Welcome!

Statistical Modeling and Analysis of Neural Data
(NEU 560)
Fall 2020

Instructor: Jonathan Pillow
Syllabus & Course Expectations
Instructors

Jonathan Pillow (instructor)
Professor, PNI & Psychology
Center for Statistics & Machine Learning (CSML)

Orren Karniol-Tambour (AI)
3rd year PhD student, PNI
(Pillow lab)

Office hours: Tuesday, 9a-10a, and by appt.
Description

This course aims to introduce students to advanced statistical and machine learning methods for analyzing of neural data, with an emphasis on methods derived from regression (supervised) and latent factor (unsupervised) models. Each technique will be illustrated via applications to neural datasets. The course will have a heavy emphasis on programming, and a substantial portion of the grade will come from homework assignments that involve writing python code to implement various statistical methods and apply them to data. The topics will focus on methods for the analysis of single and multi-neuron spike train data, calcium imaging, and fMRI datasets. Sample topics to be covered:

• neural encoding models
• logistic regression
• generalized linear models
• Poisson processes
• regularization (ridge, lasso)
• empirical Bayes / evidence optimization
• information theory
• Gaussian processes
• mixture models and the EM algorithm
• PCA, factor analysis and its extensions
• Hidden Markov Models
• Kalman filter & linear dynamical systems (LDS)
• approximate inference methods (Laplace approximation, variational inference)
• variational auto-encoders (VAEs)
Pre-requisites

- Calculus, Linear Algebra, and familiarity with basic ideas in probability / statistics (random variables, common probability distributions, mean, variance).
- No previous experience with neural data is required
- Programming experience in python is helpful given that homework assignments require programming.
Format

This is a lecture course (2 lectures per week), but students will be expected to keep up with course readings, homework, and to participate actively in class. I believe that the best way to learn mathematical concepts is by actively putting them to use.

The majority of the grade will therefore be based on homework assignments, which will involve both programming (implementation of algorithms) and paper-and-pencil problem solving.

Students will also complete a final project, alone or in collaboration with another student, and will make a 15-minute presentation on this project during the final portion of the course.
Pre-requisites

• Calculus, Linear Algebra, and familiarity with basic ideas in probability / statistics (random variables, common probability distributions, mean, variance).

• No previous experience with neural data is required

• Programming experience in python is helpful given that homework assignments require programming.
**Requirements:**
Homework will be assigned approximately every 2 weeks. Assignments should be sent to the AI by the stated due date. Late assignments will lose 10% credit per day late. Students are encouraged to work together on homework problems, but are expected to write computer code in groups of not more than two. In-class quizzes will take place once per week. There is no final exam.

**Grading:**
Grades for will be based on
- Homework: 60%
- Quizzes: 10%
- Participation: 10%
- Final project (presentation): 10%
- Final project (writeup): 10%
**Homework:**

Homework problem sets will involve a mix of programming assignments and paper-and-pencil math problems (with substantially more of the former). The goal of these assignments is to force students to put the mathematical concepts from class into practice. I believe that writing a computer program to implement or test a mathematical concept provides a much deeper understanding than merely writing down an analytical derivation. Many of the functions you write will also serve as prototypes for data analysis problems you will face in real neuroscience research.

Homeworks will be submitted in the form a Jupyter notebook. (Students will use LaTeX markup within the notebook to answer to analytical / paper-and-pencil math problems). Each homework assignment will count equally, so the homework grade will be the average over all assignments.
**Quizzes:**

Every Tuesday, class will begin with a 5-minute quiz about previous material. These quizzes should be easy for students who have followed recent lectures (or successfully completed previous homework assignments). The goal of these quizzes is to make sure students retain and have an active grasp of key concepts, without assistance from AIs or fellow students. Each student can drop their three lowest quiz scores.

Makeups will not be given; students who need to miss a class can simply count any missed quizzes among their lowest three.
Course Projects:

All students will complete a course project on some topic related to statistical modeling and analysis of neural data, which will count for 20% of the final grade. Students may work alone or in groups of two. I will provide a list of suggested topics / project ideas, but students are also free to come up with their own proposals based on course readings, problem sets, or their own interests. Example projects might involve the implementation and extension of a method discussed in class, or the application one of these methods to neural data. I will meet at least twice with all groups to approve project proposals, and to monitor progress and offer guidance / suggestions. A special session devoted to student presentations of their projects will take place during reading period (date TBA).
Participation:

Participation in real-time virtual class sessions, and online on Piazza is strongly encouraged. Your questions (and answers) will help both you and your fellow students. Students are expected to attend each scheduled class and to participate fully in-class exercises and discussions.

**Zoom:** Please have your video (with whatever background you prefer to use) turned on, unless you have an unexpected difficulty or have arranged with me otherwise.
Piazza:

We encourage all students to post questions to Piazza instead of sending email. This will allow others to benefit from your question, and will often result in a faster and more complete answer (since your fellow students may post answers before any of the instructors can). Please participate on Piazza, and endorse questions and answers as you see fit. Piazza activity will count toward the 10% participation grade!

Piazza signup page:  piazza.com/princeton/fall2020/neu560
Collaboration and Academic Integrity:

You are welcome to work together on problem sets (I would even encourage it), but the work you submit should be uniquely your own, prepared by your own hand. Students should understand every step in their code such that they could implement it again without hep from anyone.
Python:

All homework assignments will require programming solutions in Python. To install Python, we recommend installing the Anaconda development environment, which contains Python and a collection of important / popular packages.

We will be using Python 3 (not Python 2), so when you install Anaconda, choose that one (currently Python 3.8). Installing Anaconda will also Jupyter (interactive notebooks) and Spyder (a development environment).

(stay tuned for more)
Text:

There is no official textbook, but several reference texts will be placed on hold at the library (and pdf versions of the first two can be found online):

- *Analysis of Neural Data*. Kass, Eden & Brown
Questions?
Breakout: introduce yourselves!
Course website:

http://pillowlab.princeton.edu/teaching/statneuro2020/

Statistical Modeling and Analysis of Neural Data
NEU 560 (Fall 2020), Princeton University

time: Tu/Th 10:00-11:20

location: cyberspace

instructor: Jonathan Pillow

AI: Orren Karniol-Tambour (Office hours: Tuesday 9am-10am)

prerequisites: A good working knowledge of calculus, linear algebra, and basic probability / statistics. Familiarity with python is also desirable, as homework assignments and final projects will involve programming. No prior experience with neural data is required.

brief description: This course aims to introduce students to methods for modeling and analyzing neural datasets, with an emphasis on statistical approaches to the problem of information processing in neural populations. A tentative list of topics includes: neural encoding models, Poisson processes, generalized linear models, logistic regression, Gaussian processes, latent variable models, factor analysis, mixture models, EM, Kalman filter, latent dynamical models of neural activity. The course is aimed at students from quantitative backgrounds (neuroscience, engineering, math, physics, computer science, statistics, psychology) who are interested in the modeling and analysis of neural data.

syllabus: pdf

piazza site: link

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading</th>
<th>Slides / Notes</th>
<th>HW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu 9.01</td>
<td>Course Intro</td>
<td>spikes_intro.pdf</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Th 9.03| Linear algebra review, SVD          | linalg handouts:
E. Simoncelli handout (pdf)
M. Jordan chapter (pdf) |                              |    |
| Tu 9.08| SVD applications, least squares regression |                                |                              |    |
| Th 9.10| Low rank matrices, determinant      |                                  |                              |    |
| Tu 9.15| PCA                                 |                                  |                              |    |
suggested reading for today:

Introduction to *Spikes*

Bill Bialek
course poll

Please fill out before Thursday!

https://forms.gle/5s9fBGzXBrB9SWfm6
Brief introduction to computational & statistical neuroscience

Jonathan Pillow

Lecture #1
Statistical Modeling and Analysis of Neural Data
Fall 2020
What is computational neuroscience?

1. Computational/statistical tools to study the brain.
   - Extract structure from noisy data
   - Build models that capture behavior of neurons

2. Study how the brain behaves *as a computer*
   - Brain is a machine for processing information & computing relevant outputs
   - Machine for *statistical inference*
Mind-Brain Problem

What is the relationship of the mind to the brain?
The brain as a computer:

“The brain computes! This is accepted as a truism by the majority of neuroscientists engaged in discovering the principles employed in the design and operation of nervous systems. What is meant here is that any brain takes the incoming sensory data, encodes them into various biophysical variables, such as the membrane potential or neuronal firing rates, and subsequently performs a very large number of ill-specified operations, frequently termed computations, on these variables to extract relevant features from the input. The outcome of some of these computations can be stored for later access and will, ultimately, control the motor output of the animal in appropriate ways.”

- Christof Koch, *Biophysics of Computation*
Short history of brain metaphors:

• hydraulic device (Descartes, 17th C.)
• mill (Leibniz, 17th C.)
• telegraph (Sherrington, early 20th C.)
• telephone switchboard (20th C.)
• digital computer (late 20th C.)
• quantum computer? (Penrose, 1989)
• convolutional neural network? (21st C.)
What does it mean to claim the brain is a computer?

- The physical parts of the brain are important only insofar as they represent steps in a formal calculation.
- Any physical device implementing the same formal system would have the same “mind properties” as a brain.
What does it mean to claim the brain is a computer?

My claim: Most neuroscientists take it for granted that the brain is a computer.

They are devoted to finding out which computer (i.e., what formal structure? what algorithms does the brain implement?).
What is (some of) the evidence that the brain is a computer?
Mathematical model of sensory neurons

The retina

- Photoreceptors
- Bipolar cells
- Retinal ganglion cells

Detect light

Output cells (send all visual information to the brain)

To brain!
Mathematical model of sensory neurons

the retina

photoreceptors

bipolar cells

retinal ganglion cell

Difference of light in “center” and light in the “surround”

what mathematical operation?
Mathematical model of sensory neurons

- Photoreceptors
- Bipolar cells
- Retinal ganglion cell

Difference of light in “center” and light in the “surround”

Stimulus

Lots of spikes!
Mathematical model of sensory neurons

Difference of light in “center” and light in the “surround”

stimulus

few spikes
Mathematical model of sensory neurons

Difference of light in “center” and light in the “surround”

more spikes
Mach Bands

Each stripe has constant luminance

Then why does it look like there’s a gradient?
Mach Bands

Each stripe has constant luminance.

Then why does it look like there's a gradient?
Lightness Illusion
Hermann illusion
This magical slide can track where you’re looking
The Neural Coding Problem

Questions: • How are stimuli and actions encoded in neural activity?
• How are representations transformed between brain areas?
The Neural Coding Problem

Approach:  
• develop flexible statistical models of $P(y|x)$  
• quantify information coding strategies and mechanisms
Color Computations

Beau Lotto
Color Computations

Beau Lotto
color after-images

• neurons adjust their response properties after prolonged exposure to an image
• we can compute (and predict) these changes!
  • red —> green after-image
  • blue —> yellow after-image
  • black —> white after-image
Bayesian Models for Perception

**Bayes’ rule:** \( P(B \mid A) \propto P(A \mid B) P(B) \)

Formula for computing:

- \( P(\text{what’s in the world} \mid \text{sensory data}) \)

(This is what our brain wants to know!)

\[ P(\text{world} \mid \text{sense data}) \propto P(\text{sense data} \mid \text{world}) P(\text{world}) \]

- **Posterior** (resulting beliefs about the world)
- **Likelihood** (given by laws of physics; ambiguous because many world states could give rise to same sense data)
- **Prior** (given by past experience)
Helmholtz: perception as “optimal inference”

“Perception is our best guess as to what is in the world, given our current sensory evidence and our prior experience.”

\[
P(\text{world} | \text{sense data}) \propto P(\text{sense data} | \text{world}) \cdot P(\text{world})
\]

- **Posterior**
  - (resulting beliefs about the world)

- **Likelihood**
  - (given by laws of physics; ambiguous because many world states could give rise to same sense data)

- **Prior**
  - (given by past experience)
Helmholtz: perception as “optimal inference”

“Perception is our best guess as to what is in the world, given our current sensory evidence and our prior experience.”

\[
P(\text{world} | \text{sense data}) \propto P(\text{sense data} | \text{world}) \cdot P(\text{world})
\]

- **Posterior**: (resulting beliefs about the world)
- **Likelihood**: (given by laws of physics; ambiguous because many world states could give rise to same sense data)
- **Prior**: (given by past experience)
Many different 3D scenes can give rise to the same 2D retinal image.

The Ames Room
Many different 3D scenes can give rise to the same 2D retinal image.

The Ames Room

How does our brain go about deciding which interpretation?

\[ P(\text{image} \mid A) \text{ and } P(\text{image} \mid B) \text{ are equal!} \quad (\text{both } A \text{ and } B \text{ could have generated this image}) \]

Let's use Bayes' rule:

\[
P(A \mid \text{image}) = \frac{P(\text{image} \mid A) \cdot P(A)}{Z} \\
P(B \mid \text{image}) = \frac{P(\text{image} \mid B) \cdot P(B)}{Z}
\]
Hollow Face Illusion

http://www.richardgregory.org/experiments/
\[ P(\text{convex}|\text{video}) = P(\text{video}|\text{convex})P(\text{convex face}) \]

\[ P(\text{concave}|\text{video}) = P(\text{video}|\text{concave})P(\text{concave face}) \]

- very strong prior for convex (“outward”) faces
Neural prostheses:

Neurons can be replaced by other entities (silicon chips) that have different physical structure but carry out the same (or similar) mathematical operations, allowing the organism to produce (“compute”) the same behavior.
Cochlear implants
(using a “different computer” to encode auditory signals)
Direct neural control of movement
Direct neural control of movement
If we understand the mathematical operations carried out by different parts of the brain, we could (in theory) replace them with new parts that perform the same computations!
Our goal: figure out how the brain works.
There are about 10 billion cubes of this size in your brain!
Tungsten Electrode
“Utah” array (96 channels)
neuropixel probe (~1K electrodes)
This is a great time to study computational / statistical neuroscience

• We are now getting incredible data.
• Computers are getting extremely fast.
• Advances in statistical/mathematical techniques are allowing us to gain a deep understanding of neural data and neural information processing capabilities
For Next Time

• Please fill out class poll: https://forms.gle/5s9fBGzXB9Swf6

• Review Linear Algebra basics

• Install Python (instructions will be posted online)