

课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	统计学习的基本原理 (Introduction to Statistical Learning)				
2.	授课院系 Originating Department	数学系 (Mathematics)				
3.	课程编号 Course Code					
4.	课程学分 Credit Value	32				
5.	课程类别 Course Type	专业选修课 Major Elective Courses				
6.	授课学期 Semester	夏季 Summer				
7.	授课语言 Teaching Language	中英双语 English & Chinese				
8.	授课教师、所属学系、联系方式 (如属团队授课, 请列明其他授课教师) Instructor(s), Affiliation & Contact (For team teaching, please list all instructors)	彭衡 (PENG Heng), 香港浸会大学数学系 (Department of Mathematics, Hong Kong Baptist University)				
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	无 NA (请保留相应选项 Please only keep the relevant information)				
10.	选课人数限额(可不填) Maximum Enrolment (Optional)					
11.	授课方式 Delivery Method	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other (Please specify)	总学时 Total
	学时数 Credit Hours	32				32

12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	概率论与数理统计 或者 概率论 Probability and Statistics or Probability Theory
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	
14. 其它要求修读本课程的学系 Cross-listing Dept.	

教学大纲及教学日历 SYLLABUS

15. **教学目标 Course Objectives**

统计学习是指的基于统计原理针对复杂数据建模，分析解释和预测的一类统计方法。 这些方法混合和伴随着计算机科学，特别是机器学习的发展。随着大数据时代的到来，统计学习的方法和理论正被推广和应用到科学研究，商务市场分析，以及金融行业等各行业进行数据分析和信息挖掘。 本次课程的主要目的是介绍统计学习方法的基本统计原理和思想，以及概述几种常见的统计学习方法，包括稀疏回归，分类，随机森林，boosting 和向量机等具体实现及应用。

Statistical learning refers to a class of statistical methods based on statistical principles for complex data modeling, analytical interpretation and prediction. These methods are mixed and accompanied by the development of computer science, especially machine learning. With the arrival of large data age, statistical learning methods and theories are being promoted and applied to scientific research, business market analysis, and financial industry and other industries for data analysis and information mining. The main purpose of this course is to introduce the basic statistical principles and ideas of statistical learning methods, and to outline several common statistical learning methods, including sparse regression, classification, random forest, boosting and vector machine, and their realization and applications.

16. **预达学习成果 Learning Outcomes**

通过课程的学习，学生可以掌握统计学习方法的基本原理，能够应用统计软件 比如 R，实现基本的统计学习算法，进行有效的数据分析和信息挖掘。

Through the course study, students can master the basic principles of statistical learning methods, can apply statistical software such as R, implement basic statistical learning algorithms, and conduct effective data analysis and information mining.

17. **课程内容及教学日历**（如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）
Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

- 1、统计学习的基本介绍（2 学时），什么是统计学习。
实验课（2 学时），介绍统计学习软件 R 及其编程语言。
- 2、回归模型在统计学习的应用（2 学时）。
实验课（2 学时），讲解如何使用统计学习软件 R 基于回归分析进行统计学习。
- 3、分类判别算法在统计学习的应用（3 学时）。
- 4、再抽样方法在统计学习中的应用（2 学时）。
实验课（2 学时），讲解如何使用统计学习软件 R 对数据进行分类判别， 以及如何进行在抽样方法进行推断。
- 5、正则方法在线性回归模型中的应用（2 学时）。
实验课（2 学时）， 讲解如何使用统计学习软件 R 如何实现正则方法的应用。
- 6、非参数回归方法（3 学时）。
- 7、基于回归树的统计学习方法（2 学时）。
实验课（2 学时），讲解如何使用统计学习软件 R 如何实现非参数回归方法的应用，基本基于回归树的 R 软件包的应用。
- 8、支持向量机的简单介绍（2 学时）。
支持向量机在统计学习软件 R 中的实现及应用（2 学时）。
- 9、非监督统计学习方法的介绍（3 学时）。
 - 1.Basic introductions to statistical learning (2 Credit Hours), and what is the statistical study.
Lab lessons (2 Credit Hours), introduction of statistical learning software R and its programming language.
 - 2.The application of regression models in statistical learning (2 Credit Hours).
Lab lessons (2 Credit Hours), explain how to use statistical learning software R to perform statistical learning based on regression analysis.
 - 3.The application of classification and discriminant algorithms in statistical learning (3 Credit Hours).
 - 4.The application of resampling methods in statistical learning (2 Credit Hours).
Lab lessons (2 Credit Hours), explain how to use the statistical learning software R to classify and discriminate data, and how to make inferences by the resampling methods.
 - 5.The application of the regular methods in the linear regression model (2 Credit Hours).
Lab classes (2 Credit Hours), explain how to use the statistical learning software R to achieve the application of the regular methods.
 - 6.Nonparametric regression method (3 Credit Hours).
 - 7.The statistical learning methods based on tree approaches (2 Credit Hours).
Lab classes (2 Credit Hours), explain how to use the statistical learning software R to achieve the application of nonparametric regression methods, and the application of the regression tree methods by R packages.
 - 8.A brief introduction of the support vector machine (2 Credit Hours).
Lab classes (2 Credit Hours), the implementation and application of support vector machine in statistical learning by software R.
 - 9.Introduction the unsupervised statistical learning methods (3 Credit Hours).

18. 教材及其它参考资料 **Textbook and Supplementary Readings**

Text: An Introduction to Statistical Learning (统计学习导论), Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

课程评估 ASSESSMENT				
19. 评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance				

课堂表现 Class Performance				
小测验 Quiz				
课程项目 Projects				
平时作业 Assignments		40		
期中考试 Mid-Term Test				
期末考试 Final Exam				
期末报告 Final Presentation		60		
其它（可根据需要 改写以上评估方 式） Others (The above may be modified as necessary)				

20. 记分方式 **GRADING SYSTEM**

<input checked="" type="checkbox"/> A. 十三级等级制 Letter Grading <input type="checkbox"/> B. 二级记分制（通过/不通过） Pass/Fail Grading

课程审批 **REVIEW AND APPROVAL**

21. 本课程设置已经过以下责任人/委员会审议通过
This Course has been approved by the following person or committee of authority

--

