DATA IV Introduction to Statistical Learning
TERM INFORMATION
INSTRUCTOR NAME
University of Kansas

Contents

1 Preliminary Details

1.1 Course Information

Course Number: PSYC 610
Course Name: Introduction to Machine and Statistical Learning
Location: TBD
Time: TBD
Start/End Date: TBD
Final Exam Time: TBD

1.2 Professor Information

Professor: TBD
Office: TBD
Office Hours: TBD
Email: someone@ku.edu

2 Course Description, Objectives and Textbook

2.1 Course Objectives

Course Description:
Large-scale data sets ("big data") have become ubiquitous across many applied areas. The goal of this course is to provide an introduction to methods that are well-suited to deal with these big data sets. We focus on statistical learning techniques and high-dimensional statistics, and show how they can be applied in context-specific domains. The class starts with a review of data management and data visualization, before considering the first concepts in the textbook. As the course moves through the textbook, we consider regularization and data-driven choices for tuning parameters. We will motivate shrinkage estimators in different ways, and will prove the famous result that shrinkage estimators can uniformly dominate conventional estimators. Students will learn how to program statistical methods in Python, as well as how and when to use the common libraries.

Course Objectives/Student Learning Outcomes:
Upon successful completion of this course, students will be able to:

- describe a wide array of the fundamental methods in modern machine and statistical learning

- use those techniques to carry out basic analyses on real data

- create meaningful results and visual illustrations

- compare and contrast different algorithms

- analyze and discuss the meaning of their codes in Python

- become informed and critical consumers of machine learning research

Prerequisite: Data III (PSYC 599)

2.2 Textbook

Gareth, James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning.* Open Access Online.
The instructor may provide readings through Canvas.

2.3 Technology

We will use computational and statistical programming during this class. We will base our textbook lab demonstrations in Python and R, but you can also use Stata, Matlab, or other (even Excel if you are feeling adventurous!) for the homework assignments and final project.

3 Assignments

Workshops: 400 points. There will be 8 workshops throughout the course:

- Chapter 2: Statistical Learning

- Chapter 3 and Chapter 4: Linear Regression and Classification

- Chapter 5: Resampling Methods

- Chapter 6: Linear Model Selection and Regularization

- Chapter 7: Beyond Linearity

- Chapter 8: Tree-Based Methods

- Chapter 12: Unsupervised Learning

- Chapter 13: Multiple Testing

The text of the workshop will be provided on Canvas at the appropriate time. Completing each workshop for submission will earn you up to 50 points. Grading will assess for comprehensiveness and accuracy of the submission.

Exam I: 100 points. The first exam will cover chapters 2 through 6. The exam will be in-person. You will have the full class period to complete the exam.

Exam II: 100 points. The second exam will cover chapter 7, chapter 8, chapter 12, and chapter 13. The exam will be in-person. You will have the full class period to complete the exam.

Final Exam: 200 points. The final exam is comprehensive and covers all chapters discussed during the course of the semester.

Final Data Project: 150 points. Information about writing the final paper will be given in the course. The final paper will describe all parts of the research project. It will include a separate summary of the entire project (abstract), followed by an introduction, data section, result section, and

conclusions. The length of the paper will be 4000-5000 words.

The paper will be prepared in the software of your choice. Good options would be Microsoft Word, a data analysis software like RMarkdown or Stata's dyndoc, or script editing software such as LaTeX. References should be in APA format.

Final Presentation: 50 The presentations will take place in Week 16. A presentation rubric will be provided in the course.

3.1 Grades

Grades on assignments are final. If you feel there is an error in your grade, you must submit a request to me in writing to re-evaluate your assignment. If the error is arithmetic, your grade will be updated without further action. If the error is of another kind, the professor reserves the right to re-grade the entire assignment.

3.2 Grading Scale

points	percent	letter grade
900 – 1,000 points	90-100	A
800 – 899 points	80-89	B
700 – 799 points	70-79	C
600 – 699 points	60-69	D
0 - 599 points	<59	F

4 Attendance Policy

Class attendance, whether in the classroom or in a remote environment, is vital to your academic success. You are expected to attend each class session. If you consistently miss class, it will adversely affect your grade. If you miss more than three class periods, your grade is reduced by five percentage points. Each additional absence will lower your grade by an additional percentage point.

Some other important policies to note:

- Attendance is expected. However, sometimes missing class is unavoidable. Students are responsible for obtaining missed lecture notes from their classmates. I expect that you arrive to class on time and not depart early. I also request that all cell phones and pagers are turned off for the duration of the class.

- Any student with a disability that may preclude full course participation should contact the instructor in order to discuss accommodations.

- Any student who plans to observe a religious holiday that conflicts in any way with this course should contact the instructor in order to discuss accommodations.

5   Academic Misconduct

The University community traditionally has been a place where all members are free to express and exchange ideas. Such fundamental goals of the University as intellectual growth and development are predicated on honest investigation, straightforward expression of views and opinions, and genuine dialogue. The attainment of these goals requires that all who participate in the exchange of ideas maintain intellectual integrity. The University of Kansas seeks to ensure that both the professor and the student are protected from unfair actions or accusations in cases of cheating and plagiarism. The University encourages faculty and students to adopt a responsible attitude toward one another.

Academic misconduct includes but is not confined to plagiarizing; cheating on assignment and assessments; turning in counterfeit reports, tests, and papers; stealing tests and other academic material; knowingly falsifying academic records or documents; and turning in the same work to more than one class. Students and faculty alike must recognize that none of the procedures set forth in this document operate to the exclusion of civil or criminal litigation. Likewise, no definitions in this document supersede any parties concerned to resolve the contested issues without the necessity for recourse to the law in a manner that protects the rights of the individuals involved.

Consequences of academic misconduct may include, but are not limited to, a failing grade for an assignment, a failing grade for a course, or expulsion from the University. Any form of academic misconduct which results in administrative or academic withdrawal is noted on the student's transcript.

Collaboration Policy: Discussion and the exchange of ideas are essential to academic work. You are encouraged to study with your classmates. However, you should ensure that any work you submit for evaluation is the result of your own efforts and that it reflects your own ability. You must also adhere to standard citation practices and properly cite any books, articles, websites, lectures, etc. that have helped you with your work (for instance, if you used something besides course materials to complete a homework assignment). If you received any help with your work (e.g., tutoring or feedback from someone besides the professor), you must also acknowledge this assistance.

6   Instructor's Expectations

Late work: Late homework can be made up by turning in an assignment one week later with a deduction of one letter grade from the total points of the assignment.

Missed work: Homework assignments that are missed due to illness, family emergency, or university-sponsored event can be made up one week later with no deduction of points. Please alert the instructor well in advance if travel conflicts with class. If you do not provide one of these two excuses in advance of the class period, the late work penalty will apply.

Classroom management: This course is run with expectations of professional conduct. If at any time, the instructor deems that a student is not comporting him or herself in this manner, the student will be asked to leave the class for that day and will be marked absent for the time away from class.

7   Credit Hour Definition

Consistent with best practices in higher education, the University of Kansas subscribes to the federal definition of the "credit hour" endorsed by the Higher Learning Commission. Driven by intended learning outcomes and verified by evidence of student achievement, the "credit hour" is an institutionally-established equivalency that reasonably approximates not less than one hour of the classroom (or direct faculty) instruction and a minimum of two hours of out-of-class student work per week for the duration of the course enrollment period. A 3-credit-hour course, for example, requires approximately 40-45 instruction (or instructional equivalency) hours, roughly 80-90 out-of-class work (self-directed) hours and approximately 120-135 total instructional hours over the course of the semester.

8   ADA Statement

The Academic Achievement and Access Center (AAAC) coordinates academic accommodations and services for all eligible KU students with disabilities. If you have a disability for which you wish to request accommodations and have not contacted the AAAC, please do so as soon as possible. They are located in 22 Strong Hall and can be reached at 785-864-4064 (V/TTY). Information about their services can be found at http://www.disability.ku.edu. Please contact me privately concerning your needs in this course.

9   Statement on Diversity and Inclusion

As a premier international research university, the University of Kansas is committed to an open, diverse, and inclusive learning and working environment that nurtures the growth and development of all. KU holds steadfast in the belief that an array of values, interests, experiences, and intellectual and cultural viewpoints enrich learning and our workplace. The promotion of and support for a diverse and inclusive community of mutual respect require the engagement of the entire university.

10   Weapons Policy

Individuals who choose to carry concealed handguns are solely responsible for doing so safely and securely in strict conformity with state and federal laws and KU weapons policy. Safety measures outlined in the KU weapons policy specify that a concealed handgun:

- Must be under the constant control of the carrier.

- Must be out of view, concealed either on the carrier's body, or in a backpack, purse, or bag that remains under the carrier's custody and control.

- Must be in a holster that covers the trigger area and secures any external hammer in an un-cocked position.

- Must have the safety on and have no round in the chamber.

11   Course Schedule

NOTE: the course schedule is subject to change to accommodate course content and progress on community partner projects.

Day	Concept/Activity	Reading	Project Timeline
Week 1			
Day 1	Introduction and Syllabus		
Day 2	Final Project Overview		
Week 2			
Day 1	Data Management		
Day 2	Data Visualization		Finalize final project selection
Week 3			
Day 1	Statistical Learning	Chapter 2	
Day 2	Statistical Learning		Clearly defined project
(Due Tuesday)	**Workshop 1: Chapter 2**		
Week 4			
Day 1	Linear Regression	Chapter 3	
Day 2	Linear Regression		Describe dataset and project
Week 5			
Day 1	Classification	Chapter 4	
Day 2	Classification		Complete data wrangling
(Due Tuesday)	**Workshop 2: Chapters 3 & 4**		
Week 6			
Day 1	Resampling Methods	Chapter 5	
Day 2	Resampling Methods		
(Due Tuesday)	**Workshop 3: Chapter 5**		
Week 7			
Day 1	Linear Model Selection and Regularization	Chapter 6	
Day 2	Linear Model Selection and Regularization		Complete data exploration
(Due Tuesday)	**Workshop 4: Chapter 6**		
Week 8			
Day 1	Review Day		
Day 2	Exam I		

Day	Concept/Activity	Reading	Project Timeline
Week 9	SPRING/FALL BREAK		
Week 10			
Day 1	Moving Beyond Linearity	Chapter 7	
Day 2	Moving Beyond Linearity		
(Due Tuesday)	**Workshop 5: Chapter 7**		
Week 11			
Day 1	Tree-Based Methods	Chapter 8	
Day 2	Tree-Based Methods		
(Due Tuesday)	**Workshop 6: Chapter 8**		
Week 12			
Day 1	Unsupervised Learning	Chapter 12	
Day 2	Unsupervised Learning		
(Due Tuesday)	**Workshop 7: Chapter 12**		
Week 13			
Day 1	Multiple Testing	Chapter 13	
Day 2	Multiple Testing		Finalize model, analyses, and plots
(Due Tuesday)	**Workshop 8: Chapter 13**		
Week 14			
Day 1	Review		
Day 2	Exam II		
Week 15			
Day 1	Work Day		Presentation and report preparation
Day 2	Thanksgiving		
Week 16			
Day 1	In-Class Presentations		
Day 2	In-Class Presentations		
(Due Friday)	**Final Project Paper**		
Finals Week			
	Final Exam		