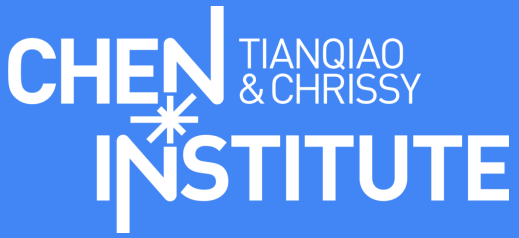
Chen Data Science & AI for Neuroscience Summer School



Caltech

Organizers

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

Adi Nair

Graduate Student

Lior S. Pachter

Bren Professor of Computational Biology and Computing and Mathematical Sciences

Pietro Perona

Allen E. Puckett Professor of Electrical Engineering

Sabera Talukder

Graduate Student

Yisong Yue

Professor of Computing and Mathematical Sciences

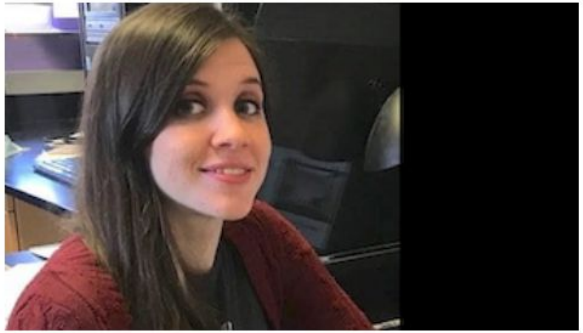
External Instructors

Eva Dyer

Assistant Professor, Georgia Tech

Jonathan Kao

Assistant Professor, UCLA



Ann Kennedy

Assistant Professor, Northwestern University

Chethan Pandarinath

Assistant Professor, Emory University and Georgia Tech

Logistics Team

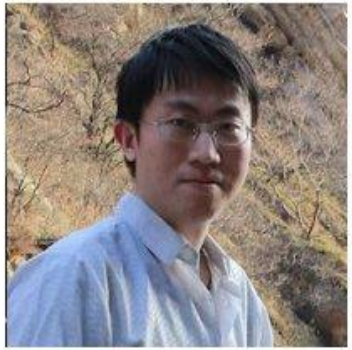
Helen O'Connor

Programs Coordinator

Mary Sikora

Executive Director, Tianqiao and Chrissy Chen Institute for Neuroscience

Teaching Assistants



Xuan Ma, Northwestern

Tara Chari, Caltech

Chris Versteeg, GA Tech

Mehdi Azabou, GA Tech

Brandon McMahan, UCLA

Daily Schedule

Today is a little different

- 9:00AM - 9:45AM: Breakfast
- 10:00AM - 11:00AM: Introductory Lecture
- 11:00AM - 1:00PM: Hands On Session
- 1:00PM - 2:00PM: Lunch
- 2:00PM - 3:45PM: Hands On Session
- 3:45PM - 4:15PM: Coffee Break
- 4:15PM - 6:15PM: Hands On Session

Daily Schedule

Today is a little different

- 9:00AM - 9:45AM: Breakfast
- 10:00AM - 11:00AM: Introductory Lecture
- 11:00AM - 1:00PM: Hands On Session
- 1:00PM - 2:00PM: Lunch
- 2:00PM - 3:45PM: Hands On Session
- 3:45PM - 4:15PM: Coffee Break
- 4:15PM - 6:15PM: Hands On Session

Typical day

- 9:00AM - 9:45AM: Breakfast
- 9:45AM - 10:45AM Introductory Lecture
- 10:45AM - 11:00AM: Bio Break
- 11:00AM - 12:00PM: Philosophy and General Questions
- 12:00PM - 1:00PM: Lunch
- 1:00PM - 2:00PM: Methods Lecture
- 2:00PM - 6:00PM: Hands On Session

Daily Schedule

Today is a little different

- 9:00AM - 9:45AM: Breakfast
- 10:00AM - 11:00AM: Introductory Lecture
- 11:00AM - 1:00PM: Hands On Session
- 1:00PM - 2:00PM: Lunch
- 2:00PM - 3:45PM: Hands On Session
- 3:45PM - 4:15PM: Coffee Break
- 4:15PM - 6:15PM: Hands On Session

Typical day

- 9:00AM - 9:45AM: Breakfast
- 9:45AM - 10:45AM Introductory Lecture
- 10:45AM - 11:00AM: Bio Break
- 11:00AM - 12:00PM: Philosophy and General Questions
- 12:00PM - 1:00PM: Lunch
- 1:00PM - 2:00PM: Methods Lecture
- 2:00PM - 6:00PM: Hands On Session

Bring your own data day

- 9:00AM - 9:45AM: Breakfast
- 10:00AM - 11:00AM Methods Lecture
- 11:00AM - 1:00PM: Hands On Session
- 1:00PM - 2:00PM: Lunch
- 2:00PM - 3:00PM: Methods Lecture
- 3:00PM - 6:00PM: Hands On Session

Ask a lot of questions,
feel free to interrupt!

Course website:

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

Github Repo:

https://github.com/SaberaTalukder/Chen_Institute_DataSAI_for_Neuroscience

Google Colab Pro+:

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

Slack Channel:

https://join.slack.com/t/cheninstitute-9181373/shared_invite/zt-1b54fb5gi-fuAhn51vD_qorBqTbXoKcQ

Bring your own data day

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

Eva Dyer, GA Tech

Ann Kennedy, Northwestern

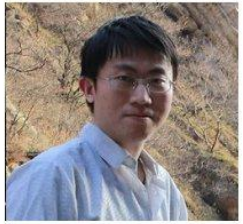
Lior Pachter, Caltech

Jonathan Kao, UCLA

Chethan Pandarinath, GA Tech

Sabera Talukder, Caltech

Aditya Nair, Caltech



Xuan Ma, Northwestern

Tara Chari, Caltech

Chris Versteeg, GA Tech

Mehdi Azabou, GA Tech

Brandon McMahan, UCLA


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

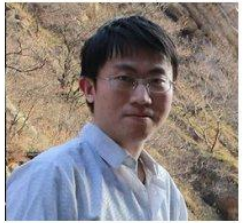

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

July 5th - 8th:

- ✦ computational basics


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

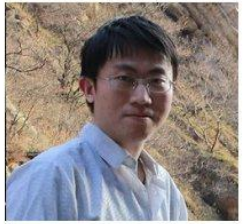

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

July 5th - 8th:

- ✦ computational basics
- ✦ dynamical time series


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

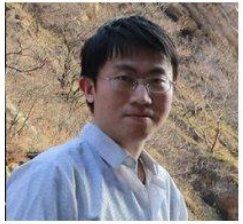

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

<u>July 5th - 8th:</u>

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

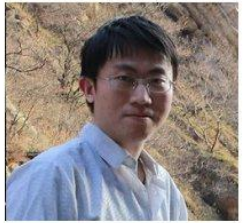

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

July 5th - 8th:

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

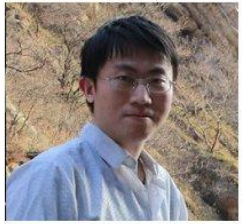

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

<u>July 5th - 8th:</u>

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning

<u>July 11th - 15th:</u>

- ✦ single cell seq & hypothesis testing


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

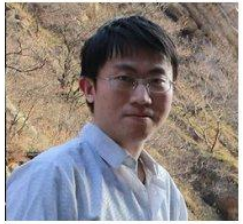

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

July 5th - 8th:

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning

July 11th - 15th:

- ✦ single cell seq & hypothesis testing
- ✦ bring your own data


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

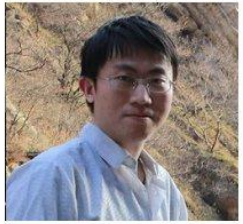

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

<u>July 5th - 8th:</u>

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning

<u>July 11th - 15th:</u>

- ✦ single cell seq & hypothesis testing
- ✦ bring your own data
- ✦ deep learning & LFADS


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

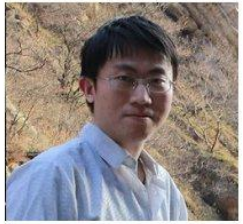

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

<u>July 5th - 8th:</u>

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning

<u>July 11th - 15th:</u>

- ✦ single cell seq & hypothesis testing
- ✦ bring your own data
- ✦ deep learning & LFADS
- ✦ generative modeling & MYOW


Eva Dyer, GA Tech


Ann Kennedy, Northwestern

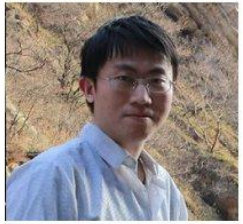

Lior Pachter, Caltech


Jonathan Kao, UCLA


Chethan Pandarinath, GA Tech


Sabera Talukder, Caltech


Aditya Nair, Caltech


Xuan Ma, Northwestern


Tara Chari, Caltech


Chris Versteeg, GA Tech


Mehdi Azabou, GA Tech


Brandon McMahan, UCLA

Topics covered:

July 5th - 8th:

- ✦ computational basics
- ✦ dynamical time series
- ✦ high dimensional data
- ✦ autoencoders & machine learning

July 11th - 15th:

- ✦ single cell seq & hypothesis testing
- ✦ bring your own data
- ✦ deep learning & LFADS
- ✦ generative modeling & MYOW
- ✦ RNNs & dynamical systems

Important notes!

- You have to wear either an N95 or KN95 indoors for all events (speakers may take them off while lecturing).
- You have to be in surveillance testing.
- Bring a reusable water bottle to the Seeley Mudd Estate.
- You cannot post pictures to social media that are taken at the Seeley Mudd Estate.

Engineering Mindset

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

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

- Architectures
- Mathematical Functions
- Code

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, because the only way you are going to learn is by doing what engineers do.

You are meant to struggle through this, because the only way you are going to learn is by doing what engineers do.

Your recourse when you don’t know what to do is:

You are meant to struggle through this, because the only way you are going to learn is by doing what engineers do.

Your recourse when you don’t know what to do is:

- ✦ Google → Stack Overflow, …

You are meant to struggle through this, because the only way you are going to learn is by doing what engineers do.

Your recourse when you don’t know what to do is:

- ✦ Google → Stack Overflow, …
- ✦ Ask your neighbor. Debug problems together!

You are meant to struggle through this, because the only way you are going to learn is by doing what engineers do.

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.

You are meant to struggle through this, because the only way you are going to learn is by doing what engineers do.

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!

Today we’re going to cover:

Computational basics!

Today we’re going to cover:

Computational basics!

Session 1
Dim Reduction via PCA

Today we're going to cover:

Computational basics!

Session 1
Dim Reduction via PCA

Session 2
Overfitting & Regularization

Today we're going to cover:

Computational basics!

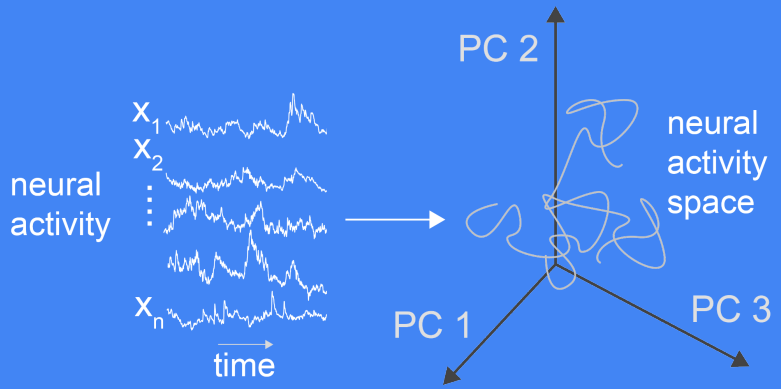
Session 1
Dim Reduction via PCA

Session 2
Overfitting & Regularization

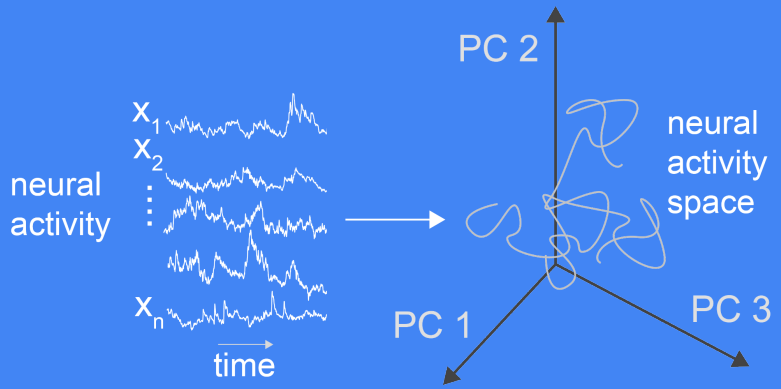
Session 3
Dataset Engineering

Session 1
Dim Reduction via PCA

Session 1
Dim Reduction via PCA

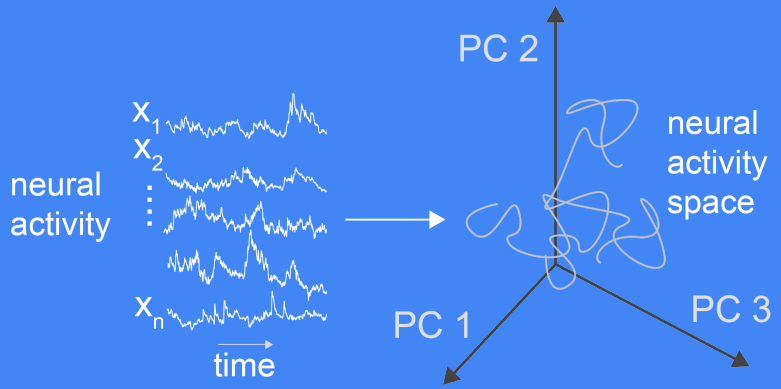


Session 1
Dim Reduction via PCA



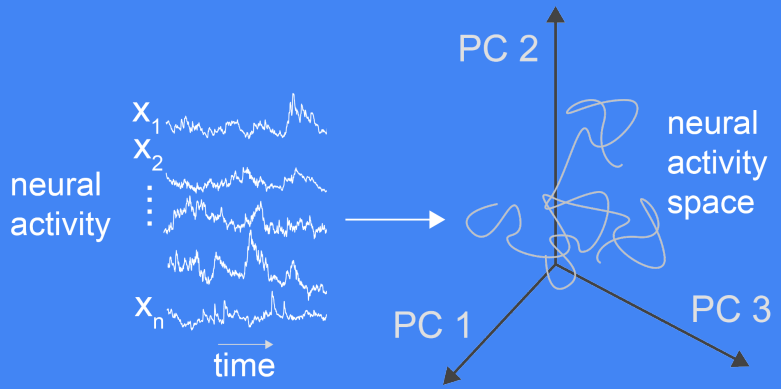
- Learn to use a pre-made Principal Component Analysis (PCA) library.

Session 1
Dim Reduction via PCA



- Learn to use a pre-made Principal Component Analysis (PCA) library.
- Implement PCA with our own matrix operations.

Session 1
Dim Reduction via PCA



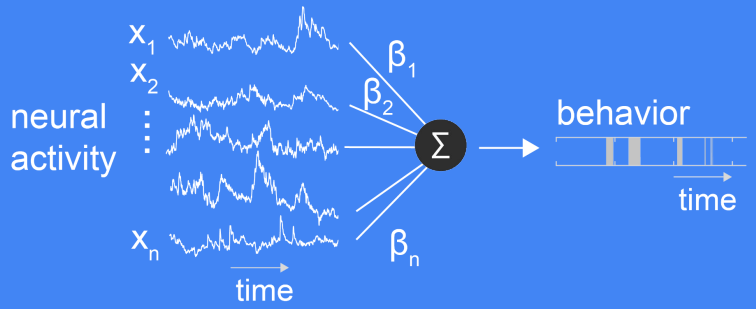
- Learn to use a pre-made Principal Component Analysis (PCA) library.
- Implement PCA with our own matrix operations.
- Use Singular Value Decomposition (SVD) to build PCA.

Session 2
Overfitting & Regularization

How can we predict behavior from neural activity?

Session 2
Overfitting & Regularization

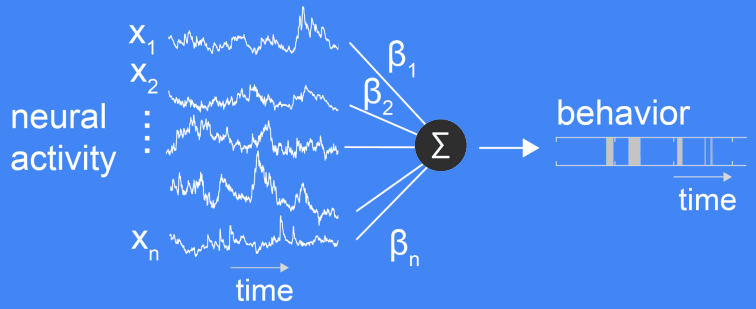
How can we predict behavior from neural activity?



Session 2
Overfitting & Regularization

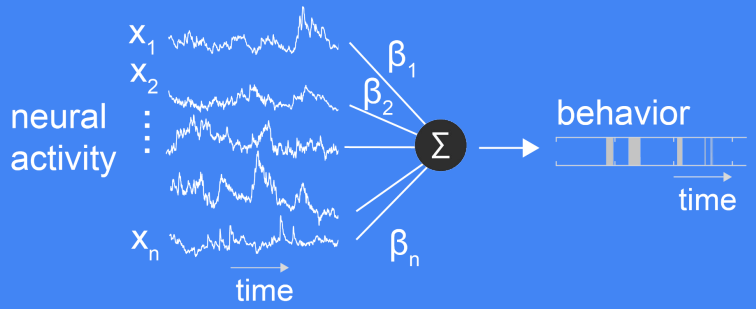
How can we predict behavior from neural activity?

- We'll fit a simple linear model (aka linear regression)



Session 2
Overfitting & Regularization

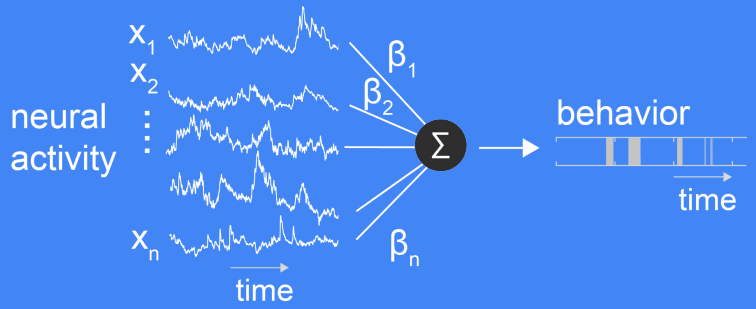
How can we predict behavior from neural activity?



- We'll fit a simple linear model (aka linear regression)
- Learn how to identify overfitting

Session 2
Overfitting & Regularization

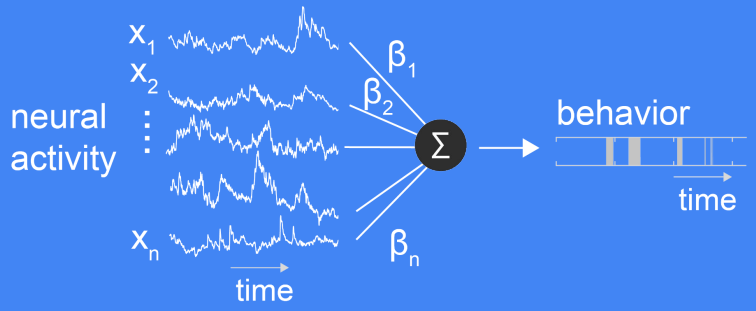
How can we predict behavior from neural activity?



- We'll fit a simple linear model (aka linear regression)
- Learn how to identify overfitting
- Combat overfitting using regularization and shuffling techniques

Session 2
Overfitting & Regularization

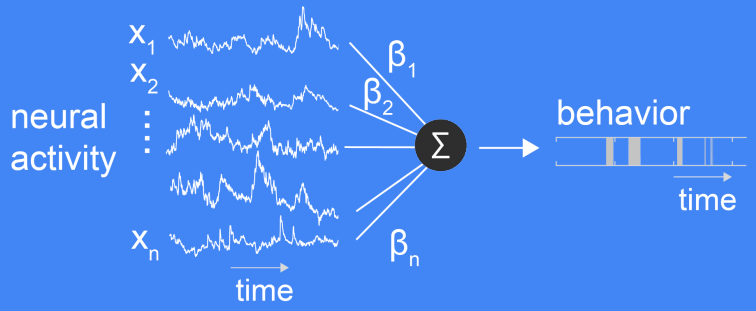
How can we predict behavior from neural activity?



- 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!

Session 2
Overfitting & Regularization

How can we predict behavior from neural activity?



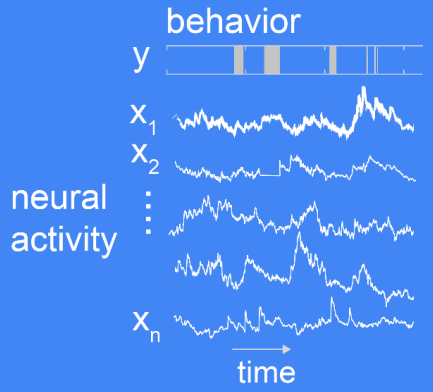
- We'll fit a simple linear model (aka linear regression)
- Learn how to identify overfitting
- Combat overfitting using regularization and shuffling techniques
- Implement linear regression from scratch!
- Learn about methods that combine dimen. reduction with regression!

Session 3
Dataset Engineering

How can we transform our data to make it better for ML methods?

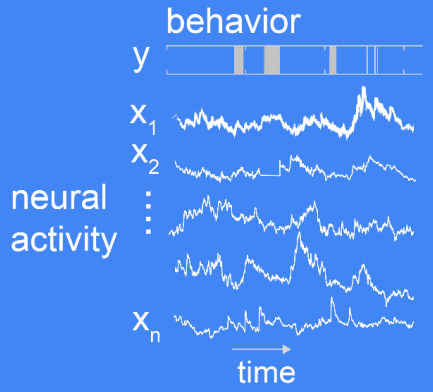
Session 3
Dataset Engineering

How can we transform our data to make it better for ML methods?



Session 3
Dataset Engineering

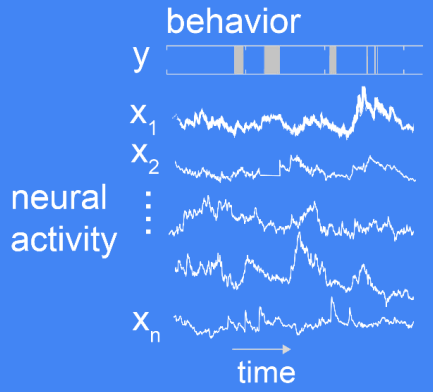
How can we transform our data to make it better for ML methods?



- Handle class imbalance using resampling methods

Session 3
Dataset Engineering

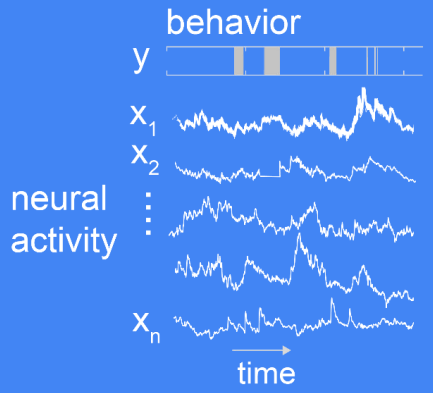
How can we transform our data to make it better for ML methods?



- Handle class imbalance using resampling methods
- Learn to identify and handle missing data using interpolation

Session 3
Dataset Engineering

How can we transform our data to make it better for ML methods?



- Handle class imbalance using resampling methods
- Learn to identify and handle missing data using interpolation
- Learn to denoise data using filters

Let’s code!

Github Repo:
https://github.com/SaberaTalukder/Chen_Institute_DataSAI_for_Neuroscience