

# STAT 4360: Introduction to Statistical Learning

Fall 2022

**Instructor:** Jiayi Wang (jiayi.wang2@utdallas.edu)

**Time:** Monday & Wednesday 1:00pm - 2:15pm

**Location:** GR 3.302

**Office Hours:** Monday & Wednesday 2:30pm - 3:30pm or by appointment

**Office Location:** PHYS 1.902

**TA:** Nisansala Wickramasinghe (wickramasinghe.wickramasinghe@utdallas.edu)

**TA office hours:** Tuesday & Thursday 4:00pm - 5:00pm or by appointment

**TA office Location:** FN 3.118

## Course Overview

### Description

Statistical and machine learning methods will be taught at an introductory level. We will cover supervised learning methods including linear regression; classification; cross-validation; bootstrap; model selection and regularization; tree-based methods; and support vector machines. We will also learn some unsupervised learning methods including principal components analysis and clustering. Computer packages such as R or Python will be used for implementation of methods and data analysis.

### Prerequisites

STAT 3355 or equivalent.

### Learning Objectives

1. Understand basic statistical and machine learning methods.
2. Analyze data using basic statistical and machine learning methods.
3. Use software packages (R or Python) for data analysis.

## Textbook and Resources

- **Textbook:**

- Required: *An Introduction to Statistical Learning with Applications in R* by James, Witten, Hastie, and Tibshirani. Visit the book's website [https://hastie.su.domains/ISLR2/ISLRv2\\_website.pdf](https://hastie.su.domains/ISLR2/ISLRv2_website.pdf) to download a PDF of the book and other related materials.

- Others:

1. *The Elements of Statistical Learning* by Hastie, Tibshirani, and Friedman. Visit <https://web.stanford.edu/~hastie/ElemStatLearn> for a PDF of the book and other related materials.
2. *Machine Learning: A Probabilistic Perspective* by Murphy, MIT Press, 2012.
3. *An Introduction to R* (<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>)

4. *Advanced R* (<http://adv-r.had.co.nz/>).

5. *Intro to Machine Learning with Python* (<https://python-course.eu/machine-learning/>).

- **Course Website:** Use eLearning ([elearning.utdallas.edu](http://elearning.utdallas.edu)) to access lecture notest, homework assignments and their solutions, the solutions of quizzes and exams, and to check your performance.
  - The course materials are available to all students registered for this class as they are intended to supplement the classroom experience. These materials are for registered students' use only. Classroom materials may not be reproduced or shared with those not in class, or uploaded to other online environments except to implement an approved Office of Student Access Ability accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct (<https://policy.utdallas.edu/utdsp5003>).

## Grading Policy

Course grade will be assigned as  $A^+$ (96-100),  $A$ (93-95),  $A^-$ (90-92),  $B^+$ (87-89),  $B$ (84-86),  $B^-$ (80-83),  $C^+$ (75-79),  $C$ (70-74),  $C^-$ (65-69),  $D^+$ (60-64),  $D$ (55-59),  $D^-$ (50-54) or  $F$ (0-49) with components and weights as follows.

\* These cutoffs may be slightly adjusted at the end of the semester.

### 1. Projects (40%)

- Six mini projects will be assigned during the semester.
- The projects will need to be submitted in class on the due date. They may not have equal weight.
- No late work will be accepted. Students who miss a homework due to an emergency or illness must provide appropriate documentation in a timely manner. Work submitted by a student as makeup work for an excused absence is not considered late work and is exempted from the late work policy.

### 2. Exams: Exam 1 (25%); Exam 2 (35%)

- The tests will be based on a specified range of course content to be announced in the class. They will involve multiple-choice problems. There may be one or two problems which require mathematical derivation. The use of a calculator is required.
- The tests are closed-book, closed-notes. A cheat sheet may be provided if deemed necessary.
- No make-up exam will be given unless there is a serious medical emergency and appropriate documentation is provided in a timely manner. It may be possible to take it early for reasons such as travel.

### 3. Homework (0%)

- Homework will be assigned but will not be collected or graded. However, students are strongly encouraged to go over the excercises in Homework, which helps in gaining mastery over the course material. Exams problems will have similar essence as the problems in the homework.

Incomplete grade may be possible only in the case of a documented serious medical emergency near the end of the semester, with at least 70% of work completed at an on-going passing grade.

## Tentative Course Schedule

Week	Topic	Textbook Chapters
1-3	Statistical Learning & Linear Regression	1-3
4-6	Classification & Resampling	4-5
7-9	Regularization & Non-linear regression	6-7
10-12	Tree-based methods & Support Vector machines	8-9
13-15	Unsupervised learning (and Deep learning if time permits)	12, 10

## Tentative Exam and Project Schedule

Due date	Item
Sep 7	Project 1
Sep 21	Project 2
Oct 12	Exam 1
Oct 19	Project 3
Nov 2	Project 4
Nov 16	Project 5
Nov 30	Project 6
Dec 7	Exam 2

## Other Policies

### Class Attendance

Regular and punctual class attendance is expected. Students who fail to attend class regularly are inviting scholastic difficulty.

### Class Participation

Regular class participation is expected. Students who fail to participate in class regularly are inviting scholastic difficulty. A portion of the grade for this course is directly tied to your participation in this class. It also includes engaging in group or other activities during class that solicit your feedback on homework assignments, readings, or materials covered in the lectures (and/or labs). Successful participation is defined as consistently adhering to University requirements, as presented in this syllabus. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

### Class Recordings

Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

The instructor may record meetings of this course. These recordings will be made available to all students registered for this class if the intent is to supplement the classroom experience. If the instructor or a UTD school/department/office plans any other uses for the recordings, consent of the students identifiable in the recordings is required prior to such use unless an exception is allowed by law.

### Academic Integrity

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

“As a Comet, I pledge honesty, integrity, and service in all that I do.”

You can learn more about academic integrity, and your rights and responsibilities at <https://conduct.utdallas.edu/integrity/>.

### Academic Support Resources

The information contained in the following link lists the University’s academic support resources for all students. Please go to <http://go.utdallas.edu/academic-support-resources>.

## **UT Dallas Syllabus Policies and Procedures**

The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus. Please review the catalog sections regarding the credit/no credit or pass/fail grading option and withdrawal from class.

Please go to <http://go.utdallas.edu/syllabus-policies> for these policies.