

第一章 概述

一、填空题

- 1 将模式识别任务看做 1 个输入输出系统 $y = f(x)$ ，则对于分类任务， y 通常为 (类别标签)；对于回归任务， y 通常为 (待预测变量)。
- 2 利用一个线性分类器对 iris 数据库进行分类，其样本空间是 (共 150 株花) 的集合，维度是 (4)；其假设空间是 (所有可能的线性分类器) 的集合，其参数空间维度是 (—)；其类别标签集合是 (—)。

二、判断题

- 3 一个数据集的特征维度总是小于该数据集的样本数量。(X)
- 4 模型容量大小与训练集样本数量无关。(✓)
- 5 明日降水概率预测属于回归问题。(✓)
- 6 班级内自由分组讨论不属于分类问题。(✓)
- 7 驯兽师用鞭打和喂食的方式教导老虎跳火圈属于有监督学习。(✓)

三、选择题

- 8 处理二维数据的单个神经元模型的参数个数是：()
A. 3 个 B. 2 个 C. 1 个 D. 不确定
- 9 从当前考场的所有考生中，找出作弊的学生，该任务属于：(A)C
A. 分类 B. 聚类 C. 异常检测 D. 以上均是
- 10 以下方法中不属于主流机器学习方法的是：(C)
A. 有监督学习 B. 无监督学习 C. 随机学习 D. 强化学习

四、简答题

- 11 什么是分类任务？分类任务有哪些实际的应用？
- 12 什么是回归任务？回归任务有哪些实际的应用？
- 13 分类任务和回归任务的区别是什么？
- 14 什么是聚类任务？请列举几个聚类的实际应用。
- 15 什么是异常检测任务？异常检测的实际应用是什么？
- 16 人脸图像伪造属于分类任务还是回归任务，为什么？
- 17 Iris 数据集有哪几个类别？每个样本都包含那些特征？
- 18 模式识别方法可以分为哪三种思路？请分别举例说明！

- 19 现代机器学习的三种主流方法是什么，各自的定义和特点是什么？
- 20 如果一个模型在训练集上表现良好，但在测试集上表现不佳，试说明该模型的容量与问题复杂度之间的关系。

五、计算(画图)题

- 21 下面是一个房屋价格预测数据集的部分数据，该任务是分类任务还是回归任务？是有监督还是无监督？试写出每个样本的特征向量？特征空间的维度是多少？

编号	占地面积/m ²	距市中心距离/km	地下室面积/m ²	房屋价格/万元
1	100	12	3	81.5
2	120	5	6	126
3	90	18	5	76.2
4	150	3	10	165.3
5	200	20	0	161
6	160	9	8	152.6

- 22 请画出有监督机器学习系统的学习阶段的系统框图

$(x_1, y_1) (x_2, y_2) \dots (x_n, y_n)$

