NANO281 - Data Science in Materials Science
Course Admin

Shyue Ping Ong

University of California, San Diego

NANO281

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.

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

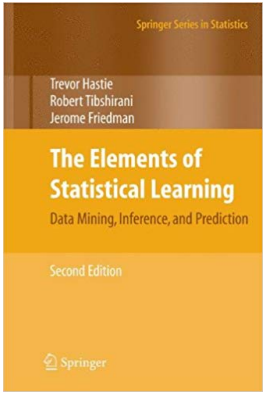
- Weeks 1 and 2: Introduction to Data Science, Python and Data Wrangling
- Week 3: Lab 1
- Weeks 4 and 5: Linear Methods and Unsupervised Learning
- Week 6: Lab 2
- Weeks 7 and 8: Kernel Methods, Trees and other Advanced Machine Learning
- Weeks 9 and 10: Final Lab

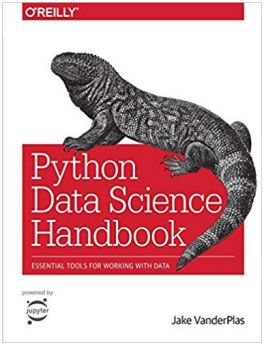
Instructors

- Lecturer: Shyue Ping Ong (ongsp@eng.ucsd.edu)
- "Teaching Assistant": Chi Chen (chc273@eng.ucsd.edu)

Recommended Textbooks (All Free)

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition
  [Amazon] [Free PDF]
- Python Data Science Handbook
  [Amazon] [Free web version]





Course Structure

- Lectures/Labs (T/T @ 930-1050)
- Conducted in-person
- Please bring your laptops.
- Grading:
  - Lab 1: 25%
  - Lab 2: 25%
  - Lab 3: 50%
- **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.

Lab Assessment Criteria

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

Class etiquette

- Highly interactive
- 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.

Pre-requisites

- Knowledge of basic statistics (e.g., Gaussian distributions, Bayes theorem, etc.)
- Knowledge of basic linear algebra (e.g., matrix multiplications, eigenvalue decompositions, inverses)
- (Optional) Some programming experience, ideally in the Python programming language, would be helpful.
- Homework:
  1. Download and install python for your OS.
  2. Go through items 1-3 in the official Python tutorial – please fire up your python and run through the actual tutorial line by line. It should not take you more than 30 mins to do the whole thing.
  3. Extra: Briefly read through item 4 in the tutorial on flow control (if and for statements, especially).

Course Admin

- Google Classroom Code
- Used for all course admin, including announcements/communications.
- Lecture slides
- Submission of labs
- Useful resources (e.g., ebooks, websites, etc.)

Questions and Feedback

- Questions welcomed at any time during or after lectures
- NANO281 is 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 in shaping the current course as well as future courses.
- Email all feedback directly to ongsp@ucsd.edu.

The End