

机器学习导论23fall期末

1. 绪论
 1. 解释"奥卡姆剃刀" (4分)
 2. 举例说明模型 (SVM, CNN等) 的归纳偏好 (4分)
2. 模型评估与选择
 1. 画出二分类混淆矩阵 (4分)
 2. 写出 precision, recall, F1 的公式 (4分)
3. 线性模型:
 1. 考虑以下多分类学习问题: 样本数量为 m , 类别数量为 N , 每个类别的样本数量一致. 假设一个二分类算法对于大小为 m 的数据训练的时间复杂度为 $O(m)$, 试分别计算该算法在 OvO、OvR 策略下训练的总体时间复杂度; 对比这两者的优劣之处;
 2. OvO和OvR存在不适用的场景吗?
 3. OvR每次将特定的类标记为正类, 其他类标记为负类, 需不需要显式地考虑
4. 决策树: 给出了一个简单数据集 D , $|D| = 8$, 特征是 $X, Y, Z \in \{0, 1\}$, 标签是 $f \in \{0, 1\}$
 1. 写出数据集的信息熵 (5分)
 2. 分别写出按照 X, Y, Z 划分的信息增益 (5分)
 3. 手搓决策树 (4分)
5. 神经网络相关
 1. 画出MP神经元, 并介绍它的输入输出分别是什么
 2. 写出两个常用的激活函数; 激活函数在神经网络中有什么作用?
 3. 如何缓解神经网络过拟合? 写出两种方法并介绍原理
6. SVM
 1. 写出软间隔SVM原问题, 和引入 ξ_i 后的形式 (2分)
 2. 写出软间隔SVM原问题的Laplacian, 并推导对偶问题 (10分)
7. 贝叶斯分类器: 见23fall PS5-3 (5+5分) (手搓贝叶斯分类器+用拉普拉斯修正再搓一遍)
8. 集成学习:
 1. 写出两种集成学习中学习器的结合方法 (4分)
 2. 为什么Random Forest优于一般的决策树? (3分)
9. 聚类:
 1. 写出两个距离度量;
 2. 距离度量的四个特性;
 3. k-means算法伪代码
10. 给一个分类任务数据集, 如何用PCA进行分类?