

重庆大学研究生《机器学习》课程试卷

2014~2015 学年 第 二 学期（春）

开课学院： 计算机 课程编号： S14010 考试日期： 20150708

考试方式： ☐ 开卷 ☒ 闭卷 ☐ 其他 考试时间： 120 分钟

题 号	一	二	三	四	五	六	七	八	考 试	平 时	课 程 成 绩
得 分											

注：1. 大标题用四号宋体、小标题及正文推荐用小四号宋体；2. 可按 A4 纸缩小打印

一、（20 分）关于机器学习方法，简要回答以下问题：

- （1）在机器学习课堂教学中，介绍了哪些机器学习方法？
- （2）将这些方法进行对比，它们有何联系？
- （3）这些机器学习方法分别具有什么特点和适用范围？
- （4）基于最小平方误差学习与贝叶斯学习有何联系？

二、（20 分）按以下要求，简述支持向量机的基本思想与算法步骤：

- （1）说明线性可分支持向量机、软最大间隔支持向量机和非线性支持向量机的概念，以及它们之间的区别与联系；
- （2）若将线性支持向量机用于分类问题，请说明训练和测试的算法步骤；

（3）请说明什么是核函数，列出至少 2 种常见核函数，写出数学表达式为优。

三、（10 分）对于布尔函数 $B_v(A \wedge C)$ ，给出表示对应的决策树。

四、（20 分）给定目标概念 EnjoySport 的 4 个训练样例，采用 ID3 算法进行决策树学习，要求写出主要过程，并画出决策树(注意： $\log_2 3 = 1.585$)。

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Sunny	Warm	High	Strong	Cool	Change	Yes
4	Rainy	Cold	High	Strong	Warm	Change	No

五（20 分）试用贝叶斯规则分析下列问题：

考虑一个医疗诊断问题，其中有两个假设：（1）病人有某种类型的癌症。（2）病人无癌症。为诊断该疾病，可用的数据来自于某化验测试，它有两种可能的输出：（+）和（-）。我们有先验知识：在所有人口中有 0.008 的人患有该疾病。另外，该化验测试只是该疾病的一个不完全预测。该测试针对确实有病的患者有 98% 的可能正确返回（+）结果，而对无该病的患者有 97% 的可能正确返回（-）结果。除此之外，测试返回的结果是错误的。

请问，假定现在有一新病人，化验测试返回了（+）结果。是否应该将病人断定为有癌症？请写出分析过程。

六（10 分）请证明，若假设空间 H 有限，且 D 为目标概念 c 的 $m \geq 1$ 个独立随机抽取的样例序列，那么对于任意 $0 \leq \epsilon \leq 1$ ，变型空间 $VS_{H,D}$ 不是 ϵ -详尽（关于 c ）的概率小于或等于：

$$|H|e^{-\epsilon m}$$

命题（组题）人：万里

审题人：何中市

命题时间：2015-06-15

研究生院制

姓名

学号

类别（学术、专业）

专业（领域）

学院

密

封

线