模式识别 第二次作业

2022 年 10 月 2 日

说明

- 作业用中文撰写，务必注明题号，鼓励使用 LaTeX；
- 文档按“学号 _ 姓名.pdf”命名；
- 本次作业截止时间为 2022 年 10 月 19 日，请到课程网站及时提交。

题目 1. 最大似然估计也可以用来估计先验概率。假设样本是连续独立地从自然状态 ω_i 中抽取的，每一个自然状态的概率为 $P(\omega_i)$。如果第 k 个样本的自然状态为 ω_i，那么就记 $z_{ik}=1$，否则 $z_{ik}=0$。

(1) 证明

$$P\left(z_{i1},\cdots,z_{in}|P\left(\omega_i\right)\right)=\prod_{k=1}^{n}P\left(\omega_i\right)^{z_{ik}}\left(1-P\left(\omega_i\right)\right)^{1-z_{ik}}$$

(2) 证明对 $P(\omega_i)$ 的最大似然估计为

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

并且简单解释这个结果。

题目 2. 设 x 的概率密度为均匀分布：

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta, & 0 \le x \le \theta \\ 0, & otherwise \end{cases}$$

(1) 假设 n 个样本 $\mathcal{D} = \{x_1, \cdots, x_n\}$ 都独立地服从分布 $p(x|\theta)$。证明对于 θ 的最大似然估计就是 $\mathcal{D}$ 中的最大值 $\max[\mathcal{D}]$。

(2) 假设从该分布中采样 5 个样本 ($n = 5$)，且有 $\max_k x_k = 0.6$，画出在区间 $0 \le \theta \le 1$ 上的似然函数 $p(\mathcal{D}|\theta)$，并解释为什么此时不需要知道其余四个点的值。

题目 3. 一种度量同一空间中的两个不同分布的距离的方式为 Kullback-Leibler 散度 (简称 KL 散度)

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

这个距离度量并不符合严格意义上的度量必须满足的对称性和三角不等式关系。假设我们使用正态分布 $p_1(\mathbf{x}) \sim N(\boldsymbol{\mu}, \Sigma)$ 来近似某一个任意的分布 $p_2(\mathbf{x})$。证明能够产生最小的 KL 散度的结果为下面这个明显的结论：

$$\begin{aligned} \boldsymbol{\mu} &= \varepsilon_2[\mathbf{x}] \\ \Sigma &= \varepsilon_2\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t\right] \end{aligned}$$

其中的数学期望是对概率密度函数 $p_2(\mathbf{x})$ 进行的。

题目 4. 数据 $\mathcal{D} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ * \end{pmatrix} \right\}$ 中的样本独立地服从二维的分布 $p(x_1, x_2) = p(x_1)p(x_2)$。其中，$*$ 代表丢失的数据，且有

$$p(x_1) = \begin{cases} \frac{1}{\theta_1} e^{-x_1/\theta_1}, & x_1 \ge 0 \\ 0, & otherwise \end{cases}$$

和

$$p\left(x_2\right) \sim U\left(0, \theta_2\right)=\begin{cases}\frac{1}{\theta_2}, & 0 \leq x_2 \leq \theta \\ 0, & otherwise\end{cases}$$

(1) 简要描述 EM 算法。

(2) 假设初始估计为 $\boldsymbol{\theta}^0=\begin{pmatrix}2 \\ 4\end{pmatrix}$，计算 $Q\left(\boldsymbol{\theta} ; \boldsymbol{\theta}^0\right)$ (EM 算法中的 E 步)。注意要对分布进行归一化。

(3) 求使得 $Q\left(\boldsymbol{\theta} ; \boldsymbol{\theta}^0\right)$ 最大的那个 θ (EM 算法中的 M 步)。

题目 5. 用前向-后向算法 (即 Baum-Welch 算法) 训练一个 HMM，已知训练序列长为 T，其中每一时刻都可能取 c 个符号中的一个。那么全部更新一次 $\hat{a}_{ij}$ 和 $\hat{b}_{jk}$ 的计算复杂度是多少？

题目 6. 假设有正态分布 $p(x) \sim N\left(\mu, \sigma^2\right)$ 和 Parzen 函数 $\varphi(x) \sim N(0,1)$。证明 Parzen 窗估计

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

有如下性质：

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

(2) $\operatorname{Var}\left[p_n(x)\right] \approx \frac{1}{2 n h_n \sqrt{\pi}} p(x)$。