BIPN 162 / BGGN 240 | Neural Data Science

Spring 2022

Instructor
Mikio Aoi, PhD
maoi@ucsd.edu

Instructional Assistant
Dexter Tsin
ttsin@ucsd.edu

Office hours
Mikio - Wed 11-11:50 am in SOLIS 111
Dexter - M/W 5-6pm

Class Schedule
Lectures: M/W/F, 10-10:50 am,
Location: SOLIS 109

Course Description: Project-based course in which students will use computational notebooks to perform exploratory data analyses and to test hypotheses in large neuroscience datasets, including the differences between unique neuron types, leveraging text mining of the neuroscience literature, and human neuroimaging analyses.

Students will be able to:

- Understand the fundamental concepts underlying modern data analysis techniques commonly used in neuroscience.
- Recognize modeling and inference techniques used in recent literature
- Know when different modeling techniques or statistical methods should be used in a given situation
- Execute and present their own data analysis project

Grading

- **Homework Assignments (40%, 4 assignments)**: Take-home coding assignments will support your progression through the course topics. Assignments will be submitted through the DataHub (http://datahub.ucsd.edu) and graded automatically using a tool called NBGrader.
  - All assignments are due **1-week after they are assigned** and are worth **10% each.**
  - These assignments should be completed individually and should take you about 2 hours.

- **Projects (groups of 2-3, 60%)** Choose a paper that uses a new data analysis technique that has been published no more than 10 years ago (most should have code and data publicly available). Implement this method using a new data set (either your own or from publicly available source).

  - Project proposal (20%) (2-pages) Project proposals will describe the main ideas of the paper, a description of the analysis method to be utilized, what question about the brain or behavior the method is meant to address, what specific results from the paper your team will be reproducing, what additional dataset is to be analyzed, the analysis to be conducted with this new dataset, and how the method is appropriate for this new analysis.

  - Project code and final report (40%) Your final project will be a written record of your project. It should document what is interesting (or not) about the method you selected and what features of it made it well suited to address the analysis you intended. Details for what your final report should include will be included in the final project handout.

Additional notes about grading:

- We will be using Canvas (http://canvas.ucsd.edu) to manage grades and assignments.
- **Late policy**: Assignments and projects will lose -10% for each day they are late.
- **Grading Scheme:** Final scores will be converted to letter grades, where A=100-90%, B=89-80%, C=79-70%, D=69-60%, and F=59-0%. For positive and minus grades, A+ = 97-100, A = 93-96.99, A- = 90-92.99, B+ = 87-89.99, B = 83-86.99, and so on.
- Grading of the final project proposal will be scaled based on whether you are enrolled in the undergraduate (BIPN 162) or graduate version of the course (BGGN 240). Graduate students will not be tasked with additional work but the expectations for quality of work will be higher. More information will be provided in the final project handout.

Course Resources

There is no official textbook for this course. Instead, there will be suggested reading listed throughout the course. In addition, you are encouraged to consult the following suggested general resources:

Coding in python for data science:

- VanderPlas, **Whirlwind Tour of Python**
- VanderPlas, **Python Data Science Handbook** (available free online or in print)
- Addison& Eldridge, **Principles of Data Science** (course notes for DSC 10)
- Adhikari & DeNero, **The Foundations of Data Science** (serves as textbook to UC Berkley's Data8 Course)
- Software Carpentry, **Plotting and Programming in Python**

You may also sign up for DataQuest. They have many free tutorials in their Data Scientist Path that are relevant to topics in this course.

Neuro-inspired statistics and machine learning textbooks:

- Kass, Eden, & Brown. Analysis of neural data. 2014.
- Eden & Kramer. Case studies in neural data analysis, 2016
- Nylen & Wallisch. Neural Data Science: A Primer with MATLAB® and PythonTM. 2017.

General introductory stats/ML textbooks:

- Bishop & Nasrabadi. Pattern recognition and machine learning. 2006.
- James, Witten, Hastie, & Tibshirani. An Introduction to Statistical Learning. 2013.
- Craigmile. All of statistics: A concise course in statistical inference. 2005

Additional resources:

- Neuromatch Academy – Computational neuroscience online tutorials

Environment of inclusivity

This course is meant to be intellectually challenging and is attended by students with a wide range of skills, levels of preparation, life histories, cultures, and ableness. It is essential to the functioning of this course that we foster an inclusive environment of mutual respect, in which we all feel we can express confusion, ask questions, and challenge each other constructively. If at any point you feel that you or others are not being provided with the resources you need to be successful in this course then please let me know immediately.

Masking Policy

Per University policy as of March 21, 2022, masks continue to be required in all indoor classroom settings. Please be considerate of your classmates and wear your mask at all times when in the classroom.

Course accommodations

If you need accommodations for this course due to a disability, please contact the *Office for Students with Disabilities* (osd@ucsd.edu) for an Authorization for Accommodation (AFA) letter. Please speak with me in the first week of class if you intend to apply for accommodations. For more information, visit http://disabilities.ucsd.edu.

Academic Integrity

You won't benefit if others do your work. If you're unclear about what constitutes cheating in this course, please ask. Cases of academic dishonesty or cheating will be first handled by me, and then by the Academic Integrity Office. If you become aware of cheating in this class, you can anonymously report it: https://academicintegrity.ucsd.edu/.

Syllabus

(subject to substantial change!)

Date	Topic	Reading
Week 1	**Introduction to Neural Data Science** Areas of active research, the neural coding problem, linear regression as a fundamental starting point.	Paninski & Cunningham, *Current Opinion in Neurobiology*, "Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience" Humphries, *The Spike*, "A Neural Data Science: How & Why"
Week 2	**The Basics** Notation and linear algebra, parameter estimation = optimization, the design matrix, convolutional regressors	Neuromatch academy linear algebra tutorial
Week 3	**Regularization, Bayes' theorem, and Beyond Gauss** Effects of regularization, Bayesian estimation, generalized linear models (GLMs)	
Week 4	**Decoding: Uses, evaluation, and methods** Linear classification, multinomial and ordinal regression, evaluating decoders	
Week 5	**Clustering and dimensionality reduction** Mixtures of Gaussians, k-nearest neighbors, PCA	
Week 6	**Making use of structure in models and experiments: Hierarchical models** Building special structure into models, mixed effects and leveraging multi-subject data	
Week 7	**Sequential data: A special kind of structure** Intro to time series, the Fourier transform, hidden Markov models	
Week 8	**Back to basics: All about matrices**	

	Recasting what we've learned in terms of matrix models. Source separation, factor analysis, and independent components analysis.	
Week 9	**The myth and mystery of neural networks** Basics of neural networks, backprop, their uses in neuroscience, and what makes them wierd.	
Week 10	**Advanced topics** Nonlinear dimensionality reduction, modeling neural dynamics, the future of neural data analysis	