

清华大学本科生考试试题专用纸

考试课程 系统工程导论 (D 卷) 2020 年 6 月 1 日

姓名 班级 学号

注意：请将所有答案写在答题纸上，考试结束试卷和答题纸一并上交)

1. 现有 4 个样本点, $x_1^T = (0,0)$, $x_2^T = (1,1)$, $x_3^T = (0,1)$, $x_4^T = (2,0)$ 。小明同学希望对样本点进行系统聚类。在查阅资料后, 小明选择用“马氏距离”衡量样本点间相似性, 以两个类簇内任意两个样本点的最小距离衡量类簇间的距离, 且距离越小, 样本点越相关。即:

$$\text{dist}(a,b) = (a-b)^T \Sigma (a-b)$$

$$\text{dist}(A,B) = \min(\text{dist}(a,b)) \quad \forall a \in A, \forall b \in B$$

$$R(a,b) = \frac{1}{\text{dist}(a,b)}$$

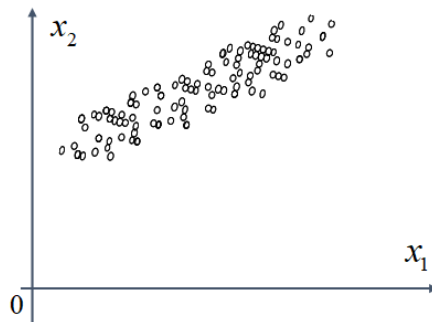
其中 $\text{dist}(a,b)$ 表示 a,b 两个样本点的距离, $\text{dist}(A,B)$ 表示 A,B 两个样本集合之间的距离, $R(a,b)$ 表示 a,b 两个样本点间的相似性, 我们取

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$$

请你替小明完成系统聚类过程, 并画出聚类谱系图。(10 分)

2. 简答题。(35 分)

- A) 在不确定性决策中, 什么是 $e(i)$? 为什么要引入 $e(i)$? (2 分)
- B) 层次分析法的优缺点是什么? 多目标规划与层次分析法之间有什么关系? (3 分)
- C) 系统工程由哪些模块构成? 请在每个模块里举出一个代表性方法及代表人物 (4 分)
- D) K-均值聚类是否一定会收敛? 为什么? 是否一定收敛到最优值? 为什么? (6 分)
- E) 给定一批二维样本, 分布如下图所示 (9 分):



- ① 请对主成分分析方法 (PCA) 的步骤进行概述;
- ② 现在要对上图的数据作 PCA, 并且只保留一个 PCA 主轴, 请在答题纸上大致画出该主轴的方向;
- ③ 试说明 PCA 与病态线性回归的关系;
- ④ 给出 PCA 数据压缩比计算公式, 并说明各变量含义

F) 系统工程发展的四个阶段及其代表案例 (4 分)

G) 利用可达矩阵获取骨架图是否可以从下至上进行, 步骤是什么? (2 分)

H) 我们在课程里学习了通过回归来拟合样本点; 事实上, 通过回归也可以进行分类任务。考虑以下的二分类任务: 假设有 n 个样本点 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 其中, $x_i \in R$, $y_i \in \{0, 1\}$, y_i 表示了每个点的真实类别。我们希望

使用 $\hat{y}_i = \frac{e^{ax_i}}{1 + e^{ax_i}}$ 来估计样本点的类别 (a 为待计算的参数, $\hat{y}_i \in (0, 1)$)。对于该任务, 可以应用极大似然法估计参数模型, 其对应的模型称为 Logistic 回归模型。我们使用交叉熵作为损失函数, 即:

$$L = \sum_{i=1}^n (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)),$$

其中 y_i 表示真实类别, \hat{y}_i 表示估计值。请计算损失函数 L 相对于 a 的梯度, 即 $\frac{\partial L}{\partial a}$ (5 分)

3. 小明在解释性结构建模上遇到了一些问题, 想请你帮帮他 (9 分)。

(1) 根据可达矩阵 A 求骨架图 (要求: 使用尽可能少的有向边完成骨架图) (6 分)

(2) 根据骨架图, 给出对应该可达矩阵的邻接矩阵 (3 分)

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

4. 随着假期的到来, 某班级共 31 名同学准备暑期郊游, 目前有 A、B、C、D 四处景点可供选择, 同学们对四处景点根据喜好顺序进行了排序, 得到如下结果:

① 4 名同学认为 $A > B > C > D$, ② 7 名同学认为 $A > C > B > D$, ③ 9 名同学认为 $C > B > D > A$

④ 4 名同学认为 $D > C > A > B$, ⑤ 3 名同学认为 $B > C > A > D$, ⑥ 4 名同学认为 $D > B > A > C$

请分别采用简单多数规则、绝对多数规则和 Borda 规则, 说明最终该班级会去哪个景点旅游。(6 分)

5. 小明高考结束, 现准备从甲、乙、丙三所大学中选择一所填写志愿, 主要考虑“所在地”和“声望”两个因素。为了合理度量声望, 小明又将“声望”因素分为“师资力量”和“科研水平”两个维度进行考察。综合考察后, 得到了如下所示的 A、B、C、D、E 五个矩阵, 其中: A 是小明对于“所在地”和“声望”两个因素的重要性的判断矩阵, B 矩阵是小明对于“师资力量”和“科研水平”两个维度的重要性的判断矩阵, C、D、E 分别是小明对这三所大学在“所在地”、“师资力量”和“科研水平”三个方面的相对优劣程度的判断矩阵。(共 15 分, 计算时请保留 3 位有效小数)

A) 请按照层次分析方法要求, 画出层次结构模型。(2 分)

B) 矩阵 B、C、D 是否一致? 请给出判断结果, 并说明理由。(3 分)

C) 请根据题目信息, 请给出小明最终会填写志愿的学校, 并说明理由。(10 分)

$$A = \begin{bmatrix} 1 & \frac{3}{7} \\ \frac{7}{3} & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{2}{3} & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & \frac{3}{4} & 1 \\ \frac{4}{3} & 1 & \frac{4}{3} \\ 1 & \frac{3}{4} & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & \frac{7}{2} & 7 \\ \frac{2}{7} & 1 & 2 \\ \frac{1}{7} & \frac{1}{2} & 1 \end{bmatrix}$$

6. 在学生节活动中，有一个抽扑克牌获得积分的游戏环节，规定如下：在桌面上有相同的 A、K、Q、J 四张扑克，每次抽一张，抽出的扑克不放回，抽到的结果与输赢规则为：

- 1) 摸到 A，得 3 分
- 2) 摸到 K，得 2 分
- 3) 摸到 Q，输 1 分
- 4) 摸到 J，输 5 分

每次抓完一个球之后，都可以选择继续玩或者不玩，也可以一开始就不玩。请用决策树的方法分析一下，是否值得花 1 个积分去玩这个游戏。（12 分）

提示：

- ① 本题决策树分支可能较大，请同学们在答题纸上合理安排空间，对于较为明显的分支可以不全部画出来
- ② 已知选择玩的话，如果第一次摸到 A，则之后的最优选择为不继续玩，且期望效益值为 3；如果第一次摸到球 J，则期望效益值为 -1（上述两个结果都还没有扣除 1 分的游戏费用）。这两个结果可以直接使用。

7. 甲同学参加一个抽奖活动，假设抽奖店老板和甲同学关于变量 x 有相同的效用函数 $u(x) = x^3$ 。现在抽奖店老板免

费得到一张抽奖券，它能够以概率 p 赢得 R ，且 $0 < p < 1$ 。那么：

- A) 画出效用函数曲线，并说明两人属于什么类型的决策者。（3 分）
- B) 如果老板想要出售该抽奖券，那么他可以接受的最低出售价格是多少？（3 分）
- C) 如果甲同学想要购买该抽奖券，那么他可以接受的最高买入价格是多少？（3 分）
- D) 请你判断这次交易能否成交，并说明理由。（4 分）