INFO251 – Applied Machine Learning

Lab 1

Suraj R. Nair, Satej Soman

Today's Agenda

- ~15 minutes: Course logistics
  - About the TAs
  - Lab goals and structure
  - Participation
  - Office hours
- ~35 minutes: statistical coding:
  - intro to using numpy and statsmodels for simulation and inference
  - Download Lab 1 materials under **files** tab on bCourses, or check out the course Github repository

About the TAs:

- Suraj R. Nair
  - 6th year PhD candidate at the School of Information
  - Background in development studies / economics
  - Development economics + Machine learning
  - Website: surajrn.github.io
- Email: suraj.nair@berkeley.edu
- Office Hours : Wednesday, 10:30am – 12pm, South Hall 107

About the TAs:

- satej soman (*suh-THAYJ SOH-munn)*
  - 4th year PhD student at the School of Information
  - background: condensed-matter physics / software engineering / policy analysis
  - using ML + GIS to study inequality, urbanization, and infrastructure
  - ischool profile | website
- Email: satej@berkeley.edu
- Office Hours : Thursday, 12:45pm – 1:45pm (right after class, with a buffer to walk over to South Hall 107)

Course Logistics / Announcements

- All labs and problem sets will be uploaded to bCourses, in the files tab
- The lab and problem set schedule for the semester is on the bCourses page
- Prioritize Ed for all class-related communications
  - Posts about logistics, etc. can be made private to the course staff
  - **The course staff will not respond to course-related inquiries on the ISchool Slack**
- Fill out the background survey by Jan 24:
  - See the quizzes tab in bCourses.
- Problem Set 1 is due on Jan 28 (8am)

Participation

- Participation counts for 4% of your grade.
- Ways to participate:
  - Attend (and ask/answer questions in) lectures
  - Attend (and ask/answer questions in) labs
  - Come to office hours
  - Answer questions on Ed
- Ed note:
  - Please read other related questions before asking your own.
  - We will do our best to respond to every question on Ed, BUT
  - Please do not post urgent HW-related questions at the last minute and expect a quick response!

A Typical Lab Session

- **First 5-15 minutes**: Review key themes and concepts from lecture and introduce lab goals.
- **Next 30-40 minutes**: Structured lab exercise
  - Will work on a Jupyter notebook
  - Some labs may contain exercise questions that you will work on together in a group (not today)
  - From time to time will walk through specific questions and solutions
- **Last 5-10 minutes**: Open Q&A, which can turn into office hours

Goals of Lab Session

- To give you practical, hands-on experience implementing the concepts discussed in class and in the readings.
- To prepare you for the problem sets.
- To help you meet fellow students and find potential study partners.
- To answer practical questions about applied machine learning.

Other Notes About Lab

- We will take attendance for the first two lab sections (please fill out the sign-in sheet!)
- You will not turn in assignments from lab – though you are welcome to complete them on your own.
- All lab material (slides, notebooks) will be uploaded to bCourses in the **files** tab.

First Three Lab Topics

- **On your own:** check out Lab 0 to make sure you're familiar with Python, pandas, and matplotlib (prerequisites)
- **Today**: Simulation and inference using numpy and statsmodels
- **Next Week**: Regression and hypothesis testing (Jan 22)
- **Two Weeks from Now**: numpy, vectorized computation, computational efficiency (Feb 29)

Office Hours

- **Instructor office hours:** 9:30am-10.30am on Tuesdays
  - **Where:** South Hall, 207 C.
  - **What:** Conceptual questions, logistical questions
- **Suraj office hours:** 10:30am – 12pm, Wednesdays
  - **Where:** South Hall, 107
  - **What:** Conceptual questions, questions from labs, questions on problem sets, logistical questions
- **Satej office hours:** 12:45 – 1:45 pm, Thursdays
  - **Where:** South Hall, 107
  - **What:** Conceptual questions, questions from labs, questions on problem sets, logistical questions

Today's Lab Topics

- Simulating random variables and drawing from probability distributions
- Understanding where OLS normal equations come from
- Dipping your toes into using statsmodels (more next week!)

source