

## 第十次作业

阅读机器学习实战 8.1, 8.2, 8.3, 9.1

### Exercise 10.1

聚类属于哪种学习方式( )。

- A. 无监督学习
- B. 监督学习
- C. 强化学习
- D. 都不属于

### Exercise 10.2

关于K均值和DBSCAN的比较，以下说法不正确的是( )。

- A. K均值很难处理非球形的簇和不同大小的簇
- B. DBSCAN使用基于密度的概念
- C. DBSCAN可以处理不同大小和不同形状的簇。
- D. K均值使用簇的基于层次的概念

### Exercise 10.3

关于kmean算法的实现描述错误的是 ( )

- A. 收敛速度慢
- B. 可以轻松发现非凸形状的簇
- C. 原理简单，实现容易
- D. 需要事先确定  $k$  的值

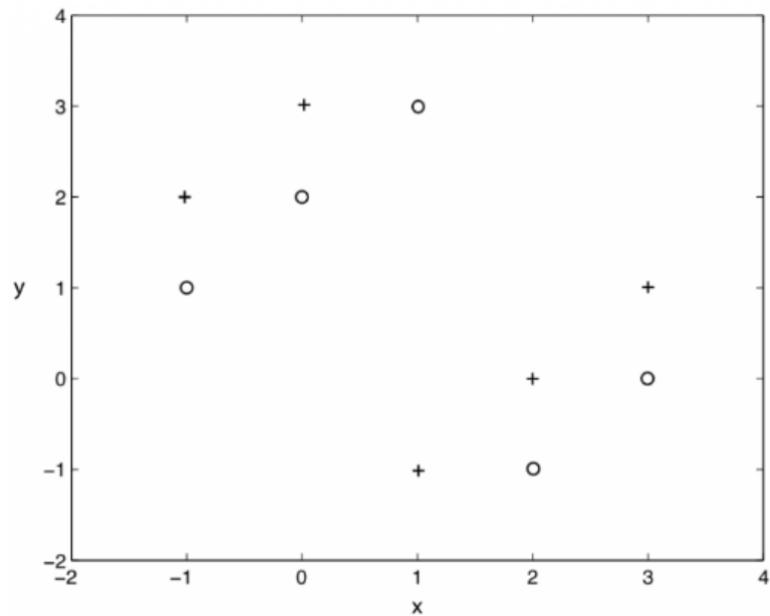
### Exercise 10.4

下列关于Kmeans聚类算法的说法错误的是( )。

- A. 对大数据集有较高的效率并且具有可伸缩性
- B. 初始聚类中心的选择对聚类结果影响不大
- C. 是一种无监督学习方法
- D. K值无法自动获取，初始聚类中心随机选择

### Exercise 10.5

假设我们使用 kNN 训练模型，其中训练数据具有较少的观测数据（下图是两个属性  $x$ 、 $y$  和两个标记为“+”和“o”的训练数据）。现在令  $k = 1$ ，则图中的 Leave-One-Out 交叉验证错误率是多少？



- A. 0%
- B. 20%
- C. 50%
- D. 100%

### Exercise 10.6

k-means是一种迭代算法，在其内部循环中重复执行以下两个步骤，哪两个？

- A、移动簇中心，更新簇中心 $u_k$
- B、分配簇，其中参数 $c^{(i)}$ 被更新
- C、移动簇中心 $u_k$ ,将其设置为等于最近的训练示例 $c^{(i)}$
- D、簇中心分配步骤，其中每个簇质心 $u_i$ 被分配（通过设置 $c^{(i)}$ ）到最近的训练示例 $x^{(i)}$

### Exercise 10.7

给定含有5个样本的集合

$$X = \begin{bmatrix} 0 & 0 & 1 & 5 & 5 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

请用  $k$  均值聚类算法将样本聚到两个类中。

### Exercise 10.8

最常用的降维算法是PCA，以下哪项是关于PCA的？

- 1、PCA是一种无监督的方法
- 2、它搜索数据具有最大差异的方向
- 3、主成分的最大数量 $\leq$  特征的数量
- 4、所有主成分彼此正交

- A、 2, 3和4
- B、 1, 2和3
- C、 1, 2和4
- D, 以上都有

### Exercise 10.9

主成分分析 (PCA) 是一种重要的降维技术, 以下对于PCA的描述不正确的是 ( ) :

- A、主成分分析是一种无监督方法
- B、主成分数量一定小于等于特征的数量
- C、各个主成分之间相互正交
- D、原始数据在第一主成分上的投影方差最小

### Exercise 10.10

应用PCA后, 以下哪项可以是前两个主成分?

- 1、 (0.5, 0.5, 0.5, 0.5) 和 (0.71, 0.71, 0.0)
- 2、 (0.5, 0.5, 0.5, 0.5) 和 (0.0, -0.71, 0.71)
- 3、 (0.5, 0.5, 0.5, 0.5) 和 (0.5, 0.5, -0.5, -0.5)
- 4、 (0.5, 0.5, 0.5, 0.5) 和 (-0.5, -0.5, 0.5, 0.5)

A、1和2

B、1和3

C、2和4

D、3和4

### Exercise 10.11

维度的诅咒是什么?

### Exercise 10.12

一旦降低了数据集的维度, 是否可以逆操作? 如果可以, 怎么做? 如果不能, 为什么?

### Exercise 10.13

如何定义聚类? 你能列举几种聚类算法吗? 聚类算法的主要应用有哪些?