数据科学与工程数学基础

作业提交规范及第 19 次作业

教师：黄定江
助教：陈诺、刘文辉

2022 年 5 月 27 日

作业提交规范

1. 作业提交形式：**练习本或笔记本**（建议统一使用一般的**练习本**即可，不接收以纸张的方式书写的作业）。
2. 作业书写说明：
   (a) 可以讨论，禁止抄袭！
   (b) 练习本封面至少包含两方面信息：**姓名**和**学号**
   (c) 每一次的作业**请另起一页**，并在**第一行标明第几次作业**。例如“第 19 次作业”；
   (d) 每一题请**标注题号**，无需抄题，直接解答；
   (e) 题与题之间**请空一行**；
   (f) 不要求字好，但要求书写整体清晰易读。
3. 作业提交途径：纸质作业交给**学习委员**，由学习委员**按学号顺序**收齐后统一在截止日期前交到**助教实验室**。**单数周**布置的作业交到助教刘文辉处**数学馆西 109**；**双数周**布置的作业交到助教陈诺处**地理馆 353**。
4. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分。

第 19 次作业

提交截至时间：**暂定 2022/06/03 周五 20:00（晚上）**

理论部分

习题 1. *(互信息)* 假设 $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ 是一个马尔科夫链，即

$$p\left(x_1, x_2, \ldots, x_n\right)=p\left(x_1\right) p\left(x_2 \mid x_1\right) \cdots p\left(x_n \mid x_{n-1}\right)$$

试化简 $I\left(X_1 ; X_2, \ldots, X_n\right)$

解.

$$\begin{aligned}
I\left(X_1 ; X_2, \ldots, X_n\right) &= H\left(X_1\right)-H\left(X_1 \mid X_2, \ldots, X_n\right) \\
&= H\left(X_1\right)-\left[H\left(X_1, X_2, \ldots, X_n\right)-H\left(X_2, \ldots, X_n\right)\right] \\
&= H\left(X_1\right)-\left[\sum_{i=1}^n H\left(X_i \mid X_{i-1}, \ldots, X_1\right)-\sum_{i=2}^n H\left(X_i \mid X_{i-1}, \ldots, X_2\right)\right] \\
&= H\left(X_1\right)-\left[\left(H\left(X_1\right)+\sum_{i=2}^n H\left(X_i \mid X_{i-1}\right)\right)-\left(H\left(X_2\right)+\sum_{i=3}^n H\left(X_i \mid X_{i-1}\right)\right)\right] \\
&= H\left(X_2\right)-H\left(X_2 \mid X_1\right) \\
&= I\left(X_2 ; X_1\right) \\
&= I\left(X_1 ; X_2\right)
\end{aligned}$$

习题 2. *(通过 KL 散度理解 MLE)* 假设 $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 来自密度为 $p(\mathbf{x})$ 的分布 P，试说明如果采用具有密度函数 $q_\theta(\mathbf{x})$ 的分布族 Q_θ 来计算 *MLE*，那么 *MLE* 将试图找到在 *KL* 散度意义上最接近真实分布 P 的分布 Q_θ 。
即证明

$$arg \max_\theta \prod_{i=1}^n q_\theta\left(\mathbf{x}_i\right) \Longleftrightarrow \arg \min_\theta D_{\mathrm{kl}}\left(P \| Q_\theta\right)$$

解.

$$\begin{aligned}
& \arg \max_\theta \prod_{i=1}^n q_\theta\left(\mathbf{x}_i\right) \Longleftrightarrow \arg \min_\theta -\frac{1}{n} \sum_{i=1}^n \log q_\theta\left(\mathbf{x}_i\right) \\
& \xrightarrow{P} \arg \min_\theta -E_P \log q_\theta(\mathbf{x}) \Longleftrightarrow \arg \min_\theta -\int p(\mathbf{x}) \log q_\theta(\mathbf{x}) d\mathbf{x} \\
& \Longleftrightarrow \arg \min_\theta H\left(P, Q_\theta\right) \Longleftrightarrow \arg \min_\theta \left(H\left(P, Q_\theta\right)-H(P)\right) \\
& \Longleftrightarrow \arg \min_\theta \left\{-\int p(\mathbf{x}) \log q_\theta(\mathbf{x}) d\mathbf{x}+\int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}\right\} \\
& \Longleftrightarrow \arg \min_\theta \left\{-\int p(\mathbf{x}) \log \frac{q_\theta(\mathbf{x})}{p(\mathbf{x})}\right\} \Longleftrightarrow \arg \min_\theta D_{\mathrm{kl}}\left(P \| Q_\theta\right)
\end{aligned}$$

其实，从优化模型参数角度来说，最小化负对数似然，交叉熵 *(多分类问题中)*，*KL* 散度这 *3* 种方式是一样的。