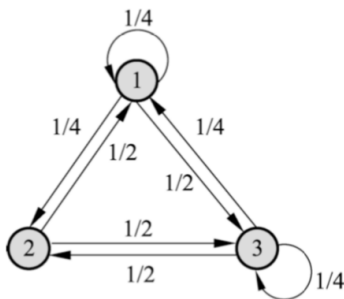


2022-2023 春夏数院数据建模与分析回忆卷

HobbitQia

1. 假设随机变量有四种取值 A, B, C, D , 其中 $P(A) + P(B) = \frac{1}{4}$ 。用最大熵模型求出 $P(A), P(B), P(C), P(D)$ (要有详细过程)
2. 随机变量 X 为 $m(m > 2)$ 维的向量, 求出主成分分析中的第 k 大主成分 y_k , 并证明 y_k 的方差是 X 的协方差矩阵的第 k 大特征值。
3. 证明下面的马尔可夫链不是可逆的。



4. 求 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 5 & 0 \end{bmatrix}$ 的奇异值分解和紧奇异值分解, 并说明奇异值分解的几何意义。

5. 简答题

- (a) 感知机的学习策略是什么? 学习得到的超平面是否唯一?
 - (b) k 近邻和 k 均值聚类中的 k 的意义是什么? 如何选取合适 k 值?
 - (c) 说明朴素贝叶斯算法的思想, 以及其优缺点。
 - (d) 写出决策树模型的过程。
6. 谱聚类中, 我们有拉普拉斯矩阵 $L = D - W$ 其中 D 为度矩阵, W 为权重矩阵。
 - (a) 证明 L 有 n 个非负特征值, 而且其中一个特征值为 0。
 - (b) 证明拉普拉斯矩阵 L 的零特征值的重数等于图 G 的连通子图的个数 k 。
 7. 假设支持向量机的约束最优化问题为如下形式

$$\begin{aligned} \min_{w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, N \\ & x_i \geq 0, i = 1, 2, \dots, N \end{aligned}$$

- (a) 求此时支持向量机的对偶形式
- (b) 求支持向量机的支持向量，并说出当去掉支持向量以外的向量时，超平面会不会改变。
- (c) 说明如何通过核方法将上述线性支持向量机转化为非线性支持向量机。