

模式识别与机器学习

考试时间 2021 年 12 月 27 日

整理 by 电智 2003 庄赖宏

一、判断 (30 分)

一共 10 题，主要看知识面的广度，了解算法特点就行，复习时要把边边角角都看一眼

1. 机器学习是为了得到泛化性能良好的模型？对
2. Logistic Regression 是分类
3. Isomap 算法是基于邻域保持思想试图保存临近点间的距离？对
- B 卷考的是 LLE 使线性关系在降维后的空间里继续保持，对的
4. SIFT 是将图像映射成局部特征向量集
5. 计算机视觉常用的图像特征包括点、边缘、直线、曲线等
6. 决策树只适用于分类任务？错
7. 高斯混合聚类是密度聚类？错
8. 有 A 和 B 两个分类器，A 的错误率比 B 小，A 比 B 好？错
9. 最小错误率贝叶斯是最小风险贝叶斯的一个特例
10. Bagging 构建的个体分类器之间具有强依赖性？错

二、简答 (30 分)

这部分都是文字叙述，平时上课老师提到的问题要仔细记下来

1. HOG 具有旋转不变性吗？不能的话，应怎么做？ B 卷考 SIFT 的不变性，从原理阐述。
2. 简述预剪枝的流程，并说出优缺点。 B 卷考后剪枝。
3. LDA 和 PCA 的区别和联系。
4. 对于回归问题，泛化误差可通过“偏差-方差分解”拆解，

$$E(f; D) = \text{bias}^2(x) + \text{var}(x) + \varepsilon^2$$

说明每个部分代表什么，并且对于学习器性能有什么意义。

5. 隐马尔科夫模型的三个基本问题中，解码问题的输入是什么？

6. 以下算法适用于监督学习还是无监督学习：

KNN 、 随机森林 、 Logistic Regression、SVM 、 决策树

三、计算 (22 分)

都是来自作业和课上的例题，考前一定要自己拿出来练练手

1. K-means 算法 (6 分)

直接给几个样本点，让你用算法流程算一遍

数据很简单，就是给的位置有点小，建议想好再落笔

2. 决策树，根据信息增益算法，构造决策树 (6 分)

和 PPT 上面的例题差不多，考试时记得带计算器

考前一定要自己手算一遍

3. 构建朴素贝叶斯分类器，带有拉普拉斯修正，画出查询表并给出测试数据的预测结果 (10 分)

和 PPT 上的例题一模一样，考前一定要重点看老师讲过的例题

四、综合 (18 分)

(1) SVM 核心模型

1. 给你三个图，选择是硬间隔还是软间隔、核函数，然后在图上画出来

2. 给你一个图像，对于图上标出来的几个点，判断软间隔的 ξ 范围

3. 半监督 SVM 的算法流程

(2) Adaboost 模型

1. 更新样本权重是怎么更新的

2. 学习器的权重是怎么来的

3. 从偏差-方差的角度，Adaboost 可以降低偏差，可以对泛化性能弱的学习器构造出很强的学习器，试从中归纳原因。

八、课程考核及成绩评定

1. 课程考核方式

课程目标	考核与评价方式及成绩比例 (%)			成绩比例 (%)
	课后作业	分组任务	期末考试	
课程目标 1	5	0	10	15
课程目标 2	5	10	15	30
课程目标 3	5	10	10	25
课程目标 4	5	20		
合计	20	40	40	100

序号	教学内容	学时	教学要点	教学方式	课程目标
1	第一章 绪论	4	① 了解机器学习发展沿革 ② 掌握机器学习基本术语 ③ 掌握模型评估与选择方法	① 讲授 ② 视频观看 ③ 课后作业	1
2	第二章 CV 领域特征描述子	4	① SIFT ② HOG	① 讲授 ② 视频观看 ③ 课后作业	2、3
3	第三章 线性模型	4	① 线性模型的基本形式 ② 线性回归 ③ 对数几率回归 ④ 线性判别分析 ⑤ 多分类学习	① 讲授 ② 视频观看 ③ 课后作业	2、3
4	第四章	4	① 间隔与支持向量	① 讲授	2、3

序号	教学内容	学时	教学要点	教学方式	课程目标
	支持向量机		② 支持向量机模型的对偶问题 ③ 软间隔与正则化 ④ 核函数与核方法 ⑤ 支持向量回归	② 视频观看 ③ 课后作业	
5	第五章 决策树	4	① 决策树基本流程 ② 属性划分策略 ③ 剪枝 ④ 连续值处理 ⑤ 缺失值处理 ⑥ 多变量决策树	① 讲授 ② 视频观看 ③ 课后作业	2、3
6	第六章 集成学习	4	① 个体与集成 ② Bagging 与随机森林 ③ AdaBoost 算法 ④ 结合策略 ⑤ 多样性	① 讲授 ② 视频观看 ③ 课后作业	2、3
7	第七章 聚类	2	① 无监督学习基本概念 ② 聚类任务 ③ 性能度量与聚类计算 ④ 常用聚类方法	① 讲授 ② 视频观看 ③ 课后作业	2、3
8	第八章 降维方法	4	① 降维基本概念 ② 低维嵌入 ③ 主成分分析 ④ 非线性降维 ⑤ 流形学习 ⑥ 度量学习	① 讲授 ② 视频观看 ③ 课后作业	2、3
9	第九章 半监督学习	4	① 半监督学习基本概念 ② 半监督 SVM ③ 基于分歧的方法 ④ 半监督聚类	① 讲授 ② 视频观看 ③ 课后作业	2、3
10	第十章 贝叶斯分类器	6	① 贝叶斯决策论 ② 朴素贝叶斯分类器 ③ 图模型基本概念 ④ 贝叶斯网 ⑤ EM 算法	① 讲授 ② 视频观看 ③ 课后作业 ④ 编程实训	2、3
11	第十一章 神经网络基础	4	① 前馈神经网络与反向传播算法 ② 卷积神经网络简介	① 讲授 ② 课后作业	2、3
12	第十章 机器学习方法的应用	4	① 应用问题描述 ② 研究现状 ③ 应用案例 ④ 期末总结	① 讲授 ② 分组展示	4