

简答题(50 分)

1. 简述监督学习和无监督学习的概念，并至少各给出一个模型举例
2. 简述过拟合和欠拟合的概念，并至少给出一种各自的缓解策略
3. 阐述留出法、K 折交叉验证法、自助法的原理
4. 阐述预剪枝和后剪枝有什么优缺点
5. 谈谈你对核函数的认识(例如定义，基本原理，常见的核函数),至少两方面

思考题(20 分)

1. 给出聚类 K 均值算法的伪代码描述
2. 阐述集成学习 bagging 和 boosting 算法的原理，并两者在偏差-方差上关注的差异

计算题(30 分)

1. 神经网络，两个输入，一个隐层两个神经元，两个输出。给定的是矩阵值，sigmoid 激活函数，有 e 参考值。(15 分)

(1) 画出神经元结构图

(2) 计算隐层和输出层的输入和输出。

2. 计算两个保安抓小偷的查准率和查全率，并讨论在要求尽可能抓到所有小偷的情况下，谁的表现更好? (8 分)

A 保安:

	实际上是小偷	实际上是好人	合计
抓住的小偷	5	1	6
放走的好人	15	79	94
合计	20	80	100

B 保安:

	实际上是小偷	实际上是好人	合计
抓住的小偷	20	20	40
放走的好人	0	60	60
合计	20	80	100

B 保安第一列的数据是 10, 10, 20

3. 计算朴素贝叶斯概率，有点像下面这个，但是具体数据有差异，有参考公式:

二、(朴素贝叶斯)试由下表的训练数据学习一个朴素贝叶斯分类器并确定  $x = (2, S, T)$  的类判别结果  $y$ 。表中  $X^{(1)}$ ,  $X^{(2)}$ ,  $X^{(3)}$  为特征,  $Y$  为类标记。(10 分)

	1	2	3	4	5	6	7	8	9	10
$X^{(1)}$	1	1	1	2	2	1	2	2	3	3
$X^{(2)}$	S	M	M	S	S	L	M	M	L	S
$X^{(3)}$	T	T	F	F	F	T	F	T	T	F
$Y$	-1	-1	1	1	-1	-1	-1	1	1	1