

# Caltech

datasai\_2023

*VOL. 2*

## overview & computational basics

aditya nair

CHEN TIANQIAO  
& CHRISSY  
\*  
INSTITUTE

# course organizers



**Helen O'Connor**

*Programs Coordinator*



**Ralph Adolphs**

*Bren Professor of Psychology, Neuroscience, and Biology*



**David J. Anderson**

*Seymour Benzer Professor of Biology; Tianqiao and Chrissy Chen Institute for Neuroscience Leadership Chair; Investigator, Howard Hughes Medical Institute; Director, Tianqiao and Chrissy Chen Institute for Neuroscience; Interim Director, T&C Chen Center for Systems Neuroscience*



**Mary Sikora**

*Executive Director, Tianqiao and Chrissy Chen Institute for Neuroscience*



**Adi Nair**

*Graduate Student*



**Daniel Wagenaar**

*Research Professor of Biology and Biological Engineering*

## external faculty



**Maryam Shanechi**

*Professor and Viterbi Early Career Chair in Electrical and Computer Engineering, Computer Science, and Biomedical Engineering, University of Southern California*



**Chethan Pandarinath**

*Assistant Professor in the Wallace H. Coulter Department of Biomedical Engineering at Georgia Tech and Emory University*



**Liam Paninski**

*Professor of Statistics and Neuroscience, Columbia University*

# external speakers & TAs



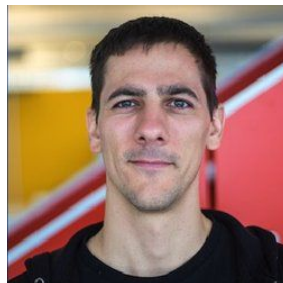
**Gabriel Rojas-Bowe**

Grad Student, Churchland Lab, UCLA



**Maxwell Mellin**

Grad Student, Churchland Lab, UCLA



**Matthew Whiteaway**

Postdoc, Paninski Lab, Columbia



**Iman Wahle**

Grad Student, Princeton



**Omid Sani**

Postdoc, Shanechi Lab, USC



**Domenick Mifsud**

Research Specialist, Emory



**Frank Lanfranchi**

Grad Student, Tsao Lab, UC Berkeley

## internal speakers & faculty



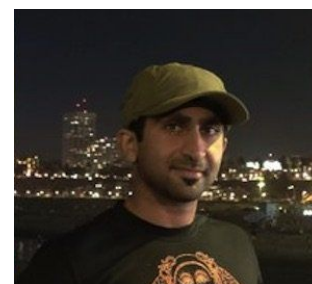
**Prof Frederick Eberhardt**  
Professor of Philosophy



**James Gornot**  
Grad Student, Thompson Lab



**Tara Chari**  
Grad Student, Pachter Lab



**Tarun Sharma**  
Grad Student, Dickinson Lab

# course schedule

## foundations & essential tools (10th - 14th)

- July 10th:  
*ethics, replication and generalizability of neural data analysis*  
Ralph Adolphs, Caltech
- July 11th:  
*machine learning basics and dynamical systems.*  
Adi Nair, James Gornet, Caltech
- July 12th:  
*encoding models in neuroscience*  
Maxwell Melin, Gabriel Rojas-Bowe, Churchland lab, UCLA
- July 13th:  
*adapting Your recordings for big data techniques*  
Daniel Wagenaar, Caltech
- July 14th:  
*dimensionality reduction & hypothesis testing*  
Tara Chari, Caltech

◆ computational &  
machine learning basics

◆ dynamical systems

◆ encoding & decoding

◆ pre-processing

◆ dimensionality reduction  
& statistics

# course schedule

## the frontier & deep learning (18th - 21st)

- July 17th:  
*state space models in neuroscience*  
Liam Paninski & Matt Whitley, Columbia
- July 18th:  
*understanding behavior & bring your own data day!*  
Tarun Sharma, Caltech, Frank Lanfranchi, UC Berkeley
- July 19th:  
*latent variable dynamical models in neuroscience*  
Maryam Shanechi, USC
- July 20th:  
*causal modelling from observational data*  
Frederick Eberhardt, Caltech
- July 21st:  
*transformers in neuroscience*  
Chethan Pandarinath & Domenick Mifsud, Emory

- ◆ state-space models
- ◆ behavior & pose estimation
- ◆ recurrent neural networks
- ◆ causal learning
- ◆ transformers

## daily schedule

- 8:30AM - 9:30AM:  
**breakfast**
- 9:45AM - 10:45AM:  
**introductory lecture**
- 10:45AM - 11:00AM:  
**bio break**
- 11:00AM - 12:00PM:  
**philosophy and general  
questions**
- 12:15PM - 1:15PM:  
**lunch**
- 1:30PM - 2:30PM:  
**methods lecture**
- 2:30PM - 6:00PM:  
**hands-on session**

## on BYOD (18th)

- 8:30AM - 9:30AM:  
**breakfast**
- 9:45AM - 10:45AM:  
**introductory lecture**
- 10:45AM - 11:00AM:  
**bio break**
- 11:00AM - 12:00PM:  
**hands-on session**
- 12:15PM - 1:15PM:  
**lunch**
- 1:30PM - 2:30PM:  
**lecture on behavior**
- 2:30PM - 6:00PM:  
**hands-on session**



Ask a lot of questions,  
feel free to interrupt!

resources:

**course website:**

<https://neuroscience.caltech.edu/about/research-centers/chen-center-for-datasai/2023-data-science-and-ai-for-neuroscience-summer-school>



**github repo:**

[https://github.com/cheninstitutecaltech/Caltech\\_DATASAI\\_Neuroscience\\_23](https://github.com/cheninstitutecaltech/Caltech_DATASAI_Neuroscience_23)



resources:

**google colab pro+:**

<https://colab.research.google.com/signup>

**slack channel:**

[https://join.slack.com/t/cheninstitute-9181373/shared\\_invite/zt-1b54fb5gi-fuAhn51vD\\_qorBqTbXoKcQ](https://join.slack.com/t/cheninstitute-9181373/shared_invite/zt-1b54fb5gi-fuAhn51vD_qorBqTbXoKcQ)



**bring your own data day (BYOD) - 18th**

google colab pro+:

first go to: <https://www.prepaidgiftbalance.com/>

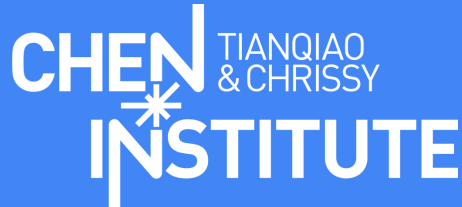
register your card with the following address:

1200 E California Blvd

MSC 2-59

Pasadena, CA - 91125

then go to: <https://colab.research.google.com/signup>



Caltech

# computational mindset

you will see many course concepts 2-3 times across multiple days, at different levels of depth.



- architectures
- mathematical functions
- code



- applying a method
- implementing a method
- practicing the math behind the method

you are meant to struggle through this

don't be afraid to 'break' things while you code

your recourse when you don't know what to do is:



- ♦ google → [stack overflow](#), ...
- ♦ ask your neighbor. debug problems together!
- ♦ Ask the TAs for help if neither you nor any of your neighbors can figure it out.
- ♦ as a last resort you can check out the solution code also in the notebook!

# resources beyond this course

datasai teaches concepts, covers a broad overview of tools and provides practical training on using tools

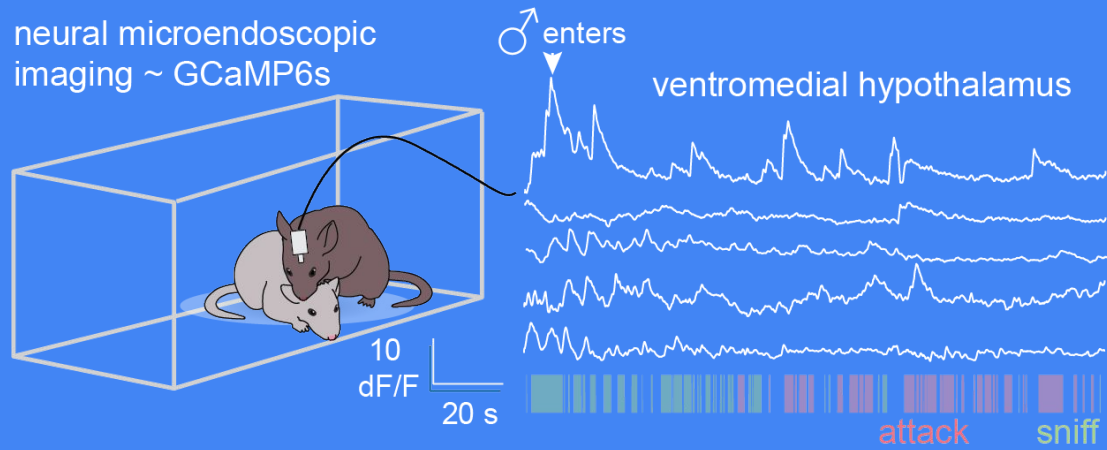
!! does not go into in-depth mathematical treatments of methods

## ML courses at Caltech:

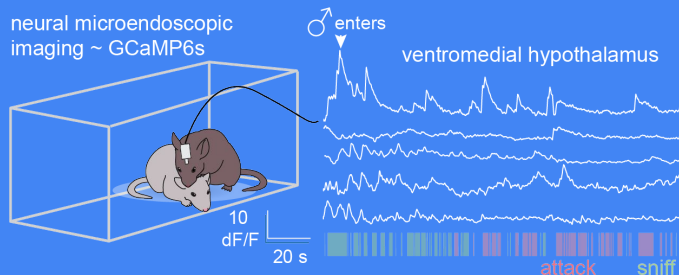
CS156: Learning from data  
Yaser Abu-Mostafa

CS155: Machine Learning & Data Mining  
Yisong Yue



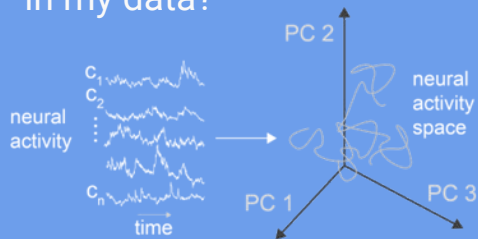


so you have some data.....



# discovering brain-behavior relationships

what are the major signals in my data?



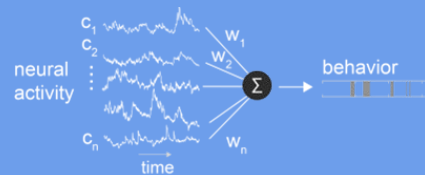
dimensionality reduction

what are single neurons tuned to?



linear encoding models

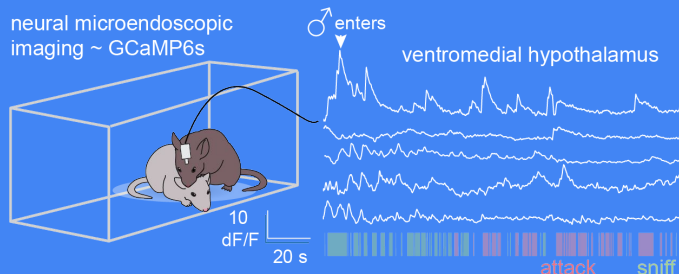
is information about behavior present in my data?



linear decoding models

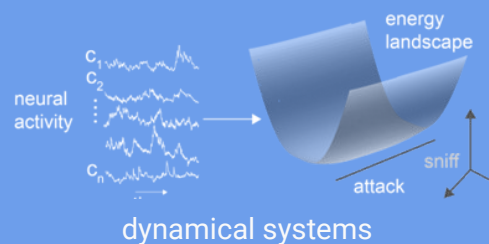
we'll cover these methods in week 1

foundations & essential tools

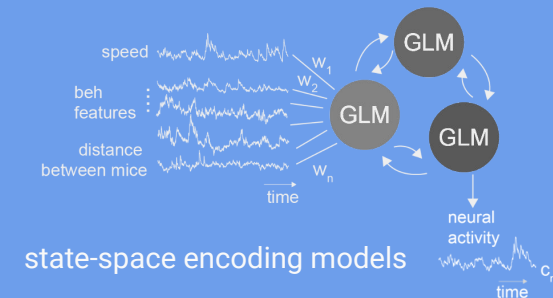


# discovering brain-behavior relationships

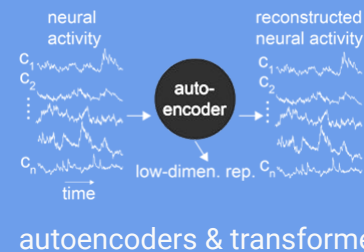
can I discover computations in data in an unsupervised manner?



what if the encoding of behavior in single neurons is dynamic?



How much non-linearly encoded behavior information is available?



we'll cover these methods in week 2

the 'frontier' & deep learning

today we're going to cover:  
computational basics

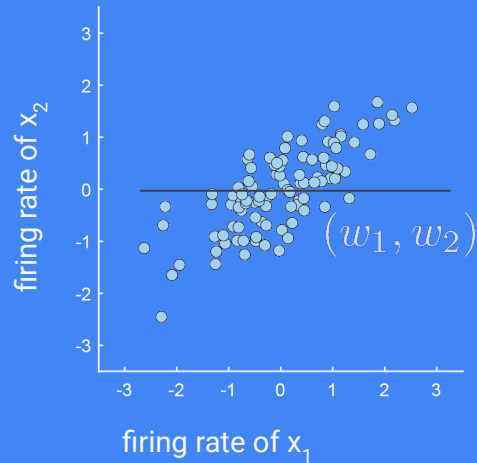
## **part 1**

dimensionality reduction via PCA  
data visualization and familiarity with python

## **part 2**

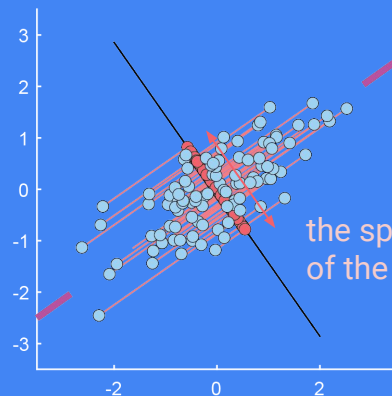
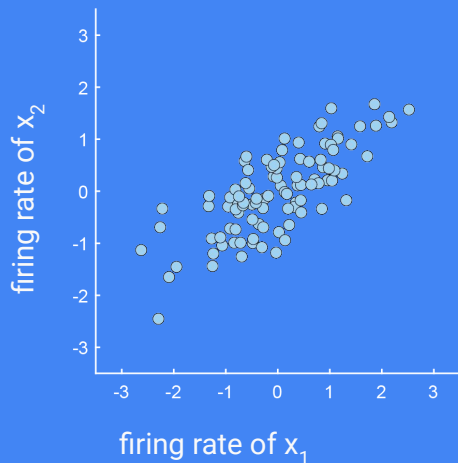
decode behavior using regression  
overfitting & regularization

let's consider a system of two neurons  
how can we best summarize this data?



make a 1-D 'projection', i.e a weighted  
combination of activities of neurons  
 $s_1(t) = w_1 x_1(t) + w_2 x_2(t)$

let's consider a system of two neurons  
how can we best summarize this data?



the sum of the pink lines tells us how well  $s$  reconstructs the data

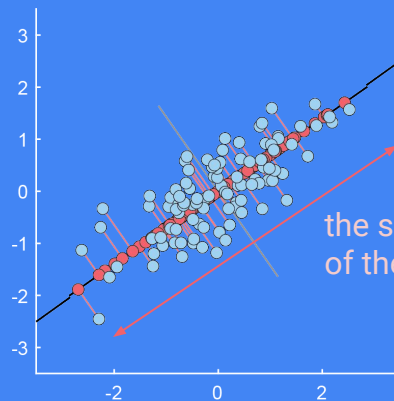
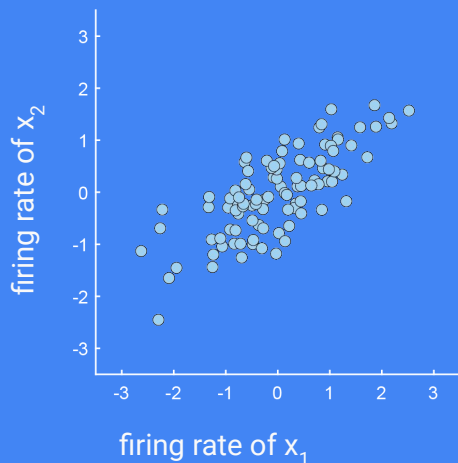
the spread of the projection of  $x$  onto  $s$  tell us of the variance captured by  $s$

an example projection

$$s_1(t) = w_1 x_1(t) + w_2 x_2(t)$$

one way to summarize the data is maximize variance or minimize reconstruction → PCA

let's consider a system of two neurons  
how can we best summarize this data?



the sum of the pink lines tells us how well  $s$  reconstructs the data

the spread of the projection of  $x$  onto  $s$  tell us of the variance captured by  $s$

an example projection

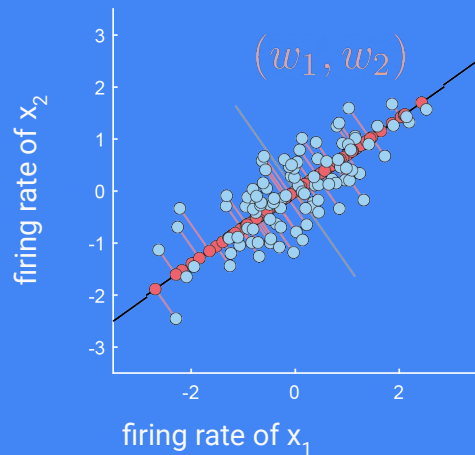
$$s_1(t) = w_1 x_1(t) + w_2 x_2(t)$$

PCA maximizes variance or minimize reconstruction

how do we find the right projections?

# principal components analysis

## through the lens of eigendecomposition



for a matrix  $X$  of size  $(t \times n)$ , where  $t$ : time and  $n$ : neurons

the projection that maximises variance in  $X$  is the first eigenvector of the covariance matrix  $C$  of mean-centered  $X$

$$C = X^T X$$

$$CV = \lambda V$$

$V$  and  $\lambda$  are eigenvectors and eigenvalues of  $C$  resp.

PCs are projection of your data onto eigenvectors of  $C$ :  $X'V$

in practice, we rarely derive  $V$  &  $\lambda$ , we instead use singular value decomposition as eigendecomposition is computationally less efficient



# principal components analysis

through the lens of singular value decomposition (SVD)

for a matrix  $X$  of size  $(t \times n)$ , where  $t$ : time and  $n$ : neurons

in eigendecomposition

$$C = X^T X$$

$$C V = \lambda V$$

$V$  and  $\lambda$  are eigenvectors and eigenvalues of  $C$  resp.

in SVD, we decompose  $X$  directly

$$X = U S V^T$$

$U$  is the left singular vector,  $S$  are singular values and  $V$  consists of right singular vectors

what's the relationship?

$$X^T X V = V S^2$$

$$C V = V S^2$$

singular values ( $S$ ) are the square root of eigenvalues and the right singular vectors ( $V$ ) are eigenvectors

we'll try both these methods and look at their equivalence in the hands-on session!

# principal components analysis

through the lens of singular value decomposition (SVD)

for a matrix  $X$  of size  $(t \times n)$ , where  $t$ : time and  $n$ : neurons

in eigendecomposition

$$C = X^T X$$

$$CV = \lambda V$$

$V$  and  $\lambda$  are eigenvectors and eigenvalues of  $C$  resp.

in SVD, we decompose  $X$  directly

$$X = USV^T$$

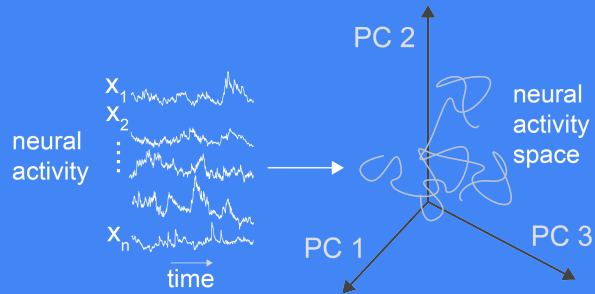
$U$  is the left singular vector,  $S$  are singular values and  $V$  consists of right singular vectors

what's the relationship?

if  $X = USV^T$ , then  $X^T X = VSU^T USV^T = VS^2V^T$

if we multiply  $V^T$  on both sides, we obtain,  $X^T X V = VS^2$ , or  $CV = VS^2$

i.e singular values ( $S$ ) are the square root of eigenvalues and the right singular vectors ( $V$ ) are eigenvectors



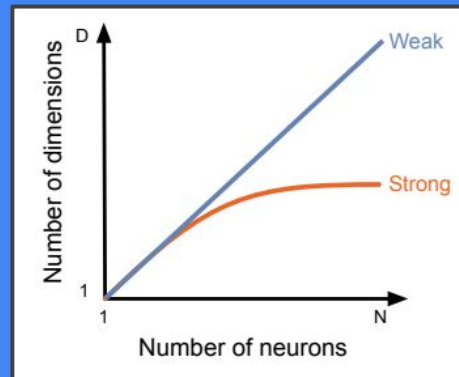
why perform  
dimensionality reduction?  
principal components analysis

**weak principle :**

dimensionality reduction is a convenient tool for making sense of complex data

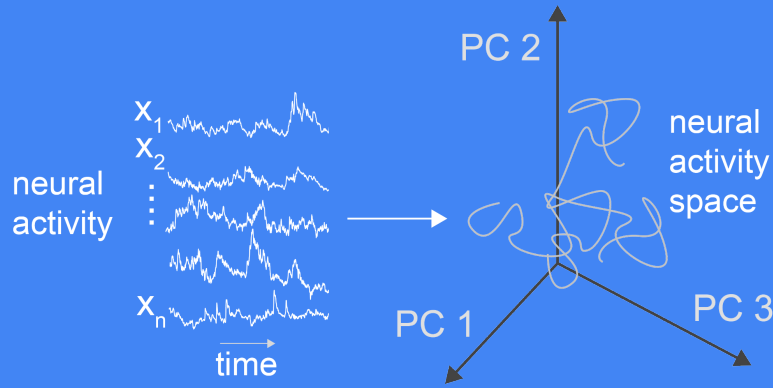
**strong principle :**

dimensionality reduction is a hypothesis for how neural circuits compute



# hands-on session part 1

## dimensionality reduction via PCA

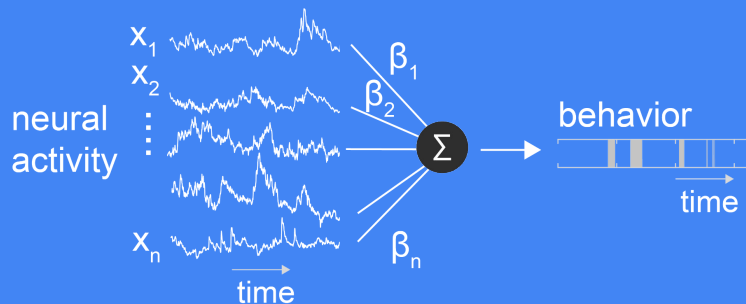


- learn to use a pre-made Principal Component Analysis (PCA) library.
- implement PCA with our own matrix operations.
- use Singular Value Decomposition (SVD) to build PCA.

# hands-on session part 2

## overfitting & regularization

how can we predict behavior from neural activity?



details of regression will be covered in day 3 & 5

- we'll fit a simple linear model (aka linear regression)
- learn how to identify overfitting
- combat overfitting using regularization and shuffling techniques
- implement linear regression from scratch.
- learn about methods that combine dimensionality reduction with regression.

# let's code!

**github repo:**

[https://github.com/cheninstitutecaltech/  
Caltech\\_DATASAI\\_Neuroscience\\_23](https://github.com/cheninstitutecaltech/Caltech_DATASAI_Neuroscience_23)

