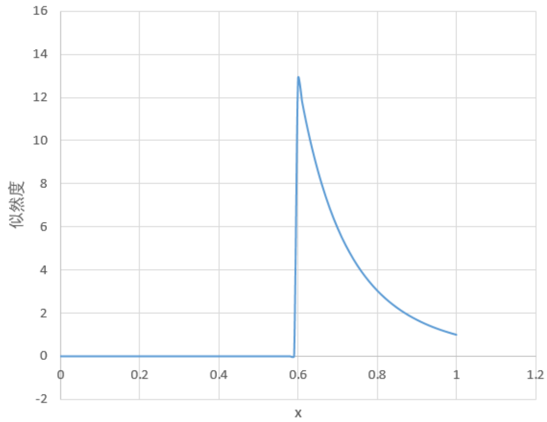
(1)

似然度函数：$L(\theta) = \prod_{i=1}^{n} p(x_i \mid \theta) = \frac{1}{\theta^n} \quad (\theta \geqslant \max\{D\})$

对似然度函数求偏导：$\frac{\partial L}{\partial \theta} = \frac{-n}{\theta^{n+1}} < 0$

由于似然度函数在定义域内单调递减，所以 $\hat{\theta} = \max\{D\}$

(2)



右图为似然度函数

由 *(1)* 式的似然度函数可知，当 $\theta < \max\{D\}$ 时，不可能出现样本 $\max\{D\}$。当且仅当 $\theta \geqslant \max\{D\}$ 时，样本集 D 的所有样本才会存在，所以不需要知道其他四个点的值。

问题 2.

(1)

后验估计：

$$\ln(l(\mu)p(\mu)) = -\frac{n}{2}\ln\left((2\pi)^d|\Sigma|\right) - \frac{1}{2}\sum_{k=1}^{n}(x_k-\mu)^\top \Sigma^{-1}(x_k-\mu)$$
$$-\frac{1}{2}\ln\left((2\pi)^d|\Sigma_0|\right) - \frac{1}{2}(\mu-m_0)^\top \Sigma_0^{-1}(\mu-m_0)$$

求偏导得：$\frac{\partial \ln(l(\mu)p(\mu))}{\partial \mu} = \sum_{k=1}^{n} \Sigma^{-1}(x_k-\mu) + \Sigma_0^{-1}(m_0-\mu)$

令偏导等于 *0* 得：$\hat{\mu} = \left(n\Sigma^{-1} + \Sigma_0^{-1}\right)^{-1}\left(\sum_{k=1}^{n}\Sigma^{-1}x_k + \Sigma_0^{-1}m_0\right)$

即均值 μ 的 *MAP* 估计为：$\hat{\mu} = \left(n\Sigma^{-1} + \Sigma_0^{-1}\right)^{-1}\left(\sum_{k=1}^{n}\Sigma^{-1}x_k + \Sigma_0^{-1}m_0\right)$

(2)

经过线性变换后，$\mu \sim N\left(Am_0, A\Sigma_0A^\top\right)$,$x \sim N\left(A\mu, A\Sigma A^\top\right)$

似然度函数：

$$\begin{aligned}
\ln(l(\mu)p(\mu)) &= -\frac{n}{2}\ln\left((2\pi)^d\left|A\Sigma A^{\top}\right|\right)-\frac{1}{2}\sum_{k=1}^{n}(A(x_k-\mu))^{\top}\left(A\Sigma A^{\top}\right)^{-1}\left(A\left(x_k-\mu\right)\right)\\
&\quad-\frac{1}{2}\ln\left((2\pi)^d\left|A\Sigma_0A^{\top}\right|\right)-\frac{1}{2}\left(A\left(\mu-m_0\right)\right)^{\top}\left(A\Sigma_0A^{\top}\right)^{-1}\left(A\left(\mu-m_0\right)\right)\\
&= K-\frac{1}{2}\sum_{k=1}^{n}\left(x_k-\mu\right)^{\top}A^{\top}\left(A^{T}\right)^{-1}\Sigma^{-1}A^{-1}A(x_k-\mu)\\
&\quad-\frac{1}{2}\left(\mu-m_0\right)^{\top}A^{\top}\left(A^{\top}\right)^{-1}\Sigma_0^{-1}A^{-1}A\left(\mu-m_0\right)\\
&= K-\frac{1}{2}\sum_{k=1}^{n}\left(x_k-\mu\right)^{\top}\Sigma^{-1}(x_k-\mu)-\frac{1}{2}\left(\mu-m_0\right)^{\top}\Sigma_0^{-1}\left(\mu-m_0\right)
\end{aligned}$$

求偏导得：$\frac{\partial\ln(l(\mu)p(\mu))}{\partial\mu}=\sum_{k=1}^{n}\Sigma^{-1}\left(x_k-\mu\right)+\Sigma_0^{-1}\left(m_0-\mu\right)$

令偏导等于 *0* 得：$\hat{\mu}=\left(n\Sigma^{-1}+\Sigma_0^{-1}\right)^{-1}\left(\sum_{k=1}^{n}\Sigma^{-1}x_k+\Sigma_0^{-1}m_0\right)$

可得 μ 经过线性变换后，*MAP* 估计值与未变换的估计值相同，即经过线性变换后，*MAP* 仍能正确估计 μ

问题 3.

(1)

$$\begin{aligned}
Q\left(\theta;\theta^0\right) &= \ln P\left(x_1\mid\theta\right)+\ln\left(x_2\mid\theta\right)+\int_{-\infty}^{+\infty}\ln p\left(x_3\mid\theta\right)\frac{P\left(2,x_{32}\mid\theta^0\right)}{\int_{-\infty}^{+\infty}p\left(2,x_{32}\mid\theta^0\right)dx_{32}}dx_{32}\\
&= -2\ln\theta_1-2\ln\theta_2-4\theta_1+\int_{-\infty}^{+\infty}\ln p\left(x_3\mid\theta\right)\frac{P\left(2,x_{32}\mid\theta^0\right)}{\frac{1}{\theta_1}e^{-2\theta_1}}dx_{32}\\
&= -2\ln\theta_1-2\ln\theta_2-4\theta_1+\int\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\cdot\frac{P\left(2,x_{32}\mid\theta^0\right)}{\frac{1}{\theta_1}e^{-2\theta_1}}dx_{32}
\end{aligned}$$

当 $\theta_2\geqslant 4$ 时，$\int\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\frac{P(2,x_{32}|\theta^0)}{\frac{1}{\theta_1}e^{-2\theta_1}}dx_{32}=\int_0^4\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\frac{1}{4}dx_{32}=-\ln\theta_1-\ln\theta_2-2\theta_1$

当 $3\leq\theta_2<4$ 时，$\int\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\cdot\frac{P(2,x_{32}|\theta^0)}{\frac{1}{\theta_1}e^{-2\theta_1}}dx_{32}=\int_0^{\theta_2}\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\cdot\frac{1}{4}dx_{32}=\frac{\theta_2}{4}\left(-\ln\theta_1-\ln\theta_2-2\theta_1\right)$

当 $\theta_2<3$ 时，$\int\ln\frac{1}{\theta_1}e^{-2\theta_1}\frac{1}{\theta_2}\cdot\frac{P(2,x_{32}|\theta^0)}{\frac{1}{\theta_1}e^{-2\theta_1}}dx_{32}=NULL$

综上所述，$Q\left(\theta;\theta^0\right)=\begin{cases}-3\ln\theta_1-3\ln\theta_2-6\theta_1 & 4\geqslant\theta_2\\ -2\ln\theta_1-2\ln\theta_2-4\theta_1-\frac{\theta_1\theta_2}{2}-\frac{\theta_2\ln\theta_1}{4}-\frac{\theta_2\ln\theta_2}{4} & 3\leqslant\theta_2<4\\ NULL & else\end{cases}$

同时由概率密度函数积分为 *1*，且 $\theta_1>0$：

$\int_0^{+\infty}\frac{1}{\theta_1}e^{-x\theta_1}dx=-\frac{1}{\theta_1^2}e^{-x\theta_1}\Big|_0^{+\infty}=\frac{1}{\theta_1^2}=1\Rightarrow\theta_1=1$

于是: $Q\left(\theta;\theta^{0}\right)=\begin{cases}-3\ln\theta_2-6 & 4\geqslant\theta_2\\ -2\ln\theta_2-4-\frac{\theta_2}{2}-\frac{\theta_2\ln\theta_2}{4} & 3\leqslant\theta_2<4\\ NULL & else\end{cases}$

(2)

当 $\theta_2\geqslant 4$ 时，$\frac{\partial Q}{\partial\theta_2}<0$，$Q$ 随着 θ_2 的增大而减小，于是此时 $\hat{\theta}_2=4, Q=-10.158883$

当 $3\leq\theta_2<4$ 时，$\frac{\partial\theta}{\partial\theta_2}=-\frac{2}{\theta_2}-\frac{1}{2}-\frac{\ln\theta_2+1}{4}<0$，此时 $\hat{\theta}_2=3, Q=-8.521183$

综上所述，$\hat{\theta}_2=3$

问题 4.

前向过程:

$\alpha_i(1)=\pi_i b_i\left(O_1\right)$

$\alpha_i(t+1)=b_i\left(O_{t+1}\right)\sum_{j=1}^{C}\alpha_j(t)a_{ji}$

式中 i 表示时刻为 t 时的状态，O_t 表示 t 时刻的观测值。$\alpha_i(t)$ 表示 t 时刻状态为 i，观测序列为 $O_1,O_2,......,O_t$ 的概率。通过上式可得，得到所有的 $\alpha_i(t)$ 复杂度为 $O(TC^2)$

反向过程:

$\beta_i(T)=1$

$\beta_i(t)=\sum_{j=1}^{C}\beta_j(t+1)a_{ij}b_j\left(O_{t+1}\right)$

$\beta_i(t)$ 表示 t 时刻下状态为 i，剩余观测序列为 $O_{t+1},O_{t+2},......,O_T$ 的概率。通过上式可得，得到所有的 $\beta_i(t)$ 复杂度为 $O(TC^2)$

$\gamma_i(t)$ 表示给定观测序列 O，在 t 时刻状态为 i 的概率，如下所示

$\gamma_i(t)=P\left(X_t=i\mid O,\theta\right)=\frac{P(X_t=i,O|\theta)}{P(O|\theta)}=\frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{C}\alpha_j(t)\beta_j(t)}$

$\xi_{ij}(t)$ 表示给定观测序列 O，在 t 时刻状态为 i，*t+1* 时刻状态为 j 的概率，如下所示

$\xi_{ij}(t)=P\left(X_t=i,X_{t+1}=j\mid O,\theta\right)=\frac{P(X_t=t,X_{t+1}=j,O|\theta)}{P(O|\theta)}=\frac{\alpha_i(t)a_{ij}\beta_j(t+1)b_j(O_t+1)}{\sum_{j=1}^{C}\alpha_j(t)\beta_j(t)}$

定义公式 $1_{O_t=v_k}=\begin{cases}1\ if\ O_t=v_k\\ 0\ otherwise\end{cases}$

则更新公式为：$a_{ij}^{*}=\frac{\sum_{t-1}^{T-1}\xi_{ij}(t)}{\sum_{t=1}^{T-1}\gamma_i(t)}$，$b_i^{*}\left(v_k\right)=\frac{\sum_{t=1}^{T}1_{O_t=v_k}\gamma_i(t)}{\sum_{t=1}^{T}\gamma_i(t)}$

计算 $\alpha_i(t)$、$\beta_i(t)$ 复杂度均为 $O(TC^2)$，更新时 a_{ij}^* 与 $b_i^*(v_k)$ 复杂度均为 $O(TC^2)$. 综上所述，更新一次全部的参数复杂度为 $O(4TC^2)$

问题 5.

(1)

$$
\begin{aligned}
\bar{p}_n(x) &= E\left(P_n(x)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{h_n}\varphi\left(\frac{x-x_i}{h_n}\right)\right) \\
&= \int \frac{1}{h_n}\varphi\left(\frac{x-v}{h_n}\right)p(v)dv \\
&= \int \frac{1}{h_n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{v-u}{\sigma}\right)^2}dv \\
&= \int \frac{1}{2\pi\sigma h_n}e^{-\frac{1}{2}\left(\frac{v^2-2uv+u^2}{\sigma^2}+\frac{x^2-2vx+v^2}{h_n^2}\right)}dv \\
&= \int \frac{1}{2\pi\sigma h_n}e^{-\frac{1}{2}\left(\frac{(\sigma^2+h_n^2)v^2}{\sigma^2+h_n^2}-\frac{2uh_n^2+2x\sigma^2}{\sigma^2h_n^2}v+\frac{u^2h_n^2+\sigma^2x^2}{\sigma^2h_n^2}\right)}dv \\
&= e^{-\frac{1}{2}\left(\frac{u^2h_n^2+\sigma^2x^2}{\sigma^2h_n^2}\right)}\cdot\frac{1}{2\pi\sigma h_n}\int e^{-\frac{1}{2}\left(\frac{\sigma^2+h_n^2}{\sigma^2h_n^2}v^2-\frac{2uh_n^2+2x\sigma^2}{\sigma^2hn^2}v\right)}dv
\end{aligned}
$$

记 $a^2 = \frac{\sigma^2h_n^2}{\sigma^2+h_n^2} \quad b = \frac{uh_n^2+x\sigma^2}{\sigma^2h_n^2}\cdot\frac{\sigma^2h_n^2}{\sigma^2+h_n^2}$

$$
\begin{aligned}
&\int e^{-\frac{1}{2}}\left(\frac{1}{a^2}v^2-2\frac{1}{a^2}bv\right)dv \\
=& \int e^{-\frac{1}{2}\frac{1}{a^2}(v^2-2bv)}dv \\
=& \int e^{-\frac{1}{2}\left(\frac{v-b}{a}\right)^2}\cdot e^{\frac{b^2}{2a^2}}dv \\
=& \quad e^{\frac{b^2}{2a^2}}\sqrt{2\pi}a
\end{aligned}
$$

$$
\begin{aligned}
\bar{P}_n(x) &= \frac{\sqrt{2\pi}\sigma h_n}{\sqrt{\sigma^2+h_n^2}}\cdot\frac{1}{2\pi\sigma h_n}e^{-\frac{1}{2}\left(\frac{x^2}{h_n^2}-\frac{\left(\frac{(x\sigma^2+uh_n^2)^2}{\sigma^4h_n^4}\right)\cdot\frac{\sigma^4h_n^4}{(\sigma^2+h_n^2)^2}}{\frac{\sigma^2h_n^2}{\sigma^2+h_n^2}}+\frac{u^2}{\sigma^2}\right)} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+h_n^2}}\cdot e^{\frac{1}{2}\frac{(x-u)^2}{\sigma^2+h_n^2}}
\end{aligned}
$$

即：$\bar{P}_n(x) \sim N\left(\mu, \sigma^2+{h_n}^2\right)$

原式得证

(2)

$$
\begin{aligned}
\operatorname{Var}(p(x)) &= \sum_{i=1}^{n} E\left[\left(\frac{1}{nh_n}\varphi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_n}\right)-\frac{1}{n}\bar{p}_n(\mathbf{x})\right)^2\right] \\
&= nE\left[\frac{1}{n^2h_n^2}\varphi^2\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_n}\right)\right]-\frac{1}{n}\bar{p}_n^2(\mathbf{x}) \\
&= \frac{1}{nh_n^2}\int\varphi^2\left(\frac{\mathbf{x}-\mathbf{v}}{h_n}\right)p(\mathbf{v})d\mathbf{v}-\frac{1}{n}\bar{p}_n^2(\mathbf{x}) \\
&\approx \frac{1}{nh_n^2}\int\varphi^2\left(\frac{x-v}{h_n}\right)p(v)dv
\end{aligned}
$$

$$
\begin{aligned}
&\int p^2\left(\frac{x-v}{hn}\right)p(v)dv \\
&= \int\frac{1}{2\pi}e^{-\left(\frac{x-v}{h_n}\right)^2}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{v-u}{\sigma}\right)^2}dv \\
&= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma}\int e^{-\left(\frac{v^2}{2\sigma^2}+\frac{v^2}{h_n^2}-\frac{uv}{\sigma^2}-\frac{2xv}{h_n^2}+\frac{x^2}{h_n^2}+\frac{u^2}{2\sigma^2}\right)}dv \\
&= \frac{\sqrt{\frac{\sigma^2h_n^2}{h_n^2+2\sigma^2}}}{2\pi\sigma}e^{-\frac{1}{2}\frac{(x-u)^2}{\frac{h_n^2}{2}+\sigma^2}}\int\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma^2h_n^2}{h_n^2+2\sigma^2}}}e^{-\frac{1}{2}\left(\frac{v-\frac{uh_n^2+2x\sigma^2}{h_n^2+2\sigma^2}}{\sqrt{\frac{\sigma^2h_n^2}{h_n^2+2\sigma^2}}}\right)^2}dv \\
&= \frac{\sqrt{\frac{\sigma^2h_n^2}{h_n^2+2\sigma^2}}}{2\pi\sigma}e^{-\frac{1}{2}\frac{(x-u)^2}{\frac{h_n^2}{2}+\sigma^2}}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{nh_n^2}\int\varphi^2\left(\frac{x-v}{h_n}\right)p(v)dv \\
&= \frac{1}{2nh_n\sqrt{\pi}}\cdot\frac{1}{\sqrt{2\pi}\sqrt{\frac{h_n^2}{2}+\sigma^2}}e^{-\frac{1}{2}\frac{(x-u)^2}{\frac{h_n^2}{2}+\sigma^2}} \\
&\approx \frac{1}{2nh_n\sqrt{\pi}}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x-u)^2}{\sigma^2}} \\
&= \frac{1}{2nh_n\sqrt{\pi}}p(x)
\end{aligned}
$$

即： $\operatorname{Var}(p(x)) \approx \frac{1}{2nh_n\sqrt{\pi}}p(x)$

原式得证