

## 选择题 (10\*2)

---

只记得几题了：

- 1.有四类人员，使用什么算法模型：多分类
- 2.关于神经网络说法错误的是
- 3.生成式算法：贝叶斯
- 4.关于聚类说法错误的是
- 5.训练集，测试集，验证集的作用（选项表述感觉不准确）
- 6.给了一个表计算查全率查准率
- 7.ovo需要训练多少个分类器：  $(k-1)k/2$
8. $p(x, c)$  是联合分布
- 9.C4.5 决策树的指标是：信息增益率
- 10.分类和回归的区别：输出为连续和离散

## 简答题 (4\*10)

---

- 1.简述假设空间和版本空间的区别
- 2.简述两种求最小值的方法
- 3.简述类型不平衡和解决办法
- 4.简述回归树和决策树的特点和区别

## 计算题 (2\*15)

---

- 1.K-means聚类给定A1,A2,A3,B1,B2,B3,C1,C2,初始选择A1, B1,C1,计算第一轮循环之后簇的分组情况和三个簇的中心坐标
- 2.贝叶斯

## 思考题 (10)

---

给出两种问题：训练集上误差大，测试集也大，和 训练集上误差小，测试集误差大

分别指出问题类型和导致原因，并分别针对决策树和神经网络给出解决办法