第一题

(a)

因为样本 $\mathcal{D} = x_1, \ldots, x_n$ 为独立同分布的，所以有：

$$p(\mathcal{D}|\theta) = \prod_{k=1}^{n} p(x_k|\theta) = \begin{cases} (\frac{1}{\theta})^n, & \forall x \in \mathcal{D}, 0 \leq x \leq \theta \\ 0, & otherwise \end{cases} \tag{1}$$

所以要想使得 $p(\mathcal{D}|\theta)$ 最大，则需要尽可能减小 θ 的值，而当 θ 小于集合 $\mathcal{D}$ 中的某个元素时，概率密度为 0，所以 θ 的最大似然估计为 $\mathcal{D}$ 中的最大值点 $max[\mathcal{D}]$。

(b)

由 (a) 可知，当 $\theta \leq max[\mathcal{D}]$ 时，概率为 0，其余 4 个点均小于最大值，对于 θ 的估计没有价值，所以无需考虑。且概率密度分布函数为：

$$p(\mathcal{D}|\theta) = \prod_{k=1}^{5} p(x_k|\theta) = \begin{cases} (\frac{1}{\theta})^5, & 0.6 \leq \theta \\ 0, & otherwise \end{cases} \tag{2}$$

函数图像如图 1所示。

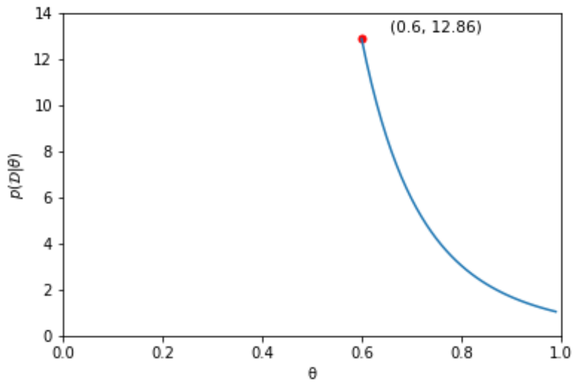

图 1: 似然函数 $p(\mathcal{D}|\theta)$ 图像

第二题

(a)

假设样本集为 $\mathcal{D}$，容量为 n，每一个样本点有 d 维特征。根据题目 $\boldsymbol{\mu}$ 的先验概率密度为：

$$p(\boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}|^{\frac{1}{2}}} exp[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{m_0})^t|\boldsymbol{\Sigma_0}|^{-1}(\boldsymbol{\mu} - \boldsymbol{m_0})] \tag{3}$$

在样本 $\boldsymbol{x} \in \mathcal{D}$ 下的条件概率密度为：

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x} - \boldsymbol{\mu})] \tag{4}$$

MAP 估计为：

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_{MAP} &= argmax_{\boldsymbol{\mu}}\, p(\boldsymbol{\mu})p(\S|\boldsymbol{\mu}) \\
&= argmax_{\boldsymbol{\mu}}\, L(\boldsymbol{\mu}|\mathcal{D}) + ln(p(\boldsymbol{\mu})) \\
&= argmax_{\boldsymbol{\mu}} \sum_{\boldsymbol{x}\in\mathcal{D}} lnp(\boldsymbol{x}|\boldsymbol{\mu}) + ln(p(\boldsymbol{\mu})) \\
&= argmax_{\boldsymbol{\mu}}\, n \times ln(\frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}) + \sum_{x\in\mathcal{D}}(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x} - \boldsymbol{\mu})) \\
&+ ln(\frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}|^{\frac{1}{2}}}) + (-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{m_0})^t|\boldsymbol{\Sigma_0}|^{-1}(\boldsymbol{\mu} - \boldsymbol{m_0})) \\
&= argmax_{\boldsymbol{\mu}}\, Z + \sum_{\boldsymbol{x}\in\mathcal{D}}(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x} - \boldsymbol{\mu})) - \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{m_0})^t|\boldsymbol{\Sigma_0}|^{-1}(\boldsymbol{\mu} - \boldsymbol{m_0})
\end{aligned} \tag{5}$$

其中 Z 为常数项，与 $\boldsymbol{\mu}$ 无关，只需要将式 (5) 中的第 2、3 项之和最大即可。

(b)

对于线性变换 $\boldsymbol{x}^{'} = \boldsymbol{A}\boldsymbol{x}$，设其均值和方差分别为 $\boldsymbol{\mu}^{'}$ 和 $\boldsymbol{\Sigma}^{'}$，则有：

$$\boldsymbol{\mu}^{'} = \varepsilon(\boldsymbol{x}^{'}) = \varepsilon(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\mu} \tag{6}$$

$$\begin{aligned}
\boldsymbol{\Sigma}^{'} &= \varepsilon[(\boldsymbol{x}^{'}-\boldsymbol{\mu}^{'})(\boldsymbol{x}^{'}-\boldsymbol{\mu}^{'})^t] \\
&= \varepsilon[(\boldsymbol{Ax}-\boldsymbol{A\mu})(\boldsymbol{Ax}-\boldsymbol{A\mu})^t] \\
&= \varepsilon\{\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})[\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})]^t\} \\
&= \varepsilon[\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{A}^t] \\
&= \boldsymbol{A\Sigma A}^t
\end{aligned} \tag{7}$$

由式 (3)、(6)、(7) 可得 $\boldsymbol{\mu}^{'}$ 的先验概率密度为:

$$\begin{aligned}
p(\boldsymbol{\mu}^{'}) &= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{\mu}^{'}-\boldsymbol{m_0}^{'})^t|\boldsymbol{\Sigma_0}^{'}|^{-1}(\boldsymbol{\mu}^{'}-\boldsymbol{m_0}^{'})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{A\mu}-\boldsymbol{Am_0})^t|\boldsymbol{A\Sigma_0A}^t|^{-1}(\boldsymbol{A\mu}-\boldsymbol{Am_0})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{m_0})^t\boldsymbol{A}^t(\boldsymbol{A}^t)^{-1}|\boldsymbol{\Sigma_0}|^{-1}\boldsymbol{A}^{-1}\boldsymbol{A}(\boldsymbol{\mu}-\boldsymbol{m_0})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma_0}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{m_0})^t|\boldsymbol{\Sigma_0}|^{-1}(\boldsymbol{\mu}-\boldsymbol{m_0})]
\end{aligned} \tag{8}$$

同理，得在样本 $\boldsymbol{x}^{'}\in\mathcal{D}^{'}$ 下的条件概率密度为:

$$\begin{aligned}
p(\boldsymbol{x}^{'}|\boldsymbol{\mu}^{'}) &= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{x}^{'}-\boldsymbol{\mu}^{'})^t|\boldsymbol{\Sigma}^{'}|^{-1}(\boldsymbol{x}^{'}-\boldsymbol{\mu}^{'})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{Ax}-\boldsymbol{A\mu})^t|\boldsymbol{A\Sigma A}^t|^{-1}(\boldsymbol{Ax}-\boldsymbol{A\mu})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{A}^t(\boldsymbol{A}^t)^{-1}|\boldsymbol{\Sigma}|^{-1}\boldsymbol{A}^{-1}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})] \\
&= \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}^{'}|^{\frac{1}{2}}}exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]
\end{aligned} \tag{9}$$

MAP 估计为:

$$\begin{aligned}
\hat{\boldsymbol{\mu}^{'}}_{MAP} &= argmax_{\boldsymbol{\mu}^{'}}\ p(\boldsymbol{\mu}^{'})p(\boldsymbol{x}^{'}|\boldsymbol{\mu}^{'}) \\
&= argmax_{\boldsymbol{\mu}^{'}}\ L(\boldsymbol{\mu}^{'}|\mathcal{D}^{'})+ln(p(\boldsymbol{\mu}^{'})) \\
&= argmax_{\boldsymbol{\mu}}\ Z^{'}+\sum_{\boldsymbol{x}\in\mathcal{D}^{'}}(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{m_0})^t|\boldsymbol{\Sigma_0}|^{-1}(\boldsymbol{\mu}-\boldsymbol{m_0})
\end{aligned} \tag{10}$$

其中 Z' 为常数项，与 $\boldsymbol{\mu}'$ 无关，只需要将式 (10) 中的第 2、3 项之和最大即可。我们发现，这个优化目标与估计 $\boldsymbol{\mu}$ 的目标是相同的，所以 MAP 能够对变换后的 $\boldsymbol{\mu}'$ 做出正确的估计。

第三题

(a)

$$
\begin{aligned}
Q(\boldsymbol{\theta};\boldsymbol{\theta^0}) &= \varepsilon_{32}[lnp(\boldsymbol{x_g},\boldsymbol{x_b};\boldsymbol{\theta})|\boldsymbol{\theta^0};\mathcal{D}_g] \\
&= \int_{-\infty}^{\infty}[\sum_{k=1}^{2} lnp(\boldsymbol{x_k}|\boldsymbol{\theta}) + lnp(\boldsymbol{x_3}|\boldsymbol{\theta})]p(x_{32}|\boldsymbol{\theta^0};\ x_{31}=2)dx_{32} \\
&= \sum_{k=1}^{2} lnp(\boldsymbol{x_k}|\boldsymbol{\theta}) + \int_{-\infty}^{\infty} ln[p(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix}|\boldsymbol{\theta})\frac{p(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix}|\boldsymbol{\theta^0})}{\int_{-\infty}^{\infty} p(\begin{pmatrix} 2 \\ x_{32}^{'} \end{pmatrix}|\boldsymbol{\theta^0})dx_{32'}}]\,dx_{32}
\end{aligned} \tag{11}
$$

其中

$$
\begin{aligned}
\int_{-\infty}^{\infty} p(\begin{pmatrix} 2 \\ x_{32}^{'} \end{pmatrix}|\boldsymbol{\theta^0})dx_{32'} &= \int_{0}^{\theta_2^0} \frac{1}{\theta_1^0}e^{-\frac{2}{\theta_1^0}}\cdot\frac{1}{\theta_2^0}dx_{32} \\
&= \int_0^4 \frac{1}{2}\cdot e^{-1}\cdot\frac{1}{4}dx_{32} \\
&= (2e)^{-1}
\end{aligned}
$$

所以

$$
\begin{aligned}
\text{式 (11)} &= \sum_{k=1}^{2} lnp(\boldsymbol{x_k}|\boldsymbol{\theta}) + 2e\cdot\int_{-\infty}^{\infty} ln[p(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix}|\boldsymbol{\theta})p(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix}|\boldsymbol{\theta^0})]\,dx_{32} \\
&= lnp(\boldsymbol{x_1}|\boldsymbol{\theta}) + lnp(\boldsymbol{x_2}|\boldsymbol{\theta}) + K \\
&= ln(\frac{e^{-1/\theta_1}}{\theta_1\theta_2}) + ln(\frac{e^{-3/\theta_1}}{\theta_1\theta_2}) + K \\
&= -\frac{4}{\theta_1} - 2ln(\theta_1\theta_2) + K
\end{aligned}
$$

对于 K 有 3 种情况。因为 $x_{22}=3$ 又由于 $p(x_2)\sim\boldsymbol{U}(0,\theta_2)$ 所以当 $\theta_2<3$ 时，$K=0$。又因为 $\theta_2^0=4$，所以如果 $\theta_2>4$，只有 $x_{32}\le 4$ 的时候，项

$$p(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix}|\boldsymbol{\theta^0})$$

才不为 0，所以：

$$Q(\boldsymbol{\theta};\boldsymbol{\theta^0}) = \begin{cases} -\frac{4}{\theta_1} - 2ln(\theta_1\theta_2) - \frac{\theta_2}{4}(ln(\theta_1\theta_2) + \frac{2}{\theta_1}), & 3 \leq \theta_2 < 4 \\ -\frac{6}{\theta_1} - 3ln(\theta_1\theta_2), & \theta_2 \geq 4 \\ -\frac{4}{\theta_1} - 2ln(\theta_1\theta_2), & otherwise \end{cases} \tag{12}$$

(b)

(1) 当 $3 \leq \theta_2 < 4$ 时，Q 为：

$$Q(\boldsymbol{\theta};\boldsymbol{\theta^0}) = -\frac{1}{\theta_1}(4 + \frac{\theta_2}{2}) - ln(\theta_1\theta_2)(2 + \frac{\theta_2}{4})$$

将 Q 对 θ_1 求偏导可得：

$$\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta^0})}{\partial\theta_1} = -\frac{1}{{\theta_1}^2}(4 + \frac{\theta_2}{2} - 2\theta_1 - \frac{\theta_1\theta_2}{4}) \tag{13}$$

令式 (13) 为 0，得：

$$(\theta_1 - 2)(-\frac{\theta_2}{4} - 2) = 0$$

则 $\theta_1 = 2$ 或 $\theta_2 = -8$，又因为 $3 \leq \theta_2 < 4$，所以 $\theta_1 = 2$。将 Q 对 θ_2 求偏导可得：

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta^0})}{\partial\theta_2} &= \frac{1}{4} + \frac{1}{2\theta_1} - \frac{2}{\theta_2} - \frac{1}{4}ln(\theta_1\theta_2) \\ \text{将}\theta_1 = 2\text{代入，则原式} &= \frac{2 - ln\theta_2}{4} - \frac{2}{\theta_2} \end{aligned} \tag{14}$$

当 $3 \leq \theta_2 < 4$ 时，式 (14) 的值小于 0，所以 $Q(\boldsymbol{\theta};\boldsymbol{\theta^0})$ 对于 θ_2 单调递减。当 $\theta_2 = 3$ 时，Q 取得最大值。此时 $Q \simeq -7.677$。

(2) 当 $\theta_2 \geq 4$ 时，Q 为：

$$Q(\boldsymbol{\theta};\boldsymbol{\theta^0}) = -\frac{6}{\theta_1} - 3ln(\theta_1\theta_2)$$

将 Q 对 θ_1 求偏导可得:

$$\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta^0})}{\partial \theta_1} = \frac{6-3\theta_1}{\theta_1^2} \tag{15}$$

令式 (15) 为 0，得 $\theta_1 = 2$。将 Q 对 θ_2 求偏导可得:

$$\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta^0})}{\partial \theta_2} = -\frac{3}{\theta_2} \tag{16}$$

当 $\theta_2 \geq 4$ 时，式 (16) 的值恒小于 0，所以 $Q(\boldsymbol{\theta};\boldsymbol{\theta^0})$ 对于 θ_2 单调递减。当 $\theta_2 = 4$ 时，Q 取得最大值。此时 $Q \simeq -9.238$。其他情况没有意义，所以综上所述，当 $\boldsymbol{\theta} = \{2, 3\}$ 时，Q 最大，最大值约为 -7.677。

第四题

假设第 t 时刻的观测状态值为 v_k，定义从状态 $\omega_i(t-1)$ 转移到状态 $\omega_j(t)$ 的概率 $\gamma_{ij}(t)$ 为：

$$\gamma_{ij}(t) = \frac{\alpha_i(t-1)a_{ij}b_{jk}\beta_j(t)}{P(\boldsymbol{V}^T|\boldsymbol{\theta})} \tag{17}$$

其中，$P(\boldsymbol{V}^T|\boldsymbol{\theta})$ 是模型产生观测序列 $\boldsymbol{V}^T$ 的总概率。$\alpha_i(t-1)$ 为前向概率，表示我们的 HMM 在 $t-1$ 时刻，位于隐状态 ω_1，并已产生观测序列 $\boldsymbol{V}^T$ 的前 $t-1$ 个观测状态的概率。$\beta_j(t)$ 为后向概率，表示我们的 HMM 在 t 时刻，位于隐状态 ω_j，并已产生观测序列 $\boldsymbol{V}^T$ 中的 $[t+1,...,T]$ 观测状态的概率。对 a_{ij} 进行估计，在任何时刻，序列中从状态 $\omega_i(t-1)$ 转移到状态 $\omega_j(t)$ 的预期值为 $\sum_{t=1}^T \gamma_{ij}(t)$。而从 ω_i 转移到任何隐状态的总预期数为 $\sum_{t=1}^T \sum_{k=1}^c \gamma_{ik}(t)$。所以，

$$\begin{aligned}\hat{a}_{ij} &= \frac{\sum_{t=1}^T \gamma_{ij}(t)}{\sum_{t=1}^T \sum_{k=1}^c \gamma_{ik}(t)} \\ &= \frac{\sum_{t=1}^T \alpha_i(t-1)a_{ij}b_{jk}\beta_j(t)}{\sum_{t=1}^T \alpha_i(t-1)\beta_i(t-1)}\end{aligned} \tag{18}$$

对于某个确定的 i，计算前向概率 $\alpha_i(t)$，需要考虑 $t-1$ 时刻的所有可能的 c 种隐状态的前向概率加权求和，即复杂度为 $\mathcal{O}(c)$。式 (18) 中对 t 求和，且对于每一个 i 都要计算一次，则 $\hat{a}_{ij}$ 中计算前向概率的总体复杂度为 $\mathcal{O}(c^2T)$。又由于后向概率计算复杂度与前向概率相同全部，所以全部更新一次 $\hat{a}_{ij}$ 的计算复杂度是 $\mathcal{O}(c^2T)$。

b_{jk} 为在隐状态 ω_i 的情况下,发出可见状态 v_k 的概率。在任意处于隐状态 ω_j 的时刻,序列中发出观测状态的预计值为 $\sum\limits_{v(t)=v_k,t=1}^{T} \sum_{l=0}^c \gamma_{jl}(t)$。而隐状态 ω_j 发出所有可见状态的总预期数为 $\sum_{t=1}^T \sum_{l=0}^c \gamma_{jl}(t)$。所

以，

$$
\begin{aligned}
\hat{b}_{jk} &= \frac{\sum\limits_{v(t)=v_k,t=1}^{T} \sum_{l=0}^{c} \gamma_{jl}(t)}{\sum_{t=1}^{T} \sum_{l=0}^{c} \gamma_{jl}(t)} \\
&= \frac{\sum_{t=1}^{T} \alpha_j(t)\beta_j(t)\delta(t, v_k)}{\sum_{t=1}^{T} \alpha_j(t)\beta_j(t)}
\end{aligned}
\tag{19}
$$

其中，$\delta(t, v_k)$ 是关于 t 和 v_k 的冲击函数，当在 t 时刻观测状态为 v_k 时，函数值为 1，否则为 0。同上，全部更新一次 $\hat{b}_{jk}$ 的计算复杂度是 $\mathcal{O}(c^2T)$。

第五题

根据题意，得

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \tag{20}$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} exp[-\frac{x^2}{2}] \tag{21}$$

(a)

$$\begin{aligned}
\bar{p_n}(x) &= \varepsilon[p_n(x)] \\
&= \frac{1}{nh_n}\sum_{i=1}^{n}\varepsilon[\varphi(\frac{x-x_i}{h_n})] \\
&= \frac{1}{h_n}\int_{-\infty}^{\infty}\varphi(\frac{x-v}{h_n})p(v)\,dv \\
&= \frac{1}{2\pi h_n\sigma}\int_{-\infty}^{\infty} exp[-\frac{1}{2}(\frac{x-v}{h_n}^2) - \frac{1}{2}(\frac{v-\mu}{\sigma}^2)]\,dv \\
&= Z\cdot\int_{-\infty}^{\infty} exp\{-\frac{1}{2}[v^2(\frac{1}{{h_n}^2}+\frac{1}{\sigma^2}) - 2v(\frac{x}{{h_n}^2}+\frac{\mu}{\sigma^2})]\}\,dv \\
&= Z^{'}\cdot\int_{-\infty}^{\infty} exp[-\frac{1}{2}(v\sqrt{\frac{1}{{h_n}^2}+\frac{1}{\sigma^2}} - \frac{\frac{x}{{h_n}^2}+\frac{\mu}{\sigma^2}}{\sqrt{\frac{1}{{h_n}^2}+\frac{1}{\sigma^2}}})^2]\,dv \\
&= Z^{'}\cdot\int_{-\infty}^{\infty} exp[-\frac{1}{2}(\frac{v-\alpha}{\theta})^2]\,dv
\end{aligned} \tag{22}$$

其中 $Z^{'}$ 是与 v 无关的项，且:

$$\theta = \frac{1}{\sqrt{\frac{1}{{h_n}^2}+\frac{1}{\sigma^2}}}$$

$$\alpha = \theta^2\cdot(\frac{x}{{h_n}^2}+\frac{\mu}{\sigma^2})$$

$$Z^{'} = \frac{1}{2\pi h_n\sigma} exp[-\frac{1}{2}(\frac{x^2}{{h_n}^2}+\frac{\mu^2}{\sigma^2}) - \frac{\alpha^2}{\theta^2})] \tag{23}$$

换元解得积分项为:

$$\int_{-\infty}^{\infty} exp[-\frac{1}{2}(\frac{v-\alpha}{\theta})^2]\, dv = \theta \cdot \int_{-\infty}^{\infty} exp[-\frac{1}{2}(\frac{v-\alpha}{\theta})^2]\, d(\frac{v-\alpha}{\theta}) \\ = \theta \cdot \sqrt{2\pi} \tag{24}$$

将式 (23)、(24) 代入式 (22) 得,

$$\bar{p_n}(x) = \frac{\theta}{\sqrt{2\pi}h_n\sigma} exp[-\frac{1}{2}(\frac{x^2}{{h_n}^2} + \frac{\mu^2}{\sigma^2} - \frac{\alpha^2}{\theta^2})] \tag{25}$$

其中,

$$\begin{aligned} &\frac{x^2}{{h_n}^2} + \frac{\mu^2}{\sigma^2} - \frac{\alpha^2}{\theta^2} \\ &= \frac{x^2}{{h_n}^2} + \frac{\mu^2}{\sigma^2} - \frac{\frac{x^2\sigma^2}{{h_n}^2} + \frac{\mu^2{h_n}^2}{\sigma^2} + 2x\mu}{{h_n}^2 + \sigma^2} \\ &= \frac{x^2({h_n}^2 + \sigma^2)}{{h_n}^2({h_n}^2 + \sigma^2)} - \frac{x^2\sigma^2 + \frac{\mu^2{h_n}^4}{\sigma^2} + 2x\mu{h_n}^2}{({h_n}^2 + \sigma^2){h_n}^2} + \frac{\mu^2({h_n}^2 + \sigma^2)}{\sigma^2({h_n}^2 + \sigma^2)} \\ &= \frac{x^2 + 2x\mu - \mu^2}{{h_n}^2 + \sigma^2} \\ &= \frac{(x-\mu)^2}{{h_n}^2 + \sigma^2} \end{aligned} \tag{26}$$

将式 (26)、(23) 代入式 (25) 得,

$$\bar{p_n}(x) = \sqrt{\frac{1}{2\pi({h_n}^2 + \sigma^2)}} exp[-\frac{1}{2} \cdot \frac{(x-\mu)^2}{({h_n}^2 + \sigma^2)}] \tag{27}$$

所以 $\bar{p_n}(x) \sim N(\mu, \sigma^2 + {h_n}^2)$。

(b)

根据方差公式，

$$\begin{aligned} Var[p_n(x)] &= Var[\frac{1}{nh_n}\sum_{i=1}^{n}\varepsilon[\varphi(\frac{x-x_i}{h_n})]] \\ &= \frac{1}{n{h_n}^2}Var[\varepsilon[\varphi(\frac{x-x_i}{h_n})]] \\ &= \frac{1}{n{h_n}^2}\cdot\{\varepsilon[\varphi(\frac{x-x_i}{h_n})^2]-(\varepsilon[\varphi(\frac{x-x_i}{h_n})])^2\} \end{aligned} \tag{28}$$

其中，

$$\begin{aligned} \varepsilon[\varphi(\frac{x-x^{'}}{h_n})^2] &= \int_{-\infty}^{\infty}\varphi(\frac{x-x^{'}}{h_n})^2 p(x^{'})\,dx^{'} \\ &= \int_{-\infty}^{\infty}\frac{1}{2\pi}exp[(-\frac{x-x^{'}}{h_n})^2]\cdot\frac{1}{\sqrt{2\pi}\sigma}exp[-\frac{1}{2}(\frac{x^{'}-\mu}{h_n})^2]\,dx^{'} \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-x^{'}}{h_n/\sqrt{2}}\right)^2\right]exp\left[-\frac{1}{2}\left(\frac{x^{'}-\mu}{\sigma}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma}\,dx^{'} \\ &= \frac{h_n/\sqrt{2}}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right] \end{aligned} \tag{29}$$

结合式 (28)，则有，

$$\begin{aligned} \frac{1}{n{h_n}^2}\cdot\varepsilon[\varphi(\frac{x-x_i}{h_n})^2] &= \frac{1}{nh_n^2}\frac{h_n/\sqrt{2}}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right] \\ &= \frac{1}{nh_n}\frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2/2+\sigma^2}\right] \end{aligned} \tag{30}$$

由窗函数性质，当 n 趋近于无穷大时，$\sqrt{h_n^2/2}+\sigma^2\simeq\sigma^2$，所以，

$$\frac{1}{n{h_n}^2}\cdot\varepsilon[\varphi(\frac{x-x_i}{h_n})^2]\simeq\frac{1}{2nh_n\sqrt{\pi}}p(x) \tag{31}$$

同理计算均值的平方，

$$
\begin{aligned}
\frac{1}{n{h_n}^2}\cdot(\varepsilon[\varphi(\frac{x-x_i}{h_n})])^2 &= \frac{1}{nh_n^2}\left[\int\varphi\left(\frac{x-x'}{h_n}\right)p\left(x'\right)dx'\right]^2 \\
&= \frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right] \\
\text{同理，}\quad &\simeq \frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \\
&\simeq 0
\end{aligned}
\tag{32}
$$

证毕。

第六题

(a)

假设这些样本点服从均匀分布，且设超立方体的体积为 V，则每个样本点的邻域可以看成是一个体积为 V/n 的小立方体，该小立方体的边长为 $(V/n)^{\frac{1}{d}}$。所以当使用题目中的部分距离表示，投影到 r 维空间时，小超立方体体积变为 $(V/n)^{\frac{r}{d}}$。所以在 r 维空间中，某样本点落入体积 V 的概率为：

$$P\left(x = x^{r'}\right) = (1/n)^{r/d} \tag{33}$$

所以错分概率为：

$$p\left(x = x^{d'} \mid x = x^{r'}\right) = \frac{(1/n)}{(1/n)^{\frac{r}{d}}} = (1/n)^{1-\frac{r}{d}} = (1/n)^{1-\frac{r}{10}} \tag{34}$$

当 $n = 10$ 时函数图像如下：

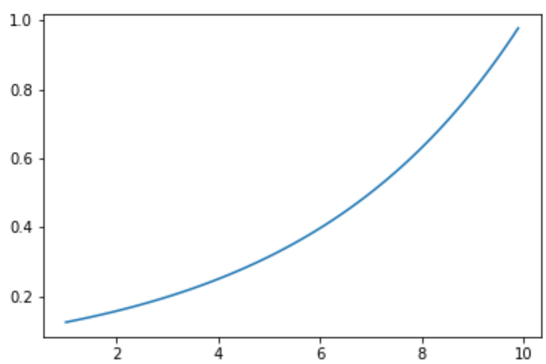

图 2: n=10

当 $n = 100$ 时函数图像如下：

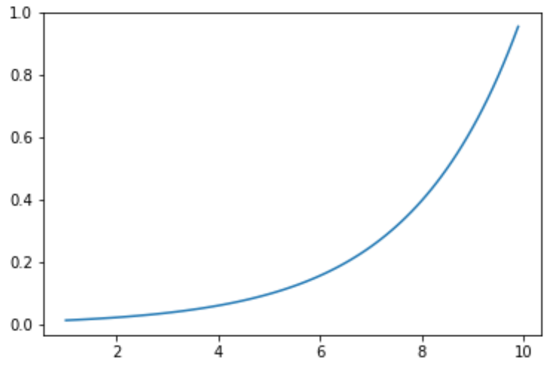

图 3: n=100

当 $n = 1000$ 时函数图像如下:

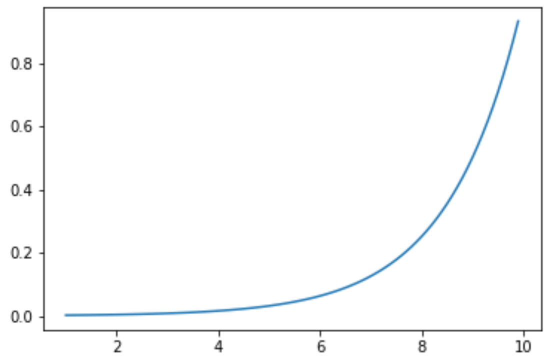

图 4: n=1000

当 $n = 10000$ 时函数图像如下:

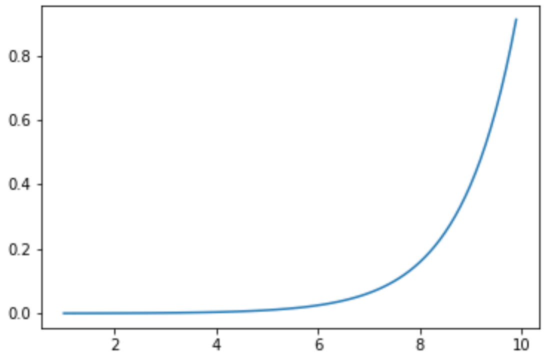

图 5: n=10000

可以看到，随着 n 的增大，估计会越来越准确。

(b)

线性斜坡函数定义为：

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{35}$$

显然在一维特征下的贝叶斯决策面为

$$x^* = 0.5$$