NANOx81- Data Science in Materials Science
Course Admin

Shyue Ping Ong

University of California, San Diego

Fall 2023

Course Objectives

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 (Introduction to Python for Data Science and Data Wrangling)
- Weeks 4 and 5: Linear Methods and Unsupervised Learning
- Week 6: Lab 2 (Exercises in linear methods and clustering for materials science)
- 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: Keivan Rahmani (kerahmani@ucsd.edu)

Recommended Textbooks (All Free)

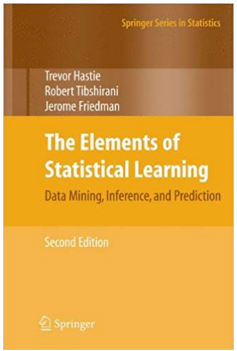

Figure: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition [Amazon][Free PDF]

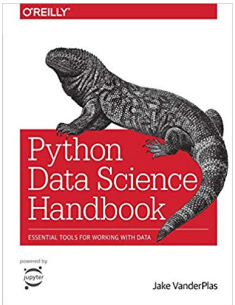

Figure: Python Data Science Handbook [Amazon] [Free web version]

Course Structure

- Lectures/Labs (Tues/Thurs @ 1100-1220) in person.
- 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 NANO181 and NANO281
  - Lab 2: 25% - Same for NANO181 and NANO281
  - Lab 3: 50% - Different for NANO181 and NANO281

Lab Assessment Criteria

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

- **Collaboration policy**: Working together is highly encouraged, but each student must submit his / her own work.
- To make the best use of this course, you should make sure you can do the exercises and not over-rely on your course mates.

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 distributions, 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).

Course Admin

- Canvas for all course admin, including announcements/communications, lecture slides, submission of labs, useful resources (e.g., ebooks, websites, etc.)
- NANOx81 Github Page - all labs with instructions.

Questions and Feedback

- Questions welcomed at any time during or after lectures
- NANOx81 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