

2013 教育部-IBM 产学合作专业综合改革项目 厦门大学《数据挖掘原理及实践》课程习题

第 7 章 分类和预测

1. 请阐述贝叶斯分类方法的思路。
2. 请阐述 Fisher 线性判别方法的思想。
3. 请阐述分类和聚类的主要区别有哪些。
4. 考虑如下二元分类问题的数据集:

A	B	类标号
T	F	+
T	T	+
T	T	+
T	F	-
T	T	+
F	F	-
F	F	-
F	F	-
T	T	-
T	F	-

按照属性 A 和属性 B 划分时的信息增益。决策树归纳算法将会选择哪个属性？

5. 请阐述多层前馈神经网络与支持向量机的区别和联系有哪些。

6. 请写出混淆矩阵, 并计算准确率 (accuracy), 精度 (precision), 敏感度 (sensitivity)和特异度 (specificity) 参数。
7. 什么是过拟合现象? 可以采用哪些方法措施避免神经网络的过拟合。
8. 支持向量机分类的基本思想是什么?
9. 时间序列可分为几大类, 它们各自有哪些特点?
10. 请阐述简单移动平均法预测时间序列的步骤。
11. 考虑下表数据集:

记录	A	B	C	类
1	0	0	0	+
2	0	0	1	-
3	0	1	1	-
4	0	1	1	-
5	0	0	1	+
6	1	0	1	+
7	1	0	1	-
8	1	0	1	-
9	1	1	1	+
10	1	0	1	+

- (1) 估计条件概率 $P(A|+)$, $P(B|+)$, $P(C|+)$, $P(A|-)$, $P(B|-)$, $P(C|-)$ 。
- (2) 根据(1) 中的条件概率, 使用朴素贝叶斯方法预测测试样本(A=0, B=1, C=0) 的类标号。

(3) 根据(1) 中的条件概率, 使用朴素贝叶斯方法预测测试样本(A=1, B=1, C=1) 的类标号。

12. 给定下图所示的数据集, 解释在此数据集上, 决策数. 朴素贝叶斯是怎么样工作的。

