COGS 138 | Neural Data Science

Meeting: Winter 2021, TuTh 2:00p-3:20p
Instructor: Professor Bradley Voytek (bvoytek@ucsd.edu)
Teaching Assistants (TAs): Kwonjoon Lee (kwl042@ucsd.edu)

Course Piazza*: http://piazza.com/ucsd/winter2021/cogs138
Course podcast/screencast: https://podcast.ucsd.edu/
*You will be able to post anonymously on Piazza; however, you will only be anonymous to your classmates. Your Instructor and TAs will be able to see who you are.

Overview

Neuroscience is a rapidly changing field that is increasingly moving towards ever larger and more diverse datasets that are analyzed using increasingly sophisticated computational statistical methods. There is a strong need to neuroscientists who can think deeply about problems that incorporate information from a wide array of domains including psychology and behavior, cognitive science, genomics, pharmacology and chemistry, biophysics, statistics, and AI/ML. With its focus on combining many large, multidimensional, heterogeneous datasets to answer questions and solve problems, data science provides a framework for achieving this goal.

Determining what data one needs, and how to effectively combine datasets, is a creative process. For example, a neural data scientist might be tasked with combining:

> 1) *demographic information* and 2) *multiple cognitive and behavioral measures*, from people from whom we might collect;
> 3) *biometric data*, 4) *motion capture data* to understand motor control, and 5) *eye-tracking* to study attention, along with;
> 6) *structural connectomic* and 7) *functional brain imaging* data collected using methods with different spatial and temporal resolution (such as fMRI and EEG), and then place those results into context relative to;
> 8) *average human brain gene expression patterns* and 9) *the existing knowledge embedded within the peer-reviewed neuroscience literature* (>3,000,000 papers).

These types of data are very different: continuous and ordinal, time-series, video and images, directed graphs, spatial, high-dimensional categorical / nominal, and unstructured natural language. What is the appropriate way to aggregate and synthesize these data? What are the benefits and caveats for, say, aggregating spatially versus temporally? Being able to conceptualize how to carry out this integration is necessary before leveraging any technical skills will even be useful.

Learning objectives

In this class you will learn:

- How to think from a "data first" perspective: what data *would* you need to answer your scientific questions of interest?
- How to develop hypotheses specific to big data environments in neuroscience.
- How to work with many different neuroscience data types that might include data on brain structure and connectivity, single-unit spiking, field potential, gene expression, and even text-mining of the peer-reviewed neuroscientific literature.
- How to read and analyze data stored in standard formats (e.g., Neurodata Without Borders and Brain Imaging Data Structure).
- How to integrate multiple heterogeneous datasets in scientifically meaningful ways.
- Choose statistical model(s) informed by the underlying data.
- Design a big data experiment and excavate data from multiple open data sources.
- Consider alternative hypotheses and assess for spurious correlations and results.

This is a project-based course wherein you will use computational notebooks to perform exploratory data analyses and to test hypotheses within and across large neuroscience datasets.

Course logistics **(subject to change)**

Neural Data Science will be a lecture and project-based course. It will consist of three hours of lecture (two 1.5 hour sessions per week); one that is designated for lecture, and one designated for in-class project work and discussion. Homework will consist of problem sets designed to support students with their progression through the larger class projects. Students will be evaluated on their in-class work and participation (20%), problem sets (20%), and three large projects (20% each). The final exam will be in the form of a final project. **This is all subject to change before the quarter begins!**

Grades

Grades for assignments and exams will be released approximately one week after the submission date. *It is your responsibility to ensure your assignments are submitted on time, without glitches or errors, and to check your grades and get in touch if any are missing or if you think there is a problem*.

Regrades. We work hard to grade fairly and to return assignments quickly. We know that you also work hard and want to receive the grade you believe you've earned. Occasionally, grading mistakes do happen, and it's important for us to correct them. If you think there's a mistake in your grade for an assignment, submit a regrade request within 72 hours of when the grade published. This request should include evidence of why you think your answer was correct (i.e., a specific reference to something said in lecture) and should point to the specific part of the assignment for us to reconsider.

Course Resources

There is no official textbook for this course. Instead we'll rely on free online resources:

- **Whirlwind Tour of Python**: https://github.com/jakevdp/WhirlwindTourOfPython/
- **Python Data Science Handbook**: https://jakevdp.github.io/PythonDataScienceHandbook/
    - This book is available free online or in print.

Schedule **(subject to change!)**

Week 1 | Introduction to Data Science

To set the foundation for this course, we'll introduce tools that are commonly used to analyze large neuroscientific data sets including Python and Jupyter. The first two weeks will serve as a crash course to bring students up to speed on the basics of using these tools to analyze data.

- What is data science, and how does it intersect with neuroscience?
- Asking the right types of questions of your data
- Exploratory data analysis in Jupyter

Week 2 | Tools for Data Science

- NWB and BIDS Data I/O
- GitHub and version control

Week 3 | Data cleaning and visualization

- Choosing your plot types
- Looking at computed electrophysiology cell metrics in the Allen Brain Database

Week 4 | Time series analysis and signal processing

How do we perform analyses on open source datasets for electrophysiology (single cell, EEG, and MEG) and imaging (two-photon) data? Students will perform time series analyses on electrophysiology and imaging data, and learn about the various types of signal processing used for different data types.

- Field potential and/or calcium data

Week 5 | Spike sorting, PCA

- Introduction to Neuropixels dataset

Week 6 | Parameterizing heterogeneous datasets

Modern neuroscience incorporates various types of data, both physiological and behavioral. This portion of this course will address how we integrate diverse types of data (e.g., analog signals, video, trial pulses, text data, etc.) to address an experimental question.

- Handling of multiple data types in a Neuropixels dataset (running data, pupil data, electrophysiology data)

Week 7 | Statistical data analysis

- What are correlations?
- Using 2-photon imaging data to calculate correlations between cells and generate heatmaps

Week 8 | Modeling single trial data

With many simultaneously recorded neurons, we can avoid signal averaging and actually observe patterns in signal trial data. Here, we'll explore different approaches to investigating single trials of population data.

Weeks 9 & 10 | Final projects

In the last two weeks of the course, students will work on final projects in which they use any of the open source datasets that we have encountered to address pressing questions in neuroscience, such as the network properties of the connectome, the role of different cell types in the cortex, or the representation of language?

Other good stuff

Class conduct. In all interactions in this class, you are expected to be respectful. This includes following the UC San Diego principles of community .

This class will be a welcoming, inclusive, and harassment-free experience for everyone, regardless of gender, gender identity and expression, age, sexual orientation, disability, physical appearance, body size, race, ethnicity, religion (or lack thereof), political beliefs/leanings, or technology choices.

At all times, you should be considerate and respectful. Always refrain from demeaning, discriminatory, or harassing behavior and speech. Last of all, take care of each other.

If you have a concern, please speak with Dr. Voytek, your TAs, or IAs. If you are uncomfortable doing so, that's ok! The OPHD (Office for the Prevention of Sexual Harassment and Discrimination) and CARE (confidential advocacy and education office for sexual violence and gender-based violence) are wonderful resources on campus.

Academic integrity. Don't cheat.

You are encouraged to (and at times will have to) work together and help one another. However, you are personally responsible for the work you submit. For group assignments, it is also your responsibility to ensure that you understand everything your group has submitted, and to make sure that the correct file has been uploaded, that the uploaded file is uncorrupted, and that it renders correctly. Projects may include ideas and code from other sources—but these other sources must be documented with clear attribution. Please review academic integrity policies here.

We anticipate you will all do well in this course! However, if you are feeling lost or overwhelmed, that's ok. Should that occur, we recommend: (1) asking questions in class, (2) attending office hours and/or (3) asking for help on Piazza.

Cheating and plagiarism have been and will be strongly penalized. If, for whatever reason, something else prohibits you from being able to turn in an assignment on time, immediately contact me by emailing your assignment (bvoytek@ucsd.edu), or else it will be graded as late.

Disability access. Students requesting accommodations due to a disability must provide a current Authorization for Accommodation (AFA) letter. These letters are issued by the Office for Students with Disabilities (OSD), which is located in *University Center 202* behind Center Hall. Please make arrangements to contact Dr. Voytek privately to arrange accommodations.

Contacting the OSD can help you further:
858.534.4382 (phone)
osd@ucsd.edu (email)
http://disabilities.ucsd.edu

How to Get Your Question(s) Answered and/or Provide Feedback. It's *great* that we have so many ways to communicate, but it can get tricky to figure out who to contact or where your question belongs or when to expect a response. These guidelines are to help you get your question answered as quickly as possible *and* to ensure that we're able to get to everyone's questions.

That said, to ensure that we're respecting their time, TAs and IAs have been instructed they're only obligated to answer questions between normal working hours (M-F 9am-5pm). However, I *know* that's not when you may be doing your work. So, please feel free to post whenever is best for you while knowing that if you post late at night or on a weekend, you may not get a response until the next weekday. As such, do your best not to wait until the last minute to ask a question.

Final thoughts. If you have…

- **Questions about course content**: These are awesome! We want everyone to see them and have their questions answered too… so post these to Piazza!
- **Questions about course logistics**: First, check the syllabus. If you can't find the answer, ask a classmate. If still unsure, post on Piazza.
- **Questions about a grade**: If for an assignment, submit a regrade request. For anything else, post as a question on Piazza, address it to "Instructors," and select the folder "regrades".
- **Something super cool to share related to class**: Feel free to email Dr. Voytek (bvoytek@ucsd.edu) or come to office hours. Be sure to include COGS138 in the email subject line and your full name in your message.
- **Something you want to talk about in-depth**: Come to office hours or schedule a time to meet by email (bvoytek@ucsd.edu). Be sure to include COGS138 in the email subject line.