

数据科学与工程数学基础

作业提交规范及第 8 次作业

教师：黄定江

助教：陈诺、刘文辉

2022 年 11 月 25 日

作业提交规范

1. 作业提交形式：使用 Word 或 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ 编写所得到的电子文档。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。
2. 作业命名规范：提交的电子文档必须命名为：“学号_姓名”。命名示例：10175501112_陈诺。
3. 作业提交途径：点击打开每次作业的传送门网址：[第 8 次作业提交传送门](#)，无需注册和登录，直接上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 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 8 次作业



提交截至时间：2022/12/5 周一 12:00（中午）

理论部分

习题 1. 同时抛 2 颗骰子, 事件 A, B, C 分别表示为

(A) 仅有一个骰子是 3

(B) 至少一个骰子是 4

(C) 骰子上点数总和为偶数。

试计算事件 A, B, C 发生后所提供的信息量

解. $H_A = -\log \frac{5}{18}, H_B = -\log \frac{11}{36}, H_C = -\log \frac{1}{2}$

习题 2. 一个容器里面装有 a 个红球和 a 个白球, 若从容器中取出 $k, (k \geq 2)$ 个球。对于有放回和无放回两种情况, 哪种情况的熵更大? 请回答并给予说明。

解. 考虑集合 $\{(x_1, x_2, \dots, x_n) | x_i = 0 \text{ or } 1\}$ 如果 $x_i = 0$ 则代表第 i 次取出红球否则取出白球。在有放回的情况下, 取得该集合里面任意元素的概率都是相同的, 且概率和为 1。而在无放回的情况下, 则取得不同元素的概率是有可能不同的。且概率和也为 1。根据熵的极致性得, 有放回的情况下熵更大。

习题 3. 证明: 在多分类问题中, 利用交叉熵函数作为损失函数和用 KL 散度作为损失函数是等价的。

解. 真实分布: 设第 i 个样本 x_i 属于 y_i 类, 真实标签分布为 p_i , p_i 是第 y_i 个分量为 1 的 one-hot 向量。预测分布: 对于第 i 个样本 x_i , 预测标签分布是 $q_i = f(x_i; \theta)$, θ 是要学习的参数。

$$KL \text{ 散度} = (p_i^T \log p_i - p_i^T \log q_i)$$

$$\text{交叉熵} = (-p_i^T \log q_i)$$

由于真实标签是真实存在的, 不变的。所以 $\arg \min_{\theta} KL \text{ 散度} = \arg \min_{\theta} \text{交叉熵}$ 。

习题 4. (互信息) 假设 $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \dots \rightarrow X_n$ 是一个马尔科夫链, 即

$$p(x_1, x_2, \dots, x_n) = p(x_1) p(x_2 | x_1) \cdots p(x_n | x_{n-1})$$

试化简 $I(X_1; X_2, \dots, X_n)$

解.

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