Answer of Assignment 2

2020E8017782032_ 蒲尧

说明

作业用中文撰写，鼓励使用 LaTeX。
文档按 “学号 _ 姓名.pdf” 命名提交。
本次作业截止时间为 2020 年 10 月 27 日，请到课程网站及时提交。

Question 1

Let x have a uniform density

$$p(x \mid \theta) \sim V(0,\theta) = \begin{cases} 1/\theta, & 0 \le x \le \theta; \\ 0, & \text{otherwise} \end{cases}$$

1. Suppose that n samples $\mathbf{D} = \{x_1, ..., x_n\}$ are drawn independently according to $p(x \mid \theta)$. Show that the maximum likelihood estimate for θ is $\max[\mathbf{D}]$, i.e., the value of the maximum element in $\mathbf{D}$.

2. Suppose that $n = 5$ points are drawn from the distribution and the maximum value of which happens to be $\underset{k}{max}\, x_k = 0.6$. Plot the likelihood $p(x \mid \theta)$ in the range $0 \le \theta \le 1$. Explain in words why you do not need to know the values of the other four points.

Answer 1

解：

1. 我们使用指标函数 $I(\bullet)$，当括号内的逻辑为真则值为 1，否则为 0。则可以得到似然函数:

$$\begin{aligned} p(\boldsymbol{D} \mid \theta) &= \prod_{k=1}^{n} p(x_k \mid \theta) \\ &= \prod_{k=1}^{n} \frac{1}{\theta} I(0 \le x_k \le \theta) \\ &= \frac{1}{\theta^n} I\left(\theta \ge \underset{k}{max}\, x_k\right) I\left(\underset{k}{min}\, x_k \ge 0\right) \end{aligned}$$

我们可以看到似然函数随 θ 增大而减小，而 θ 最小值为 $\underset{k}{max}\, x_k$
$\therefore$ 对于 θ 的最大似然估计就是 $\mathbf{D}$ 中的最大值点 $max[\mathbf{D}]$

2.

$$\because \begin{cases} n = 5; \\ 0 \le \theta \le 1; \\ \underset{k}{max}\, x_k = 0.6 \end{cases} \quad \therefore \begin{cases} p(\boldsymbol{D} \mid \theta) = 0, & 0 \le \theta < 0.6; \\ p(\boldsymbol{D} \mid \theta) = \frac{1}{\theta^5}, & 0.6 \le \theta \le 1; \end{cases}$$

图像如下所示

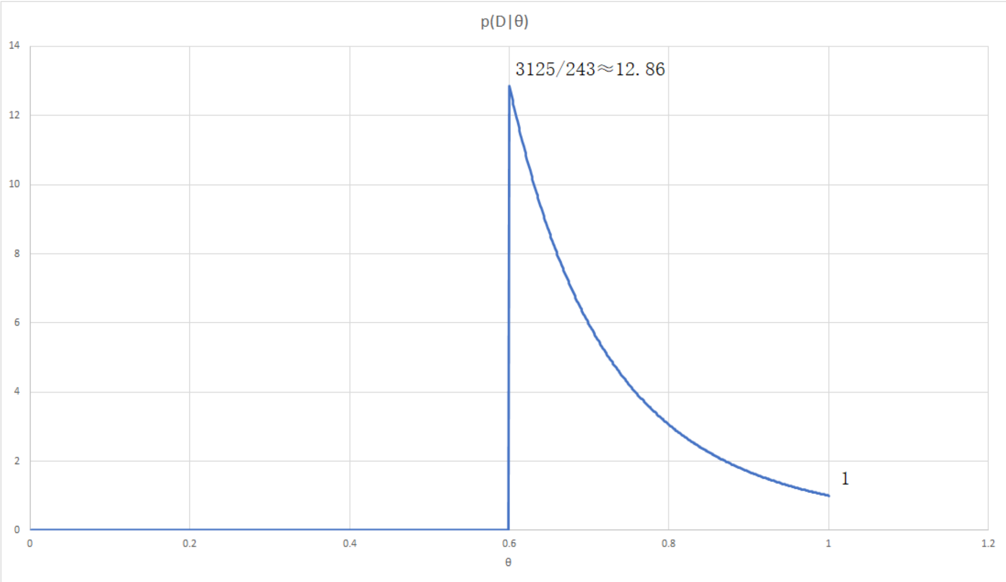

图 1: $p\left(\boldsymbol{D} \mid \theta\right) - \theta$ 图

$\because p\left(\boldsymbol{D} \mid \theta\right)$ 的值只与 $n, \theta, \max\limits_{k} x_k = 0.6$ 有关
$\therefore$ 我们不需要知道其他元素值。

Queestion 2

Assume we have training data from a Gaussian distribution of known covariance $\boldsymbol{\Sigma}$ but unknown mean $\boldsymbol{\mu}$. Suppose further that this mean itself is random, and characterized by a Gaussian density having mean $\boldsymbol{m}_0$ and covariance $\boldsymbol{\Sigma}_0$.

1. What is the MAP estimator for $\boldsymbol{\mu}$?
2. Suppose we transform our coordinates by a linear transform $\boldsymbol{x}' = \boldsymbol{A}\boldsymbol{x}$, for nonsingular matrix $\mathbf{A}$, and accordingly for other terms. Determine whether your MAP estimator gives the appropriate estimate for the transformed mean $\boldsymbol{\mu}'$. Explain.

Answer 2

解：

1. 由题可知：

$$p\left(\boldsymbol{\mu}\right) = \frac{1}{\left(2\pi\right)^{\frac{d}{2}} \left|\boldsymbol{\Sigma}_0\right|^{\frac{1}{2}}} exp[-\frac{1}{2}\left(\boldsymbol{\mu} - \boldsymbol{m}_0\right)^t \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\mu} - \boldsymbol{m}_0\right)]$$

$$
\begin{aligned}
\ln\left[p\left(\boldsymbol{D}\mid\boldsymbol{\mu}\right)\right] &= \ln\prod_{k=1}^{n} p\left(\boldsymbol{x}_k\mid\boldsymbol{\mu}\right)\\
&= \sum_{k=1}^{n}\ln p\left(\boldsymbol{x}_k\mid\boldsymbol{\mu}\right)\\
&= \sum_{k=1}^{n}\ln\{\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)]\}\\
&= -\frac{n}{2}\ln\left[(2\pi)^d\left|\Sigma\right|\right]-\sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)\right]
\end{aligned}
$$

$\therefore$ 均值 $\boldsymbol{\mu}$ 的 MAP 估计是

$$
\begin{aligned}
\hat{\boldsymbol{\mu}} &= arg\max_{\boldsymbol{\mu}}\{\ln\left[p\left(\boldsymbol{D}\mid\boldsymbol{\mu}\right)\right]\times p\left(\boldsymbol{\mu}\right)\}\\
&= arg\max_{\boldsymbol{\mu}}\left\{\left[-\frac{n}{2}\ln\left[(2\pi)^d\left|\boldsymbol{\Sigma}\right|\right]-\sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_k-\boldsymbol{\mu}\right)\right]\right.\right.\\
&\left.\left.\times\left[\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_0\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)^t\boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)]\right]\right\}
\end{aligned}
$$

2. 经过线性变换 $\boldsymbol{x}'=\boldsymbol{A}\boldsymbol{x}$ 之后，训练样本的均值和协方差矩阵变为 $\boldsymbol{\mu},\boldsymbol{\Sigma}$:

$$
\boldsymbol{\mu}'=\mathcal{E}\left[\boldsymbol{x}'\right]=\mathcal{E}\left[\boldsymbol{A}\boldsymbol{x}\right]=\boldsymbol{A}\mathcal{E}\left[\boldsymbol{x}\right]=\boldsymbol{A}\boldsymbol{\mu}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}' &= \mathcal{E}\left[\left(\boldsymbol{x}'-\boldsymbol{\mu}'\right)\left(\boldsymbol{x}'-\boldsymbol{\mu}'\right)^t\right]\\
&= \mathcal{E}\left[\left(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}\boldsymbol{\mu}\right)\left(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}\boldsymbol{\mu}\right)^t\right]\\
&= \mathcal{E}\left[\boldsymbol{A}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^t\boldsymbol{A}^t\right]\\
&= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\boldsymbol{t}}
\end{aligned}
$$

转换后的均值依然为高斯分布：

$$
\begin{aligned}
p\left(\boldsymbol{\mu}'\right) &= \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_0'\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{\mu}'-\boldsymbol{m}_0'\right)^t\boldsymbol{\Sigma}_0'^{-1}\left(\boldsymbol{\mu}'-\boldsymbol{m}_0'\right)]\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_0'\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)^t\boldsymbol{A}^t\left(\boldsymbol{A}\boldsymbol{\Sigma}_0\boldsymbol{A}^t\right)^{-1}\boldsymbol{A}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)]\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_0'\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)^t\boldsymbol{A}^t\left(\boldsymbol{A}^t\right)^{-1}\left(\boldsymbol{\Sigma}_0\right)^{-1}\left(\boldsymbol{A}\right)^{-1}\boldsymbol{A}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)]\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{A}\boldsymbol{\Sigma}_0\boldsymbol{A}^t\right|^{\frac{1}{2}}}exp[-\frac{1}{2}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)^t\left(\boldsymbol{\Sigma}_0\right)^{-1}\left(\boldsymbol{\mu}-\boldsymbol{m}_0\right)]
\end{aligned}
$$

log-likelihood:

$$\begin{aligned}
\ln\left[p\left(\boldsymbol{D}' \mid \boldsymbol{\mu}'\right)\right] &= \ln\prod_{k=1}^{n} p\left(\boldsymbol{x}_k' \mid \boldsymbol{\mu}'\right) \\
&= \sum_{k=1}^{n} \ln p\left(\boldsymbol{x}_k' \mid \boldsymbol{\mu}'\right) \\
&= \sum_{k=1}^{n} \ln\{\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}'\right|^{\frac{1}{2}}} exp[-\frac{1}{2}\left(\boldsymbol{x}_k' - \boldsymbol{\mu}'\right)^t \boldsymbol{\Sigma}'^{-1}\left(\boldsymbol{x}_k' - \boldsymbol{\mu}'\right)]\} \\
&= -\frac{n}{2}\ln\left[(2\pi)^d\left|\boldsymbol{\Sigma}'\right|\right] - \sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)^t \boldsymbol{A}^t\left(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t\right)^{-1}\boldsymbol{A}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)\right] \\
&= -\frac{n}{2}\ln\left[(2\pi)^d\left|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t\right|\right] - \sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)^t \boldsymbol{A}^t\left(\boldsymbol{A}^t\right)^{-1}(\boldsymbol{\Sigma})^{-1}(\boldsymbol{A})^{-1}\boldsymbol{A}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)\right] \\
&= -\frac{n}{2}\ln\left[(2\pi)^d\left|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t\right|\right] - \sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)^t \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)\right]
\end{aligned}$$

最新的对 $\boldsymbol{\mu}'$ 的 MAP 估计如下：

$$\begin{aligned}
\hat{\boldsymbol{\mu}}' &= arg\,\underset{\boldsymbol{\mu}}{max}\{\ln\left[p\left(\boldsymbol{D} \mid \boldsymbol{\mu}\right)\right] \times p\left(\boldsymbol{\mu}\right)\} \\
&= arg\,\underset{\boldsymbol{\mu}}{max}\left\{\left[-\frac{n}{2}\ln\left[(2\pi)^d\left|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t\right|\right] - \sum_{k=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)^t \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}_k - \boldsymbol{\mu}\right)\right]\right]\right. \\
&\left.\times\left[\frac{1}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{A}\boldsymbol{\Sigma}_0\boldsymbol{A}^t\right|^{\frac{1}{2}}} exp[-\frac{1}{2}\left(\boldsymbol{\mu} - \boldsymbol{m}_0\right)^t \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\mu} - \boldsymbol{m}_0\right)]\right]\right\}
\end{aligned}$$

通过比较 $\hat{\boldsymbol{\mu}}$ 与 $\hat{\boldsymbol{\mu}}'$，可见两者表达式相同，故可用 MAP 适当估计 $\hat{\boldsymbol{\mu}}'$。

Queestion 3

Consider data $\boldsymbol{D} = \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}2\\\star\end{pmatrix}\right\}$, sampled from a two-dimensional (separable) distribution $p(x_1, x_2) = p(x_1)p(x_2)$, with (1). As usual, $*$ represents a missing feature value.

$$p\left(x_1\right) \sim \begin{cases}\frac{1}{\theta_1}e^{-x_1/\theta_1}, & \text{if } x_1 \geq 0\\ 0, & \text{otherwise}\end{cases} \quad and \quad p\left(x_2\right) \sim \begin{cases}\frac{1}{\theta_1}, & \text{if } 0 \leq x_2 \leq \theta\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

1. Start with an initial estimate $\theta^0 = \begin{pmatrix}2\\4\end{pmatrix}$ and analytically calculate (θ, θ^0)—the E step in the EM algorithm. Be sure to consider the normalization of your distribution.

2. Find the θ that maximizes your (θ, θ^0) —the M step.

Answer 3

解：

$$\begin{aligned}
Q\left(\theta;\theta^{0}\right) &= \mathcal{E}_{x_{32}}\left[\ln p\left(x_g,x_b;\theta\right) \mid \theta^{0}, D_g\right] \\
&= \int_{-\infty}^{+\infty}\left[\ln p\left(x_1 \mid \theta\right)+\ln p\left(x_2 \mid \theta\right)+\ln p\left(x_3 \mid \theta\right)\right] p\left(x_{32} \mid \theta^{0}, x_{31}=2\right) dx_{32} \\
&= \ln p\left(x_1 \mid \theta\right)+\ln p\left(x_2 \mid \theta\right)+\int_{-\infty}^{+\infty} p\left(x_3 \mid \theta\right) p\left(x_{32} \mid \theta^{0}, x_{31}=2\right) dx_{32} \\
&= \ln p\left(x_1 \mid \theta\right)+\ln p\left(x_2 \mid \theta\right)+\int_{-\infty}^{+\infty} p\left(x_3 \mid \theta\right) \frac{p\left(\begin{pmatrix} 2 \\ x_{32} \end{pmatrix} \mid \theta^{0}\right)}{\underbrace{\int_{-\infty}^{+\infty} p\left(\begin{pmatrix} 2 \\ x_{32}' \end{pmatrix} \mid \theta^{0}\right) dx_{32}'}_{=1/(2e)}} dx_{32} \\
&= \ln p\left(x_1 \mid \theta\right)+\ln p\left(x_2 \mid \theta\right)+2e\int_{-\infty}^{+\infty} \ln\left(\frac{1}{\theta_1}e^{-\frac{2}{\theta_1}} \cdot \frac{1}{\theta_2}\right) \cdot \left(\frac{1}{2e}\frac{1}{4}\right) dx_{32} \\
&= \ln\left(\frac{1}{\theta_1}e^{-\frac{1}{\theta_1}} \cdot \frac{1}{\theta_2}\right)+\ln\left(\frac{1}{\theta_1}e^{-\frac{3}{\theta_1}} \cdot \frac{1}{\theta_2}\right)+\frac{1}{4}\ln\left(\frac{1}{\theta_1}e^{-\frac{2}{\theta_1}} \cdot \frac{1}{\theta_2}\right)\int_{-\infty}^{+\infty} 1 dx_{32} \\
&= -2\ln\theta_1\theta_2-\frac{4}{\theta_1}\underbrace{-\frac{1}{4}\left(\ln\theta_1\theta_2+\frac{2}{\theta_2}\right)\int_{-\infty}^{+\infty} 1 dx_{32}}_{\equiv K}
\end{aligned}$$

Case 1: $3 \leq \theta_2 \leq 4$,

$$\begin{aligned}
Q\left(\theta;\theta^{0}\right) &= -2\ln\theta_1\theta_2-\frac{4}{\theta_1}-\frac{1}{4}\left(\ln\theta_1\theta_2+\frac{2}{\theta_2}\right)\int_{0}^{\theta_2} 1 dx_{32} \\
&= -2\ln\theta_1\theta_2-\frac{4}{\theta_1}-\frac{\theta_2}{4}\left(\ln\theta_1\theta_2+\frac{2}{\theta_2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial\theta_1} &= -2/\theta_1+4/\left(\theta_1^2\right)-\theta_2/\left(4\theta_1\right)+\theta_2/\left(2\theta_1^2\right) \\
&= \frac{\left(-8\theta_1+16\right)+\left(-\theta_1\theta_2+2\theta_2\right)}{4\theta_1^2} \\
&= \frac{8\left(2-\theta_1\right)+\theta_2\left(2-\theta_1\right)}{4\theta_1^2} \\
&= \frac{\left(8+\theta_2\right)\left(2-\theta_1\right)}{4\theta_1^2}
\end{aligned}$$

令 $\frac{\partial Q}{\partial\theta_1}=0$，则 $\theta_1=2$ 或 $\theta_2=-8$，而 $p\left(x_2\right) \sim \begin{cases} \frac{1}{\theta_1}, & \text{if } 0 \leq x_2 \leq \theta \\ 0, & \text{otherwise} \end{cases}$ $\therefore \theta_1=2$

$Q=-2\ln 2\theta_2-2-\frac{\theta_2}{4}\left(\ln 2\theta_2+1\right)=-\ln 2\theta_2\left(2+\frac{\theta_2}{4}\right)-\left(2+\frac{\theta_2}{4}\right)=-\left(2+\frac{\theta_2}{4}\right)\left(1+\ln 2\theta_2\right)$

可见，θ_2 越大，Q 越小，又 $3 \leq \theta_2 \leq 4$，$\therefore$ 当 $\theta_2=3$，有 $Q_{max} \approx -7.677$。

Case 2: $\theta_2 \geq 4$,

$$
\begin{aligned}
Q\left(\theta;\theta^0\right) &= -2\ln\theta_1\theta_2 - \frac{4}{\theta_1} - \frac{1}{4}\left(\ln\theta_1\theta_2 + \frac{2}{\theta_2}\right)\int_0^4 1dx_{32} \\
&= -2\ln\theta_1\theta_2 - \frac{4}{\theta_1} - \left(\ln\theta_1\theta_2 + \frac{2}{\theta_2}\right) \\
&= -3\ln\theta_1\theta_2 - \frac{6}{\theta_1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial\theta_1} &= -3/\theta_1 + 6/\left(\theta_1^2\right) \\
&= \frac{6-3\theta_1}{\theta_1^2}
\end{aligned}
$$

令 $\frac{\partial Q}{\partial\theta_1} = 0$，则 $\theta_1 = 2$，

$Q = -3\ln 2\theta_2 - 3$

可见，θ_2 越大，Q 越小，当 $\theta_2 = 4$，有 $Q_{max} \approx -9.238$。

Case 3: $\theta_2 = otherwise$, K=0。

综上，

$$
Q_{max} = \begin{cases}
-7.677, & \text{if } 3 \leq \theta_2 \leq 4, \boldsymbol{\theta} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\
-9.238, & \text{if } \theta_2 \geq 4, \boldsymbol{\theta} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}
\end{cases}
$$

Queestion 4

Consider training an HMM by the Forward-backward algorithm, for a single squence of length T where each symbol could be one of c values. What is the computational complexity of a single revision of all values $\hat{a}_{ij}$ and $\hat{b}_{jk}$?

Answer 4

解：

我们定义从状态 $\omega_i(t-1)$ 转移到 $\omega_j(t)$ 的概率 $\gamma_{ij}(t)$，

$$
\gamma_{ij}(t) = \frac{\alpha_i(t-1)\, a_{ij} b_{jk} \beta_j(t)}{P\left(\boldsymbol{V}^T \mid \boldsymbol{\theta}\right)}
$$

其中，$P\left(\boldsymbol{V}^T \mid \boldsymbol{\theta}\right)$ 是模型用任意的隐状态路径产生序列 $\boldsymbol{V}^T$ 的概率。这样 $\gamma_{ij}(t)$ 就是在产生序列的条件下从状态 $\omega_i(t-1)$ 转移到 $\omega_j(t)$ 的概率。

序列从状态 $\omega_i(t-1)$ 转移到 $\omega_j(t)$ 的预期值是 $\sum_{t=1}^T \gamma_{ij}(t)$,而从 ω_i 的任何转移的总预期数为 $\sum_{t=1}^T \sum_k \gamma_{ik}(t)$，这样，隐状态转移概率 $\hat{a}_{ij}$ 和发出显状态概率 $\hat{b}_{jk}$ 可如此求出：

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

$$\hat{b}_{jk} = \frac{\sum_{\substack{t=1 \\ v(t)=v_k}}^{T} \sum_{l} \gamma_{jl}(t)}{\sum_{t=1}^{T} \sum_{l} \gamma_{jl}(t)}$$

可见，$\hat{a}_{ij}$ 和 $\hat{b}_{jk}$ 复杂度都为 $O\left(c^2 T\right)$

Queestion 5

Consider a normal $p(x) \sim N\left(\mu, \sigma^2\right)$ and Parzen-window function $\phi(x) \sim N(0,1)$. Show that the Parzen-window estimate

$$p_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \phi\left(\frac{x - x_i}{h_n}\right)$$

has the following properties:

1. $\bar{p}_n(x) \sim N\left(\mu, \sigma^2 + h_n^2\right)$
2. $Var\left[p_n(x)\right] \approx \frac{1}{2nh_n\sqrt{\pi}} p(x)$

Answer 5

解：由题可知

$$\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \\
\phi(x) &= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{x^2}{2}\right] \\
p_n(x) &= \frac{1}{nh_n} \sum_{i=1}^{n} \phi\left(\frac{x - x_i}{h_n}\right)
\end{aligned}$$

1.

$$\begin{aligned}
\bar{p}_n(x) &= \mathcal{E}\left[p_n(x)\right] \\
&= \frac{1}{nh_n} \sum_{i=1}^{n} \mathcal{E}\left[\phi\left(\frac{x - x_i}{h_n}\right)\right] \\
&= \frac{1}{h_n} \int_{-\infty}^{+\infty} \phi\left(\frac{x - v}{h_n}\right) p(v)\, dv \\
&= \frac{1}{h_n} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dv - -(1) \\
&= \frac{1}{2\pi h_n \sigma} \int_{-\infty}^{+\infty} exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2}\right) - \frac{1}{2} v^2 \underbrace{\left(\frac{1}{h_n^2} + \frac{1}{\sigma^2}\right)}_{\equiv 1/\theta^2} + v \underbrace{\left(\frac{x}{h_n^2} + \frac{\mu}{\sigma}\right)}_{\equiv \alpha/\theta^2}\right] dv
\end{aligned}$$

即令

$$\begin{cases}\theta^2=\frac{1}{1/h_n^2+1/\sigma^2}=\frac{h_n^2\sigma^2}{h_n^2+\sigma^2}\\ \alpha=\theta^2\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right)\end{cases}$$

$$\begin{aligned}\bar{p}_n\left(x\right)&=\frac{1}{2\pi h_n\sigma}exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}\right)\right]\int_{-\infty}^{+\infty}exp\left[-\frac{1}{2}\left(\frac{v^2}{\theta^2}-\frac{2v\alpha}{\theta^2}+\frac{\alpha^2}{\theta^2}-\frac{\alpha^2}{\theta^2}\right)\right]dv\\&=\frac{1}{2\pi h_n\sigma}exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}-\frac{\alpha^2}{\theta^2}\right)\right]\int_{-\infty}^{+\infty}exp\left[-\frac{1}{2}\left(\frac{v-\alpha}{\theta}\right)^2\right]dv\\&=\frac{1}{2\pi h_n\sigma}exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}-\frac{\alpha^2}{\theta^2}\right)\right]\int_{-\infty}^{+\infty}exp\left[-\left(\frac{w}{\sqrt{2}\theta}\right)^2\right]dw\\&=\frac{1}{2\pi h_n\sigma}exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}-\frac{\alpha^2}{\theta^2}\right)\right]\sqrt{\pi}\sqrt{2}\theta\\&=\frac{\theta}{\sqrt{2\pi}h_n\sigma}exp\left[-\frac{1}{2}\left(\frac{x^2}{h_n^2}+\frac{\mu^2}{\sigma^2}-\frac{h_n^2\sigma^2}{h_n^2+\sigma^2}\left(\frac{x}{h_n^2}+\frac{\mu}{\sigma^2}\right)^2\right)\right]\\&=\frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}exp\left[-\frac{1}{2}\frac{\left(x-\mu\right)^2}{h_n^2+\sigma^2}\right]--(2)\end{aligned}$$

由此可见，$\bar{p}_n\left(x\right)\sim N\left(\mu,\sigma^2+h_n^2\right)$ 。

2.

$$\begin{aligned}Var\left[p_n\left(x\right)\right]&=Var\left[\frac{1}{nh_n}\sum_{i=1}^{n}\left[\phi\left(\frac{x-x_i}{h_n}\right)\right]\right]\\&=\frac{1}{n^2h_n^2}\sum_{i=1}^{n}Var\left[\phi\left(\frac{x-x_i}{h_n}\right)\right]\\&=\frac{1}{nh_n^2}Var\left[\phi\left(\frac{x-v}{h_n}\right)\right]\\&=\frac{1}{nh_n^2}\left\{\mathcal{E}\left[\phi^2\left(\frac{x-v}{h_n}\right)\right]-\left\{\mathcal{E}\left[\phi\left(\frac{x-v}{h_n}\right)\right]\right\}^2\right\}\end{aligned}$$

其中

$$\begin{aligned}\mathcal{E}\left[\phi^2\left(\frac{x-v}{h_n}\right)\right]&=\int\phi^2\left(\frac{x-v}{h_n}\right)p\left(v\right)dv\\&=\int_{-infty}^{+infty}\frac{1}{2\pi}exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2 2\right]\frac{1}{\sqrt{2\pi}\sigma}exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dv\\&\overset{h_n/\sqrt{2}->h_n}{=}\frac{h_n/\sqrt{2}}{\sqrt{2\pi}}\frac{1}{h_n/\sqrt{2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n/\sqrt{2}}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma}exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dv\end{aligned}$$

根据式子 (1) 和 (2)，我们可以得到：

$$\begin{aligned}\mathcal{E}\left[\phi^2\left(\frac{x-v}{h_n}\right)\right]&=\frac{h_n/\sqrt{2}}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2/2+\sigma^2}\right]\\&=\frac{h_n}{2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2/2+\sigma^2}\right]\end{aligned}$$

所以可以得到 $Var\left[p_n\left(x\right)\right]$ 的第 1 项：

$$\begin{aligned}
\frac{1}{nh_n^2}\mathcal{E}\left[\phi^2\left(\frac{x-v}{h_n}\right)\right] &= \frac{1}{nh_n^2}\frac{h_n}{2\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2/2+\sigma^2}\right] \\
&= \frac{1}{2nh_n\sqrt{\pi}}\frac{1}{\sqrt{2\pi}\sqrt{h_n^2/2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2/2+\sigma^2}\right] \\
&\approx \frac{1}{2nh_n\sqrt{\pi}}\left\{\frac{1}{\sqrt{2\pi}\sigma}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{\sigma^2}\right]\right\} \\
&= \frac{1}{2nh_n\sqrt{\pi}}p\left(x\right)
\end{aligned}$$

上式子 $\approx$ 是因为 $\sqrt{h_n^2/2+\sigma^2}\approx\sigma$。

类似地 $Var\left[p_n\left(x\right)\right]$ 的第 2 项：

$$\begin{aligned}
\frac{1}{nh_n^2}\left\{\mathcal{E}\left[\phi\left(\frac{x-v}{h_n}\right)\right]\right\}^2 &= \frac{1}{nh_n^2}\left[\int_{-\infty}^{+\infty}\phi\left(\frac{x-v}{h_n}\right)p\left(v\right)dv\right]^2 \\
&= \frac{1}{nh_n^2}\left\{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-v}{h_n}\right)^2\right]\frac{1}{\sqrt{2\pi}\sigma}exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dv\right\}^2 \\
&= \frac{1}{nh_n^2}\left\{h_n\frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2+\sigma^2}\right]\right\}^2 \\
&= \frac{h_n^2}{nh_n^2}\left\{\frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{h_n^2+\sigma^2}\right]\right\}^2 \\
&\approx \frac{1}{n}\left\{\frac{1}{\sqrt{2\pi}\sigma}exp\left[-\frac{1}{2}\cdot\frac{\left(x-\mu\right)^2}{\sigma^2}\right]\right\}^2 \\
&= \frac{1}{n}p^2\left(x\right) \\
&\approx 0
\end{aligned}$$

$\therefore Var\left[p_n\left(x\right)\right]\approx\frac{1}{2nh_n\sqrt{\pi}}p\left(x\right)$

Queestion 6

Explore the effect of r on the accuracy of nearest-neighbor search based on par- tial distance. Assume we have a large number n of points randomly placed in a d- dimensional hypercube. Suppose we have a test point x, also selected randomly in the hypercume, and find its nearest neighbor. By definition, if we use the full d-dimensional Euclidean distance, we are guaranteed to find its nearest neigh- bor. Suppose though we use the partial distance

$$D_r\left(x,x'\right)=\left(\sum_{i=1}^{r}\left(x_i-x_i'\right)^2\right)^{1/2}$$

1. Plot the probability that a partial distance search finds the true closest neighbor of an arbitrary point x as a function of r for fixed $n\left(1\leq r\leq d\right)$ for $d=10$.

2. Consider the effect of r on the accuracy of a nearest-neighbor classifier. Assume we have n/2 prototypes from each two categories in a hypercube of length 1 on a side. The density for each category is separable into the product of (linear) ramp functions, highest at one side, and zero at the other side of the range. Thus the density for category ω_1 is highest at $(0,0,...0)^t$ and zero at $(1,1,...1)^t$, while the density for ω_2 is highest at $(1,1,...1)^t$ and zero at $(0,0,...0)^t$. State by inspection the Bayesian decision boundary.

Answer 6

解：

1. 假设 n 个样本在 d 维相互独立，均匀分布，
   则每个维度的样本数为：$n^{1/d}$
   r 维上的样本数为：$n^{r/d}$
   $\therefore r$ 维上选中正确数据作为最近邻的概率为：

$$p = n^{r/d}/n = n^{\frac{r}{d}-1}$$

   $\because d = 10$
   $\therefore p = n^{\frac{r}{10}-1}$
   当 $n = 2, 3, 4, ..., 10$，$p - r$ 图如下所示：

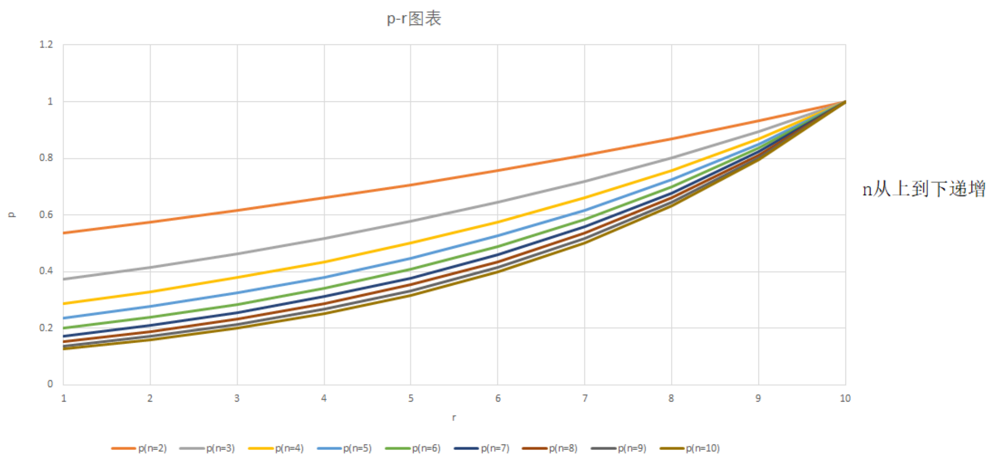

图 2: $p - r$ 图

2. 1 维：$x_1^* = 0.5$
   2 维：$x_2^* = 1 - x_1$
   3 维：$x_1x_2x_3 = (1 - x_1)(1 - x_2)(1 - x_3) \Rightarrow x_3^* = \frac{(1-x_1)(1-x_2)}{1-x_1-x_2+2x_1x_2}$
   ...
   d 维：$\prod_{i=1}^{d} x_i = \prod_{i=1}^{d} (1 - x_i)$