NANOx81R Data Science in Materials Science
Course Admin

Shyue Ping Ong

Aiiso Yufeng Li Family Department of Chemical and Nano Engineering
University of California, San Diego
http://materialsvirtuallab.org

To provide a comprehensive introduction into the application of data science to materials science.

What will you learn in this course?

- Enough data science (including the mathematics) to understand how to apply them to solve materials science problems.
- Best practices in using various data science techniques.
- Practical use of open-source Python packages to do data science in materials science.

What this course is NOT

- A probability and statistics course.
- A replacement for a rigorous data science course.
- A replacement for a materials science course.

Course Plan

- Weeks 1 and 2: Introduction to Data Science, Python and Data Wrangling
- Week 3: Lab 1 - Python for Data Science and Data Wrangling
- Weeks 4 and 5: Linear Methods and Unsupervised Learning
- Week 6: Lab 2 - Linear methods and clustering
- Weeks 7 and 8: Kernel Methods, Trees and Neural Networks
- Weeks 9 and 10: Final Lab (Kaggle competition)

Instructors

- Lecturer: Shyue Ping Ong (ongsp@ucsd.edu)
- Teaching Assistant: Sojung Koo (s2koo@ucsd.edu)

Recommended Textbooks (All Free)
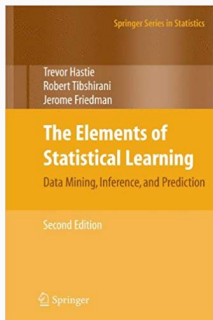
Figure 1: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition [Amazon][Free PDF]
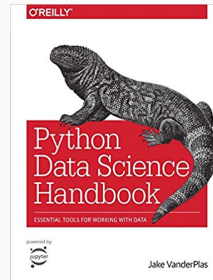
Figure 2: Python Data Science Handbook [Amazon] [Free web version]

Course Structure

- Lectures/Labs (Tues/Thurs @ 930-1050).
- Note: Please ignore scheduled lab sessions - all labs are held in lecture times, not in a separate time.
- Recordings will be available online after the class.
- **Please bring your laptops.**
- Grading:
    - Lab 1: 25% - Same for NANO181R and NANO281R
    - Lab 2: 25% - Same for NANO181R and NANO281R
    - Lab 3: 50% - Different for NANO181R and NANO281R

Model performance	30%
Materials Science Insights	30%
Data Science Technique	30%
Programming Style	10%

Class policies

- **Collaboration**: Working together is highly encouraged, but each student must submit his / her own work.
- **Use of AI**: Despite being an course in AI/ML techniques, use of AI tools such as ChatGPT or similar is **strongly discouraged**. You need to work through the exercises, including the mistakes and iterations, to learn the concepts. The instructors will use AI detection tools on your labs - we know what an AI-generated answer looks like. If the problem is severe enough, all credit for Labs 1 and 2 will be zeroed out, and all evaluation will be done only on the final lab, which is open-ended and therefore not amenable to AI cheating.
- Remember that you are here to learn an essential skill.

Class etiquette

- Interaction preferred - interruptions with questions highly encouraged.
- Please be punctual. Lectures will start on time.
- Use of laptop to follow class examples is encouraged, but please be respectful of your lecturer and classmates by not using devices for non-class applications. All devices must be on silent mode.

Prerequisites

- Knowledge of basic statistics (e.g., Gaussian distribution, Bayes theorem, etc.).
- Knowledge of basic linear algebra (e.g., matrix multiplication, eigenvalue decomposition, inverse).
- Some programming experience. Ideally, experience in the Python programming language would be helpful.
- 1st Homework (ungraded):
  1. Go to Google Colab.
  2. Create a new notebook.
  3. Go through items 1-3 in the official Python tutorial – please run through the actual tutorial line by line. It should not take you more than 30 mins to do the whole thing.
  4. Extra: Briefly read through item 4 in the tutorial on flow control (if and for statements, especially).

- Canvas for all course admin, including announcements/communications and submission of labs.
- NANOx81 for all materials, including labs with instructions.

- Questions welcomed at any time during or after lectures
- NANO181R/NANO281R is relatively very new - the instructors will try their best, but I would ask for you to be tolerant of any issues while we continue to improve the curriculum and labs.
- Your **feedback** is invaluable for shaping the current course as well as future courses.
- Email all feedback directly to ongsp@ucsd.edu.

The End