

不定项选择题

1.关于正态分布，下列说法错误的是：

- A. 正态分布具有集中性和对称性
- B. 正态分布的均值和方差能够决定正态分布的位置和形态
- C. 正态分布的偏度为 0，峰度为 1
- D. 标准正态分布的均值为 0，方差为 1

2.我们想在大数据集上训练决策树，为了使用较少时间，我们可以：

- A. 增加树的深度
- B. 减少树的深度
- C. 增加学习率（learning rate）
- D. 减少树的数量

3.关于相关系数,下列描述中正确的有：

- A. 相关系数为 0.8 时，说明两个变量之间呈正相关关系
- B. 相关系数等于 1 相较于相关系数等于-1，前者的相关性更强
- C. 相关性等于 1 相较于相关系数等于 0，前者的相关性更强
- D. Spearman 相关系数可以衡量两个定序变量之间的相关程度

4.下列哪些方法可以用来对高维数据进行降维：

- A. LASSO
- B. 主成分分析法
- C. 小波分析法
- D. 线性判别法

5.“点击率问题”是这样一个预测问题，99%的人是不会点击的，而 1%的人是会点击进去的，所以这是一个非常不平衡的数据集。假设，现在我们已经建了一个模型来分类，而且有了 99% 的预测准确率，我们可以下的结论是：

- A. 模型预测准确率已经很高了，我们不需要做什么了
- B. 模型预测准确率不高，我们需要做点什么改进模型
- C. 无法下结论
- D. 以上都不对

编程题

6.输入一棵多叉树，多叉树的每个节点有一个编号，要求对树的所有节点排序，使得：对于每个节点，它的父节点排在它的后面。输入会以边的集合的形式给出，假如树有 N 个节点，那么输入是一个大小为 $N-1$ 的列表，列表的每个元素是 `Edge` 类型，`Edge` 的定义为：

```
class Edge {
    public int u;
    public int v;
}
```

其中 u 表示父节点的编号， v 表示子节点的编号。请实现函数：

```
List<int> sort(List<Edge> edges) {
}
```

7.假设某卡牌类游戏里面，你搜集了 17 种卡牌，每种卡牌的攻击伤害率从小到大排列分别为 1, 4, 9, \dots , 289，即 n 的平方 ($1 \leq n \leq 17$, n 为正整数)，并且每种卡牌的数量足够多，现在请你计算出有多少种卡牌组合方式使得卡牌的总攻击率为 m ($m \leq 500$) 比如 $m=10$ ，那么有 4 种组合卡牌的方式，相同种卡牌可重复出现；

8.给定一个 $N \times M$ 个像素大小的黑白图片，图片上有若干个黑色连通区域。编写代码，为每个黑色连通区域找到能包含它的最小矩形，其中矩形的长宽与图片的长宽分别平行。注意连通是指 8 方向连通。

矩形类的定义为：

```
class Rect {
    public int top;
    public int bottom;
    public int left;
    public int right;
}
```

要实现的函数签名

```
List<Rect> findRects(boolean[][] image) {
}
```

其中如果 `image[x][y]` 为 `false` 则对应坐标点 (x, y) 为白色像素点，否则为黑色像素点。

9.给定一个大小为 $N \times N$ 的二维数组，数组元素有正也有负，这个二维数组有很多子数组，每个子数组有一个和，求所有子数组的和的最大值是多少。编程实现，要求算法时间复杂度不大于 $O(N^3)$ 。