

COGS 105: Introduction to Statistical Learning

Professor Ramesh Srinivasan, Cognitive Sciences

Version 0 09/25/25

Course meetings: TTh 2:00-3:20

E-mail: r.srinivasan@uci.edu

Discord: rameshUCI

Office Hours: **By Appointment**

Office: SBSG 2336

Course Description

Introduction to statistical learning as a tool for scientific research. Topics include linear regression, logistic regression, mixture models, maximum likelihood estimates, model selection, classification, cross-validation, ridge and Lasso regularization, decision trees, ensemble methods, random forests, unsupervised learning, clustering, and dimensionality reduction.

Short form: This is machine learning, exclusive of neural networks. What distinguishes the content of this course from most machine learning courses is the emphasis on interpretation of models, which is essential for scientific applications.

Prerequisites

MATH 2A/B

COGS 10A/B/C

COGS/PSYC 14P or ICS 31 or MATH 9

If you are missing a prereq and need some help, just reach out.

Homework and Tests

There will be homework that incorporates math problems and data analysis. The level of math required is PreCalc, and (possibly) a derivative.

They will have reasonable lead times, so you should be able to do them by the deadline. I

strongly discourage requesting extensions on Homework because it prevents me from releasing solutions.

You are allowed to discuss Homework with your classmates. You are not to discuss Tests with your classmates.

In order to submit your work, you will create a GitHub repository which is **Private** and add me as a collaborator. My Github username is rameshshrinivasanuci

Data Analysis Project

A major component of this course (and your grade) is a data analysis project. The goal of this project is to take the methods developed in the course and apply them to new data and to interpret the results of these analysis in a final report, submitted at the end of the quarter.

Data Sets

I will provide some options for data sets that you might use for a final project.

However, I recognize that my interests are not the same as yours, so **I am quite open to a final project that you define**. This may come from data sets that you have access to because of your participation in research on campus. It may also come from data sets you identify from a search through data archives such as kaggle.com or openneuro.org.

If you choose to use your own data, you must obtain approval for your proposed data set. The data set must be sufficiently rich and complex that you can do something interesting with it. If the data originates from active research, you must also ask for permission to use the data for this project and for me to have access to the data.

Scope of Work

The final project should address the goals of your project by using visualization and data analysis. An important goal is *designing data analysis that has predictive and explanatory power*.

Collaboration

You are allowed to collaborate with others to discuss data analysis. However, I expect to see individual data analysis notebooks and individually written final reports.

Project Proposal

The first assignment will be to define the scope of work for your project and to get it approved by me. This will be submitted as a written project proposal (1 page).

Part of the goals of this course is to acquire the skills to identify *What are the important questions to answer for my data set?* In addition to identifying the analysis that you would like to do

with the data set, your proposal should explain what you believe you will learn from the analysis.

A good thing to think about is *If everything goes perfectly, what will my final plots look like?*

Final Report

The final report should be submitted in the form of a short paper (3 pages of text + as many figures as needed) that provides:

- a brief background and explanation of the data.
- an explanation of the objective of the analysis.
- an explanation of the meaning of the graph or result.
- discussion of what your analysis learned and what additional data/analysis would be useful.

A Github repository for your project should be created. The repository should be self-contained and I should be able to run your code and produce all the figures and quantitative estimates that you present in your final report. I will be grading the clarity of your notebook.

Final Presentation

We might do it in person during Finals week if time permits. Or I'll have you make a TikTok ☺. The main idea is for you to learn how to make a brief but complete presentation (FlashTalk). Title + 3 slides and in under 5 minutes should tell us a story.

Software

Anaconda Python

All of the exercises and projects in this class make use of the Python programming language. Install Anaconda Python Libraries on your (Windows, Mac, or Linux) computer.

Anaconda Python - <https://www.anaconda.com/products/individual>

Why Anaconda Python?

Your computer may already have Python installed on it, which in principle you could configure to use for this class. It is much easier to simply install the Anaconda distribution of Python which is free and comes with an nearly complete library of software for scientific computing as well as a number of other useful tools like the *Jupyter Lab* IDE.

VS CODE

My preferred IDE for Jupyter Notebooks is Visual Studio Code.

<https://code.visualstudio.com/>

There is no requirement to use VS Code. You can use Jupyter Notebooks or Jupyter Lab, if you are already used to them and like them. Some people like PyCharm. All of the materials in this class should work in any IDE.

Visual Studio Code has really good GitHub integration.

Course Structure

Final Exam

There is no Final Exam for this course. **Final Project is due on Thursday Dec 14, 5 pm.**

Grading Policy

Homework and Tests: 50% of grade.

Final Project Proposal: 5% of grade

Final Project Repository/Code: 20% of grade Final Project Paper/Presentation: 25% of grade.

Use of AI

If you want to use AI tools to help you do the work in the class, I think it's reasonable to use it as an Assistant and a search tool. Its always reasonable to search online for explanations to concepts, and AI tools are quite good at that. *Be aware they are often wrong.*

It is not acceptable to simply dump Homework or Test questions into an AI tool (e.g., Chat-GPT or Gemini), and submit the output. I will know. If you use these tools in your assignments, you need to acknowledge exactly how they were used. Without an accurate acknowledgement, I will not accept your submission, and record a 0 for the assignment.

Disability services, academic dishonesty, and copyright policy

Disability Services link: <https://dsc.uci.edu/>

Academic Dishonesty link: <https://aisc.uci.edu/students/academic-integrity/index.php>

Copyright policy link: <http://copyright.universityofcalifornia.edu/use/teaching.html>