

湖南工业大学研究生课程教学大纲

课程编号：

编写人： 曾志高

编写日期： 2022-02-20

课程中文名称	统计机器学习				
课程英文名称	Statistical Machine Learning				
开课学期	春季	开课单位	计算机学院		
课程类别	计算机科学与技术专业、专业选修课				
主讲教师	曾志高	职称	教授	联系电话	13873392328
备讲教师	朱文球	职称	教授	联系电话	15886388595
总学时	32	其中	讲 课	28	
			实践课时	4	
学分	2	教学方式	理论教学与实践教学相结合		闭卷考试
面向专业	计算机科学与技术			考核方式	<input type="checkbox"/> 考试 <input checked="" type="checkbox"/> 考查
预修课程	概率与统计、matlab/python 编程、矩阵论、线性代数				

课程内容：

《统计机器学习》是计算机专业的一门专业课，其教学重点是使学生掌握常见机器学习算法，包括算法的主要思想和基本步骤，并通过编程练习和典型应用实例加深了解；同时对机器学习的一般理论，如假设空间、采样理论、计算学习理论，以及无监督学习和强化学习有所了解。

第一部分 决策论与信息论基础：

2 学时

- a) 损失函数、错分率的最小化、期望损失的最小化等
- b) 相对熵、互信息

第二部分 概率分布

4 学时

- a) 高斯分布、混合高斯分布、Dirichlet 分布、beta 分布等
- b) 指数分布族：最大似然估计、充分统计量、共轭先验、无信息先验等
- c) 非参数方法：核密度估计、近邻法

第三部分 回归的线性模型

6 学时

- a) 线性基函数模型

b) 贝叶斯线性回归	
c) 贝叶斯模型比较	
第四部分 分类的线性模型	6 学时
a) 判别函数：二分类和多分类的 Fisher 线性判别	
b) 概率生成模型：连续输入、离散特征	
第五部分 支持向量机	4 学时
a) 最大边缘分类器：历史回顾	
b) 用于多分类和回归的支持向量机：几何背景、各种变种	
c) 统计学习理论简介：Vapnik 等人的工作	
第六部分 演化算法	4 学时
a) 遗传算法	
b) 粒子群算法	
第七部分 特征提取	4 学时
a) PCA	
b) 其它特征提取方法	
第八部分 数据挖掘	2 学时

课程内容英文简介

Machine learning allows computers to learn and discern patterns without actually being programmed. When Statistical techniques and machine learning are combined together they are a powerful tool for analysing various kinds of data in many computer science/engineering areas including, image processing, speech processing, natural language processing, robot control, as well as in fundamental sciences such as biology, medicine, astronomy, physics, and materials.

Introduction to Statistical Machine Learning provides a general introduction to machine learning that covers a wide range of topics concisely and will help you bridge the gap between theory and practice.

课程教学目标及重点、难点：

本课程的任务是要使学生掌握统计机器学习的基本概念、基本原理、实用的算法，了解机器学习领域的发展动向，如何用统计机器学习的方法解决实际问题，

在解决问题的过程中应遵循的流程、准则和规范。本课程是一门实践性很强的课程，它是各种机器学习经验的总结与提炼，学生不但应注意概念、原理、方法、技术的掌握，也应该注重方法、技术的实际应用。本课程教学难点为机器学习算法的应用。

教学要求：

1. 决策论与信息论基础：了解并掌握统计决策理论和信息论的基础知识。
2. 熟悉常见的分布，熟悉掌握最大似然估计方法，学会利用信息先验和共轭先验简化计算，了解一些常用的非参数方法。
3. 掌握线性回归的一般方法，学会使用 matlab 中有关回归的程序包，并能将之用于解决实际问题。
4. 对分类问题有一个全面的了解，掌握一些常用的分类方法。
5. 掌握支持向量机的基本原理，面对各自研究领域中的具体问题学会使用支持向量机，粗略地了解统计学习理论。
6. 了解常见演化算法，如遗传算法和粒子群算法的工作原理，及其在解决实际问题中的应用。
7. 掌握主成份分析的特征提取方法，了解其它特征提取方法，如傅立叶描述子，纹理特征提取等等。
8. 掌握数据挖掘常用方法，如关联规则等，对大数据问题进行需求获取、分析、建模。
9. 了解机器学习中的高级课题，把握机器学习的发展动态和方向。

教材及主要参考书：

- [1] 周志华著. 机器学习[M]. 清华大学出版社, 2016.
- [2] Addisson Salazar,Luis Vergara,Jorge Igual. Statistical Pattern Recognition[M]. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2013.

大作业：

根据实际问题，选用一种机器学习方法来解决它。