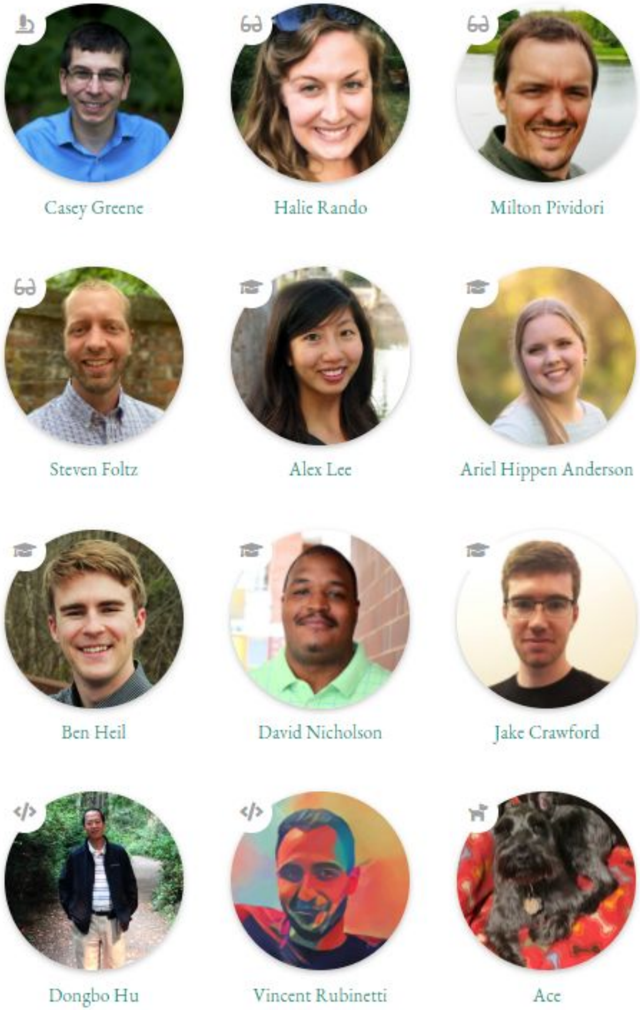
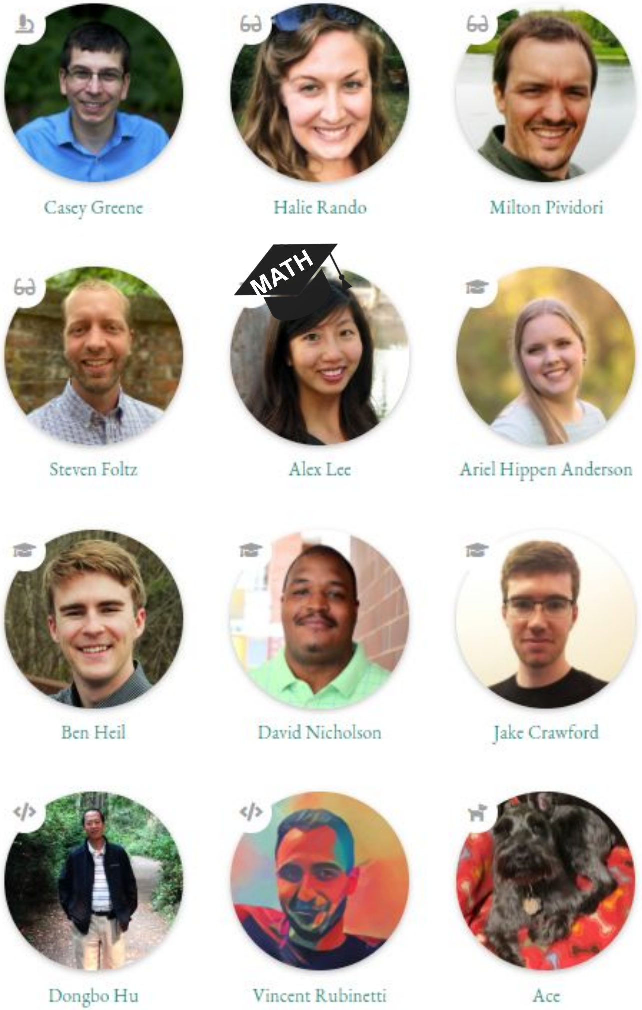
Deep Learning workshop part III

29 October 2020

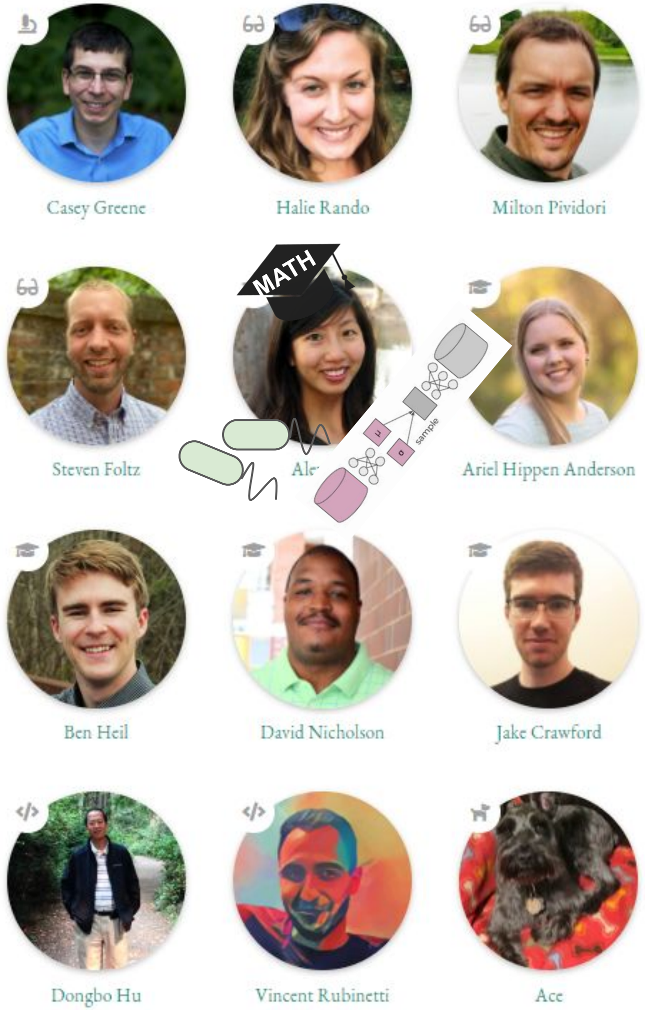
Greene Lab



- 4th year phD student in Casey Greene's lab at UPenn



- 4th year phD student in Casey Greene's lab at UPenn
- Hard core math major



- 4th year phD student in Casey Greene’s lab at UPenn
- Hard core math major
- Research uses deep learning methods to extract patterns from gene expression data

Recurrent Neural Networks (RNN)

What is sequential data?

Data where order matters

Examples:

- Time-series
- Sentences

Problems using sequential data

Task: Can you predict the next word given the context?

We finish each other's ________.

The fish are in the ________.

https://openai.com/blog/better-language-models/

Problems using sequential data

Task: Can you translate this sentence?

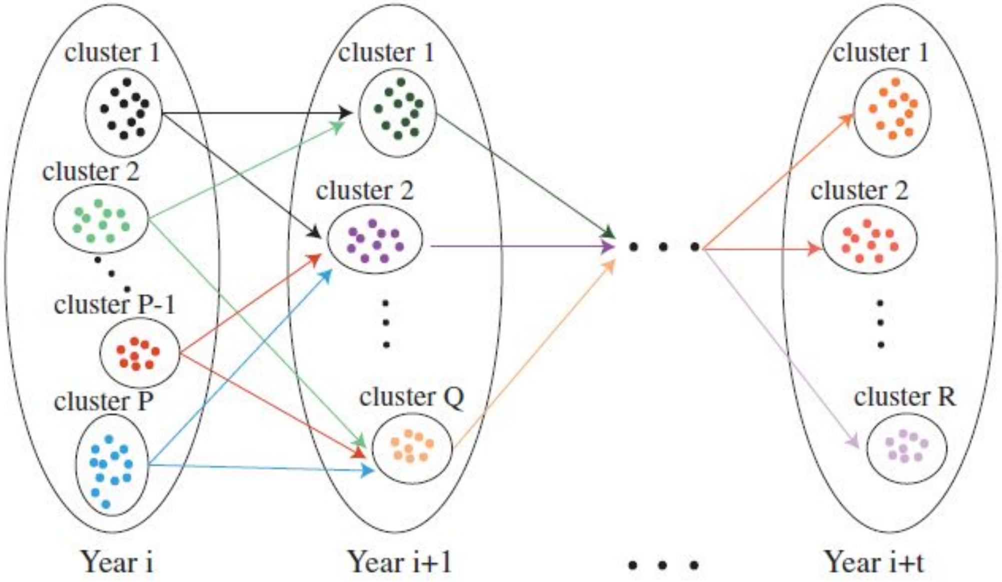
Je t'aime. **FRENCH**

__________. **ENGLISH**

https://arxiv.org/pdf/1409.3215.pdf
https://arxiv.org/pdf/1409.1259.pdf
https://arxiv.org/pdf/1409.0473.pdf

Problems using sequential data

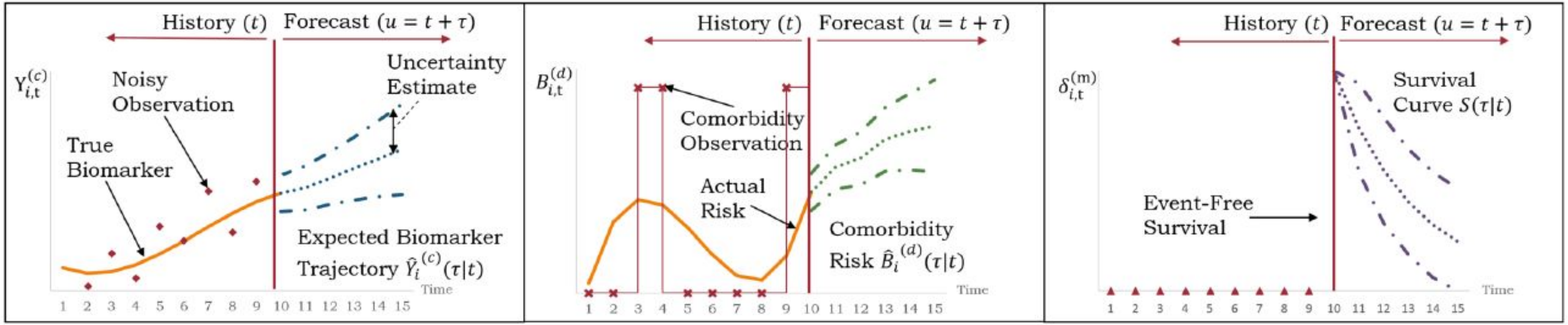
Task: Can you predict how influenza will mutate?



https://academic.oup.com/bioinformatics/article/36/9/2697/5717964

Problems using sequential data

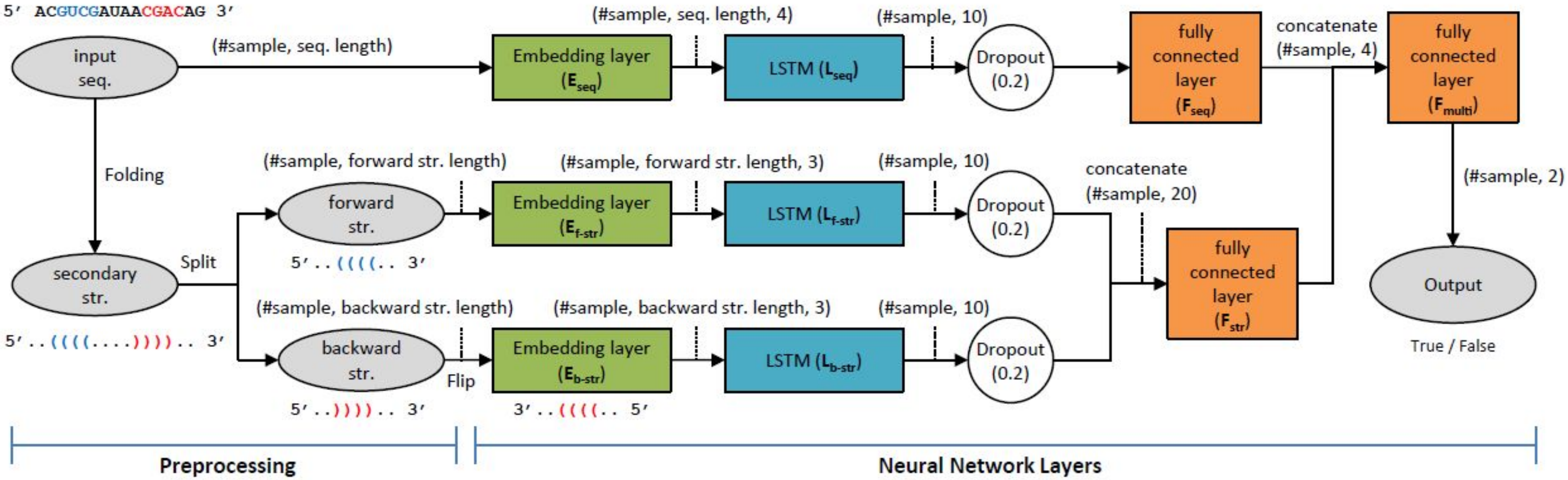
Task: Can you predict clinical outcome in the ICU?



https://arxiv.org/pdf/1803.10254.pdf

Problems using sequential data

Task: Can you predict if a sequence is a miRNA?



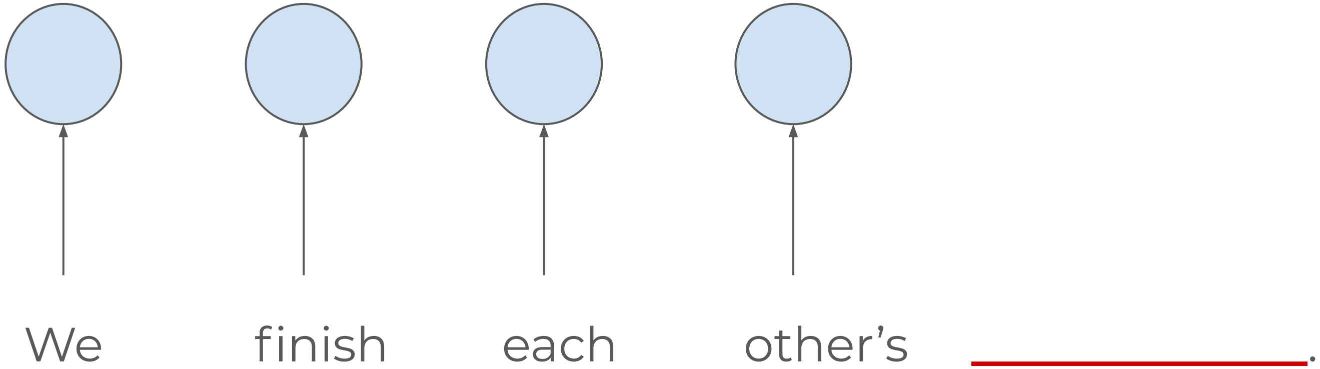
https://arxiv.org/pdf/1605.00017.pdf

We need to learn the sequence!

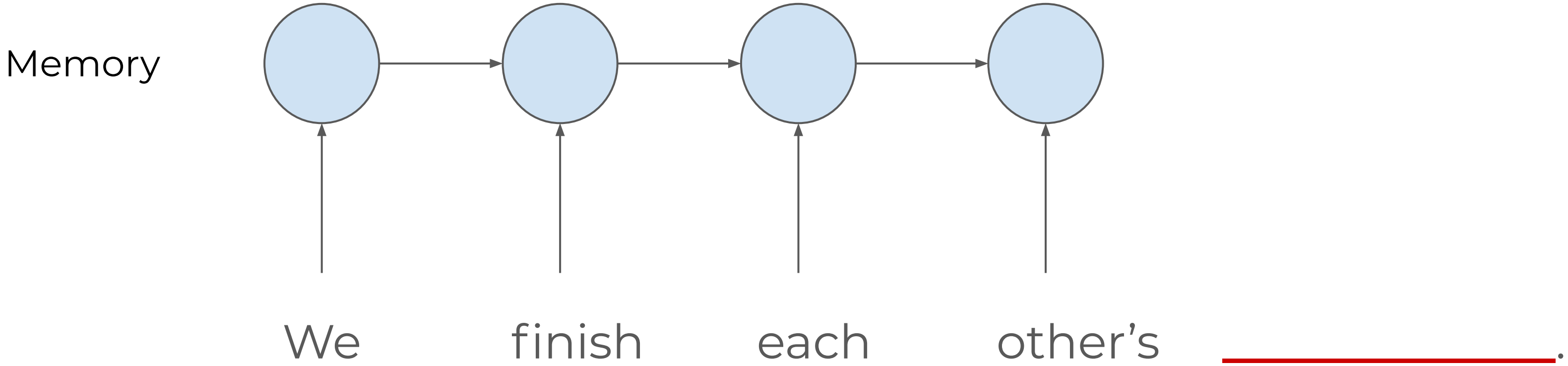
How do RNN's work?

We finish each other's ______________.

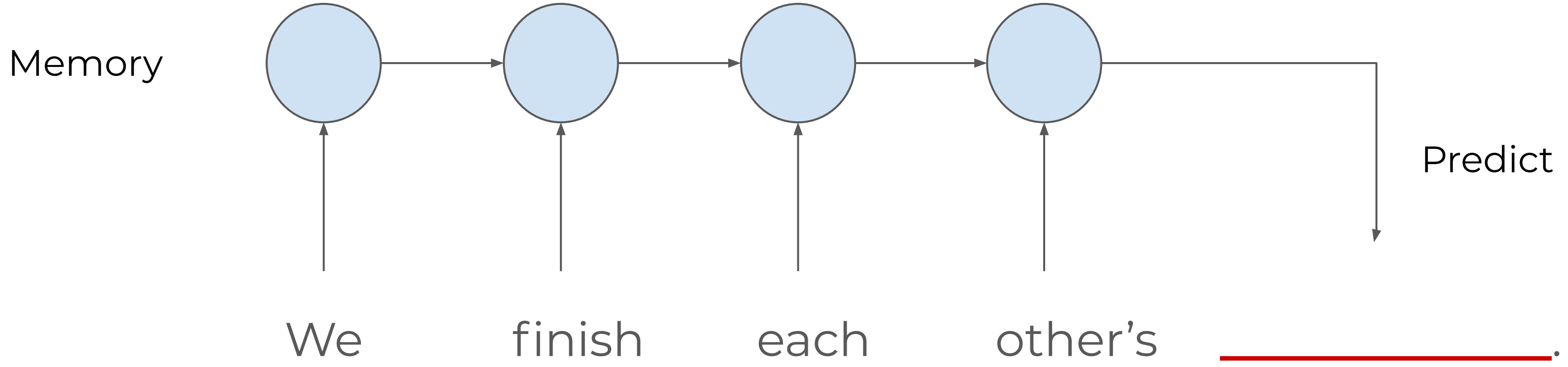
How do RNN's work?



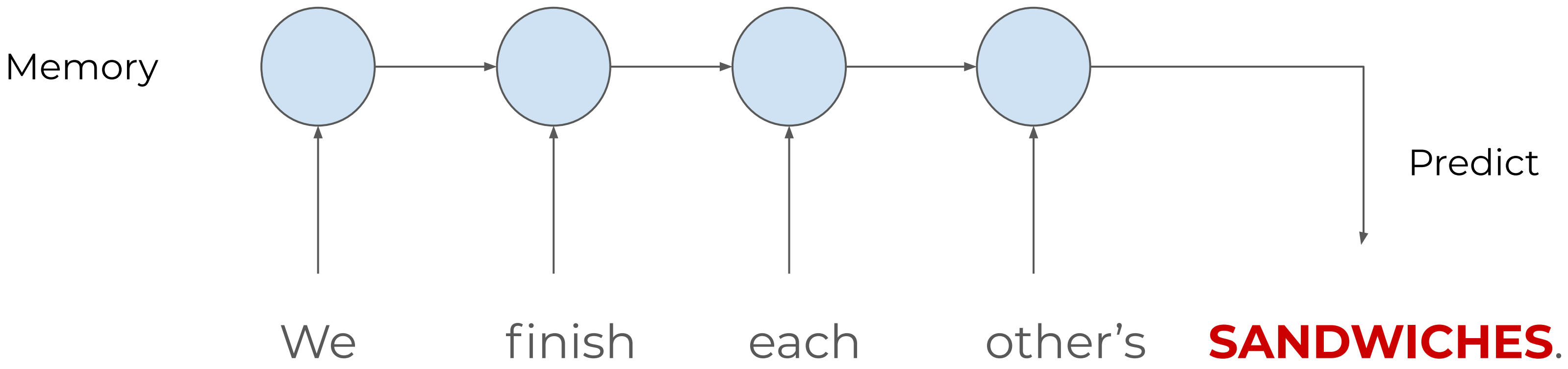
How do RNN's work?



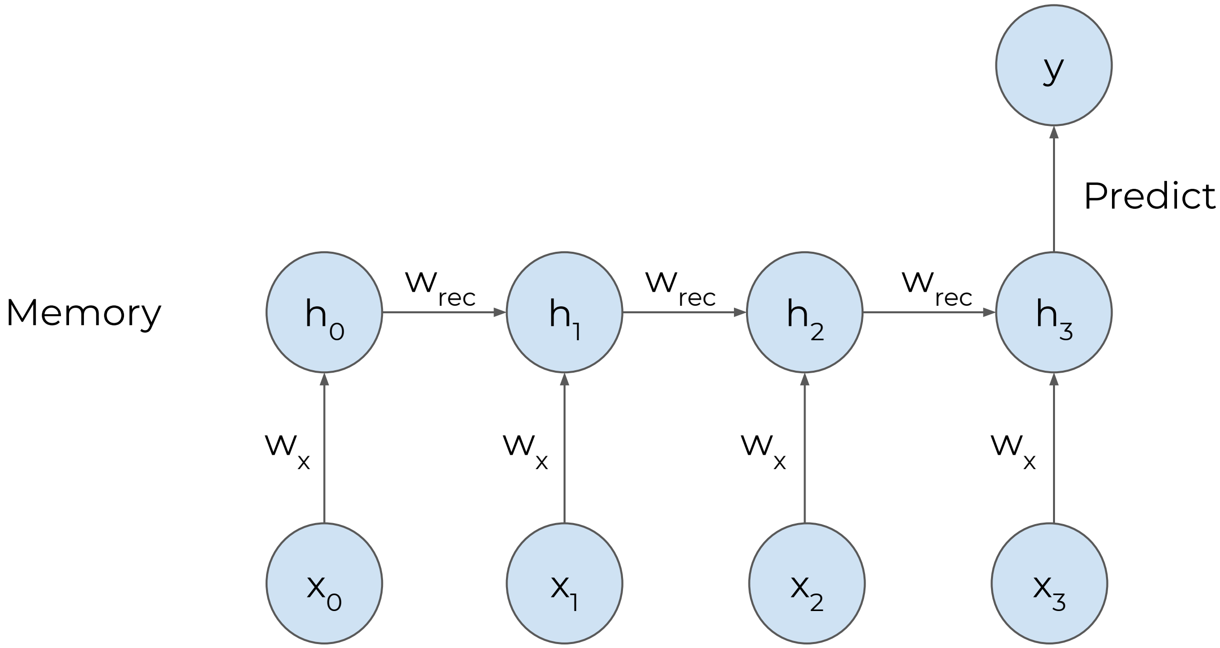
How do RNN's work?



How do RNN's work?

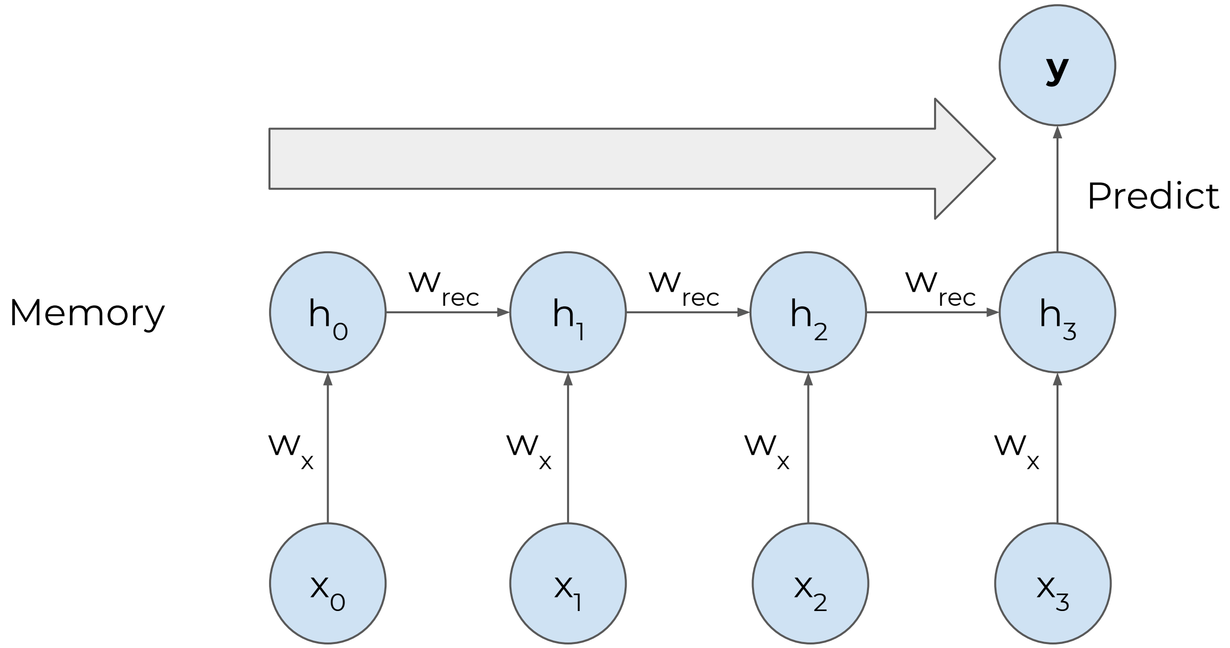


How do RNN's work?



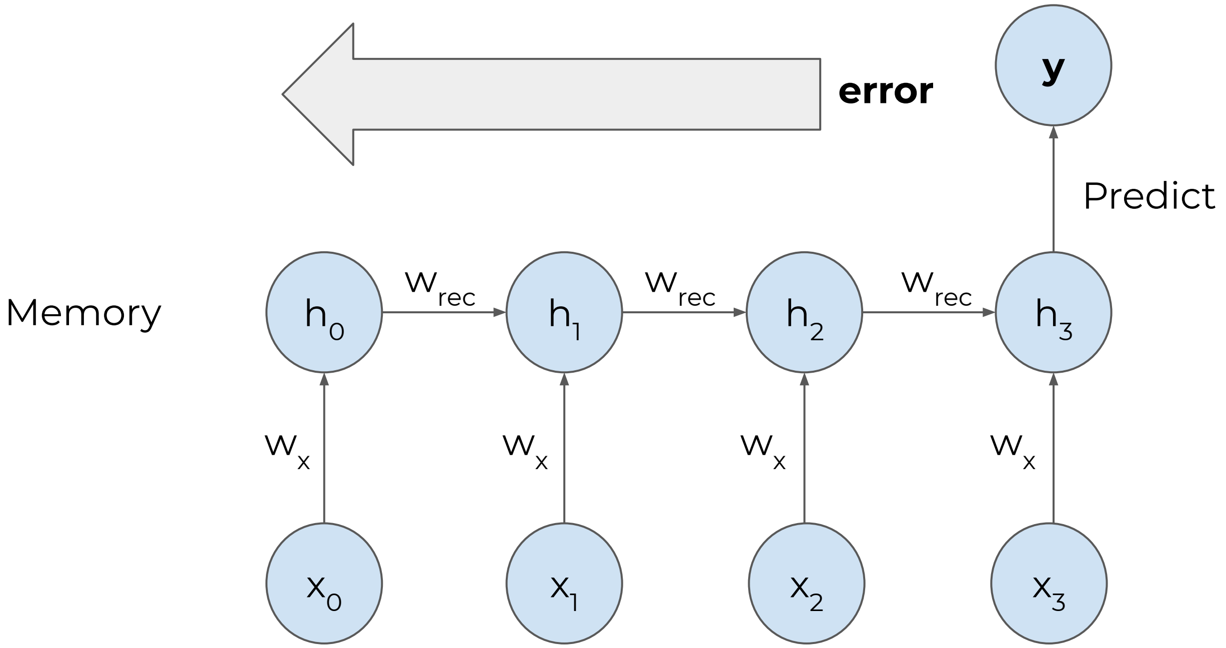
https://peterroelants.github.io/posts/rnn-implementation-part01/

How do RNN's work?



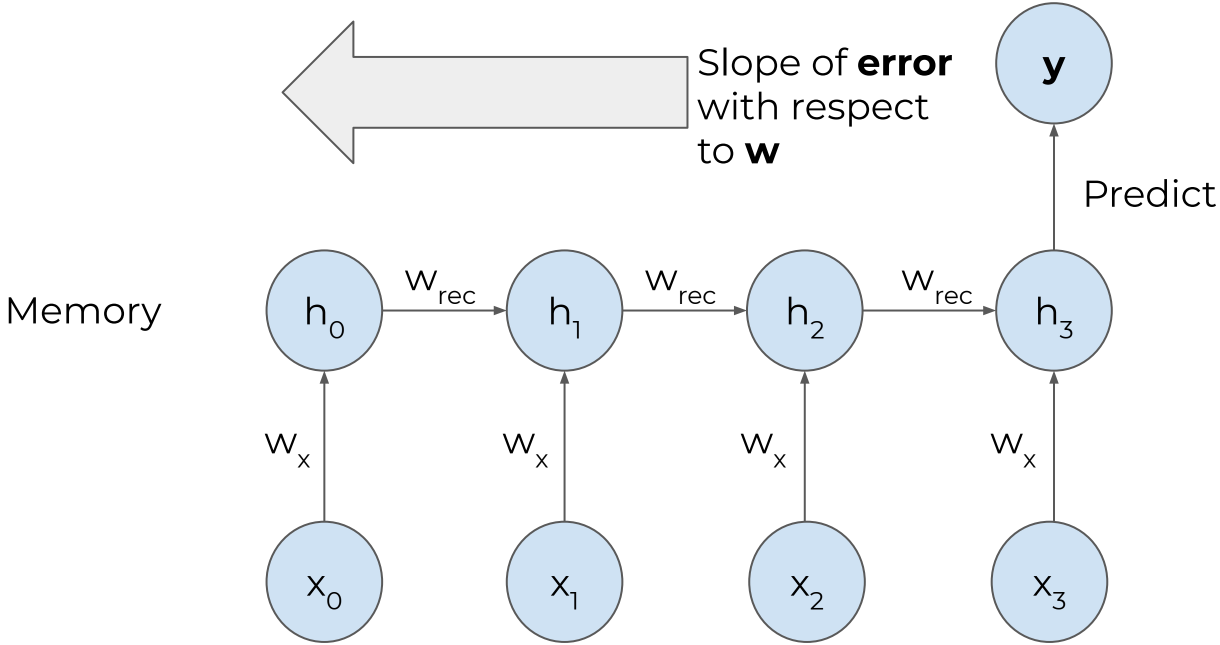
https://peterroelants.github.io/posts/rnn-implementation-part01/

How do RNN's work?



https://peterroelants.github.io/posts/rnn-implementation-part01/

How do RNN's work?



https://peterroelants.github.io/posts/rnn-implementation-part01/

Just a little bit of math

Slope of **error** with respect to **w** $= \dfrac{d\,\mathbf{error}}{d\,\mathbf{w}}$

Just a little bit of math

Slope of **error** with respect to **w** $= \dfrac{d\boldsymbol{error}}{d\boldsymbol{w}} = \dfrac{\partial\boldsymbol{error}}{\partial\boldsymbol{w}}$

Just a little bit of math

$$\frac{\partial \boldsymbol{error}}{\partial \boldsymbol{w}} = \frac{\partial \boldsymbol{error}}{\partial \boldsymbol{y}} \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{w}} \right)$$

Just a little bit of math

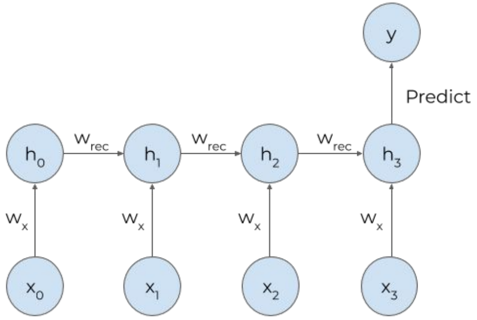
$$\frac{\partial \boldsymbol{error}}{\partial \boldsymbol{w}} = \frac{\partial \boldsymbol{error}}{\partial \boldsymbol{y}} \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{w}} \right)$$

Standard loss
function
(i.e. RMSE)

Just a little bit of math

$$\frac{\partial \boldsymbol{error}}{\partial \boldsymbol{w}} = \frac{\partial \boldsymbol{error}}{\partial \boldsymbol{y}} \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{w}} \right)$$

$$\frac{\partial}{\partial \boldsymbol{w}} \left(y = h_3 \right)$$

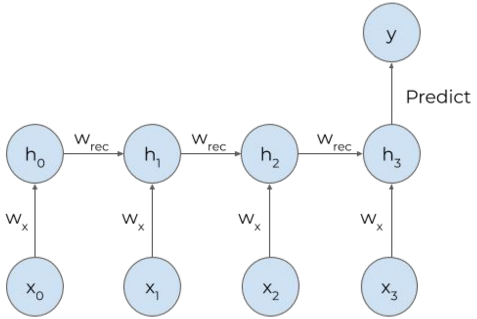


Just a little bit of math

$$\frac{\partial \mathbf{error}}{\partial \mathbf{w}} = \frac{\partial \mathbf{error}}{\partial \mathbf{y}} \left(\frac{\partial \mathbf{y}}{\partial \mathbf{w}} \right)$$

$$\frac{\partial}{\partial \mathbf{w}} \left(y = \mathbf{h_3} \right)$$

$$\frac{\partial}{\partial \mathbf{w}} \left(y = w_x x_3 + w_{rec} h_2 \right)$$



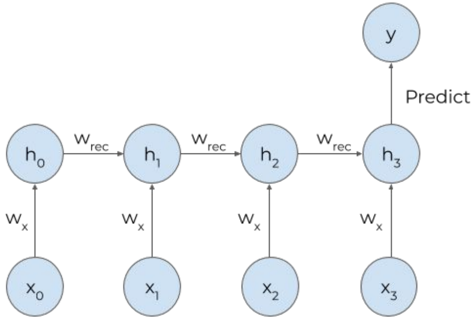
Just a little bit of math

$$\frac{\partial \mathbf{error}}{\partial \mathbf{w}} = \frac{\partial \mathbf{error}}{\partial \mathbf{y}} \left(\frac{\partial \mathbf{y}}{\partial \mathbf{w}} \right)$$

$$\frac{\partial}{\partial \mathbf{w}} \left(y = h_3 \right)$$

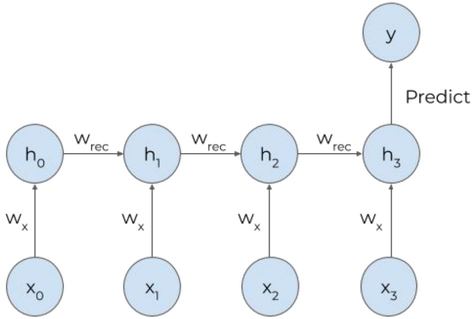
$$\frac{\partial}{\partial \mathbf{w}} \left(y = w_x x_3 + w_{rec} \boldsymbol{h_2} \right)$$

$$\frac{\partial}{\partial \mathbf{w}} \left(y = w_x x_3 + w_{rec}(w_x x_2 + w_{rec} h_1) \right)$$



Just a little bit of math

$$\frac{\partial \boldsymbol{error}}{\partial \boldsymbol{w}} = \frac{\partial \boldsymbol{error}}{\partial \boldsymbol{y}} \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{w}} \right)$$



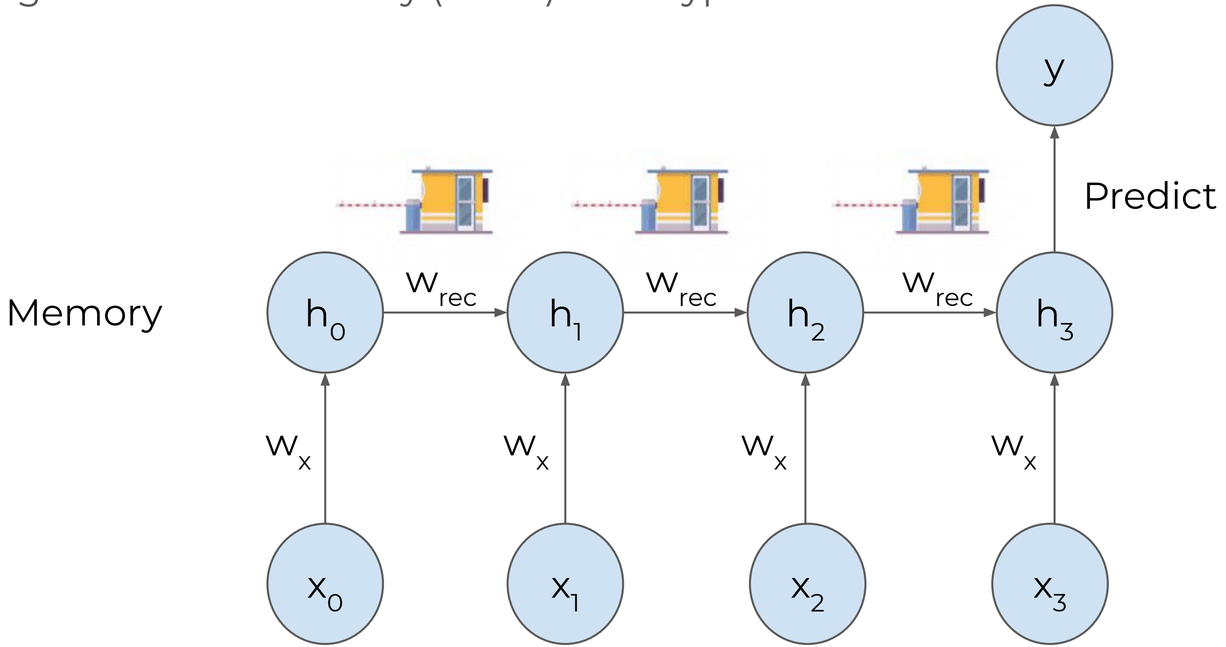
$$\frac{\partial}{\partial \boldsymbol{w}} \left(y = h_3 \right)$$

.
.
.

$$\frac{\partial}{\partial \boldsymbol{w}} \left(y = w_x x_3 + \boldsymbol{w_{rec}}(w_x x_2 + \boldsymbol{w_{rec}}(w_x x_1 + \boldsymbol{w_{rec}}(w_x x_0))) \right)$$

LSTMs to the rescue

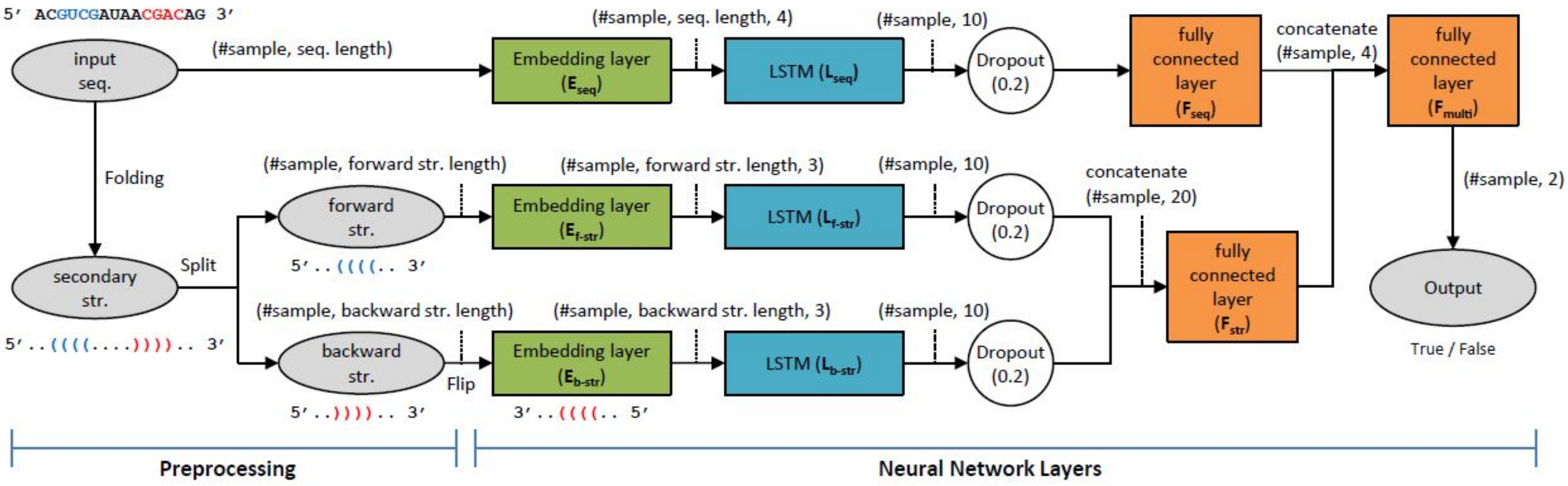
Long short term memory (LSTM) are a type of RNN



https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Problems using sequential data

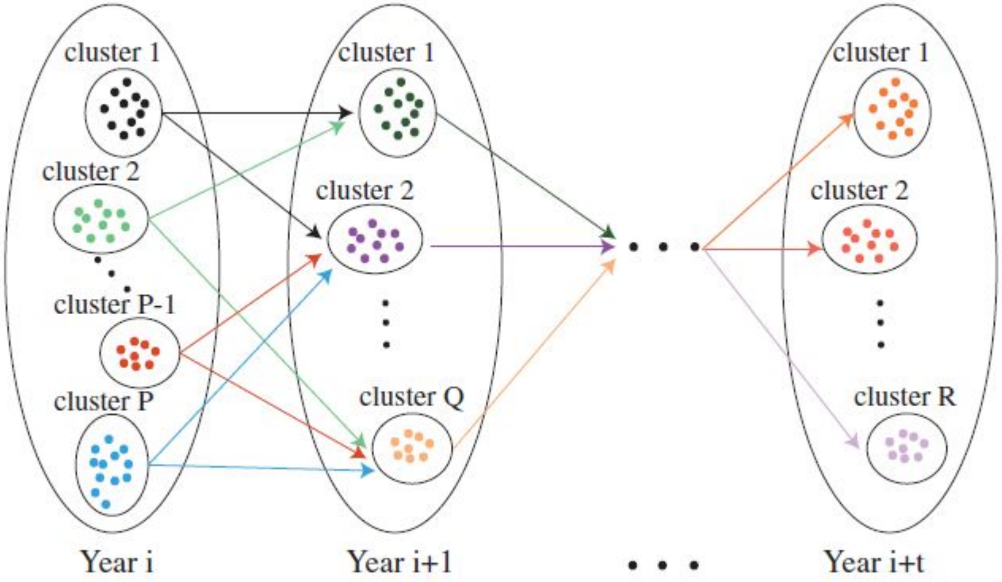
Task: Can you predict if a sequence is a miRNA?



https://arxiv.org/pdf/1605.00017.pdf

Problems using sequential data

Task: Can you predict how influenza will mutate?



https://academic.oup.com/bioinformatics/article/36/9/2697/5717964

Takeaway

- RNN-related methods can be useful for sequential data

Takeaway

- RNN-related methods can be useful for sequential data
- RNNs can take three forms
  - Vanilla - regular RNNs
  - LSTMs - Long short term memory networks
  - GRU - Gated Recurrent Unit (similar to LSTM but less parameters)

Takeaway

- RNN-related methods can be useful for sequential data
- RNNs can take three forms
  - Vanilla - regular RNNs
  - LSTMs - Long short term memory networks
  - GRU - Gated Recurrent Unit (similar to LSTM but less parameters)
- However the curse of dimensionality is real!
  - Think about time-series analysis --> node per time point
  - Too many time points --> very hard to train

Questions?

Dimensionality reduction methods

What are “dimensionality reduction methods”?

Why should we use them?

High dimensional data is everywhere

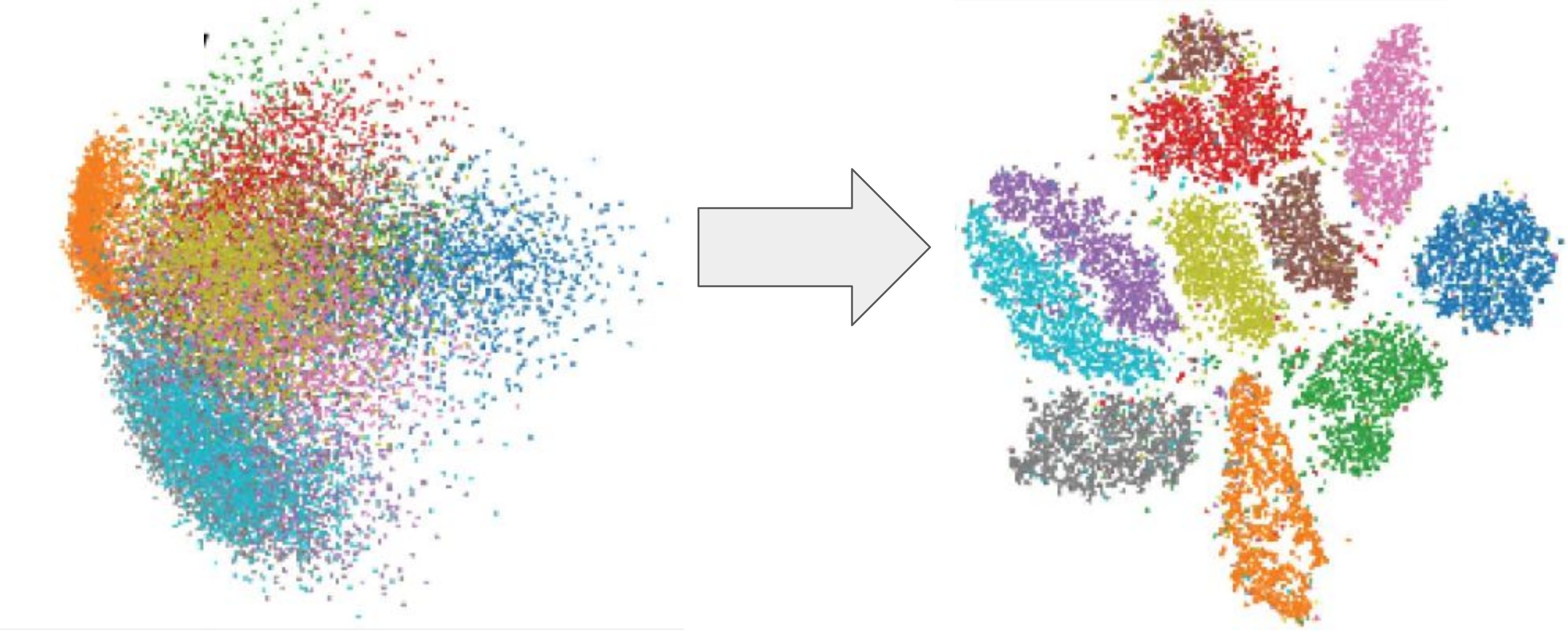
1200x1200

24 x 17788

	BTNL8	LINC01134	HEATR4	ACO1	PLPP3
SRR493937	0.316291	0.037657	0.271263	7.680846	35.5811
SRR493938	0.211909	0.089802	0.270260	7.783635	34.7091
SRR493939	0.031951	0.180184	0.242934	3.674145	9.25606
SRR493940	0.072871	0.188795	0.302474	3.471724	9.36842
SRR493941	0.314067	0.089359	0.211705	6.003360	53.3867

However, data is intrinsically low dimensional

High dimensional data **Low dimensional patterns**



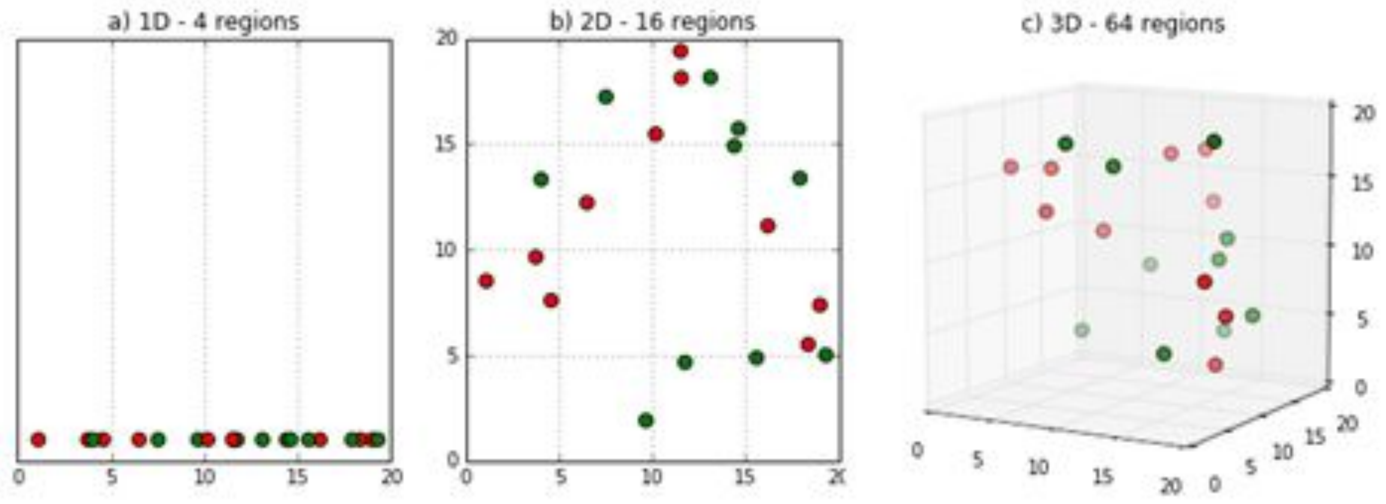
Data is intrinsically low dimensional

Data is intrinsically low dimensional

```
(24, 17788)
```

	BTNL8	LINC01134	HEATR4	ACO1	PLPP3
SRR493937	0.316291	0.037657	0.271263	7.680846	35.5811
SRR493938	0.211909	0.089802	0.270260	7.783635	34.7091
SRR493939	0.031951	0.180184	0.242934	3.674145	9.25606
SRR493940	0.072871	0.188795	0.302474	3.471724	9.36842
SRR493941	0.314067	0.089359	0.211705	6.003360	53.3867

Relationships harder to discover in high dimensions



https://deepai.org/machine-learning-glossary-and-terms/curse-of-dimensionality

What is dimensionality reduction and why use it?

Dimensionality reduction methods reduce high dimensional data into a reduced representation that captures the most salient part of the original data.

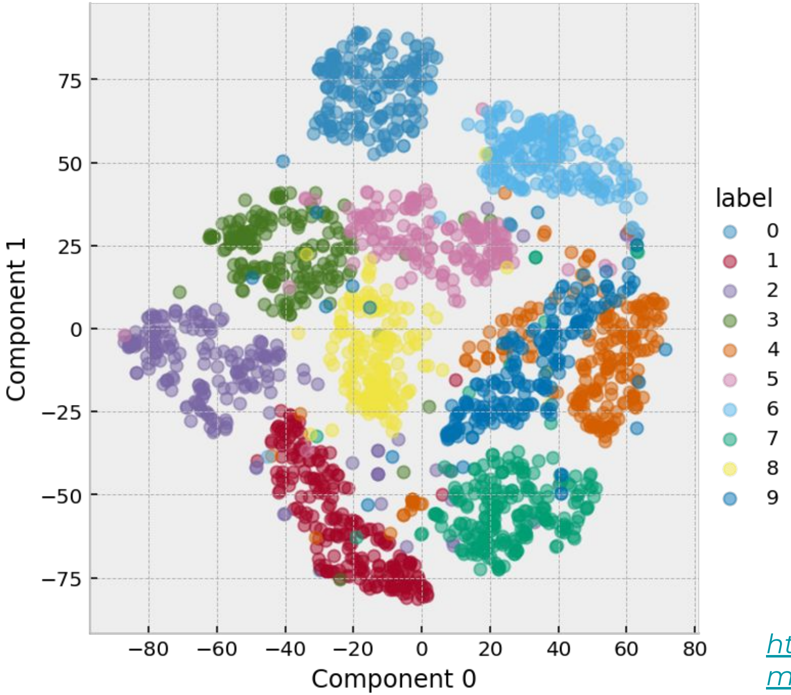
What is dimensionality reduction and why use it?

Dimensionality reduction methods reduce high dimensional data into a reduced representation that captures the most salient part of the original data.

Why use them?

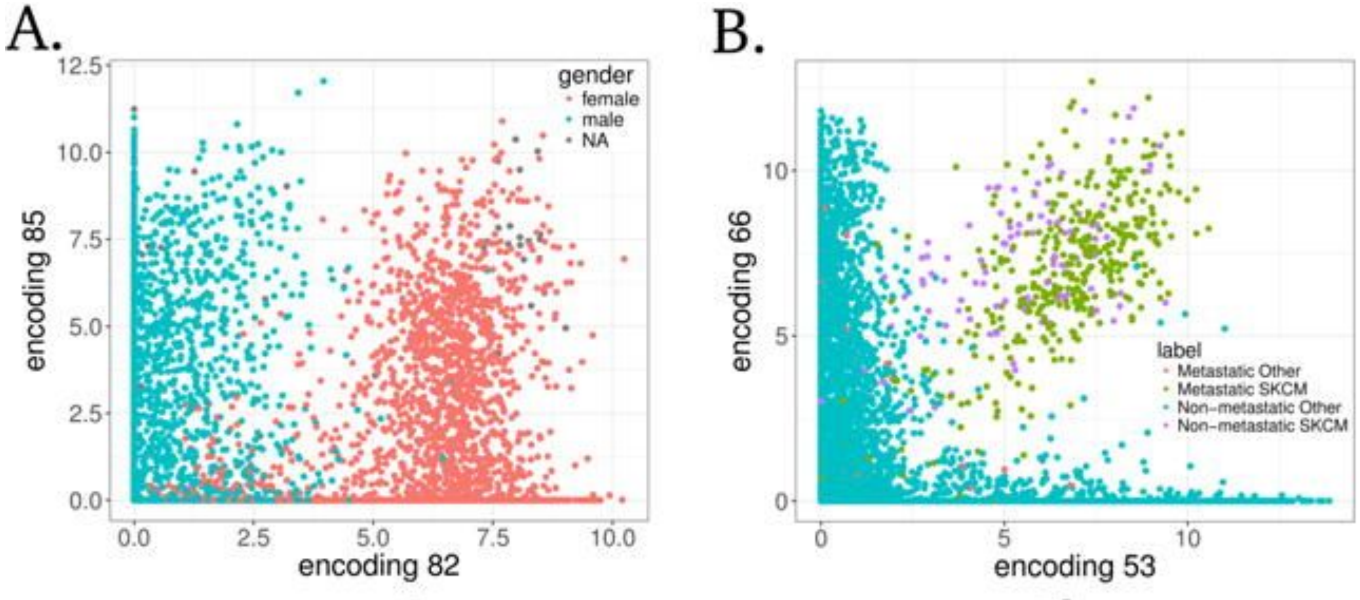
1. Patterns of interest live in low dimensions
2. Relationships are hard to find in high dimensions due to the *curse of dimensionality*

Low dimensional representation can extract clusters in MNIST



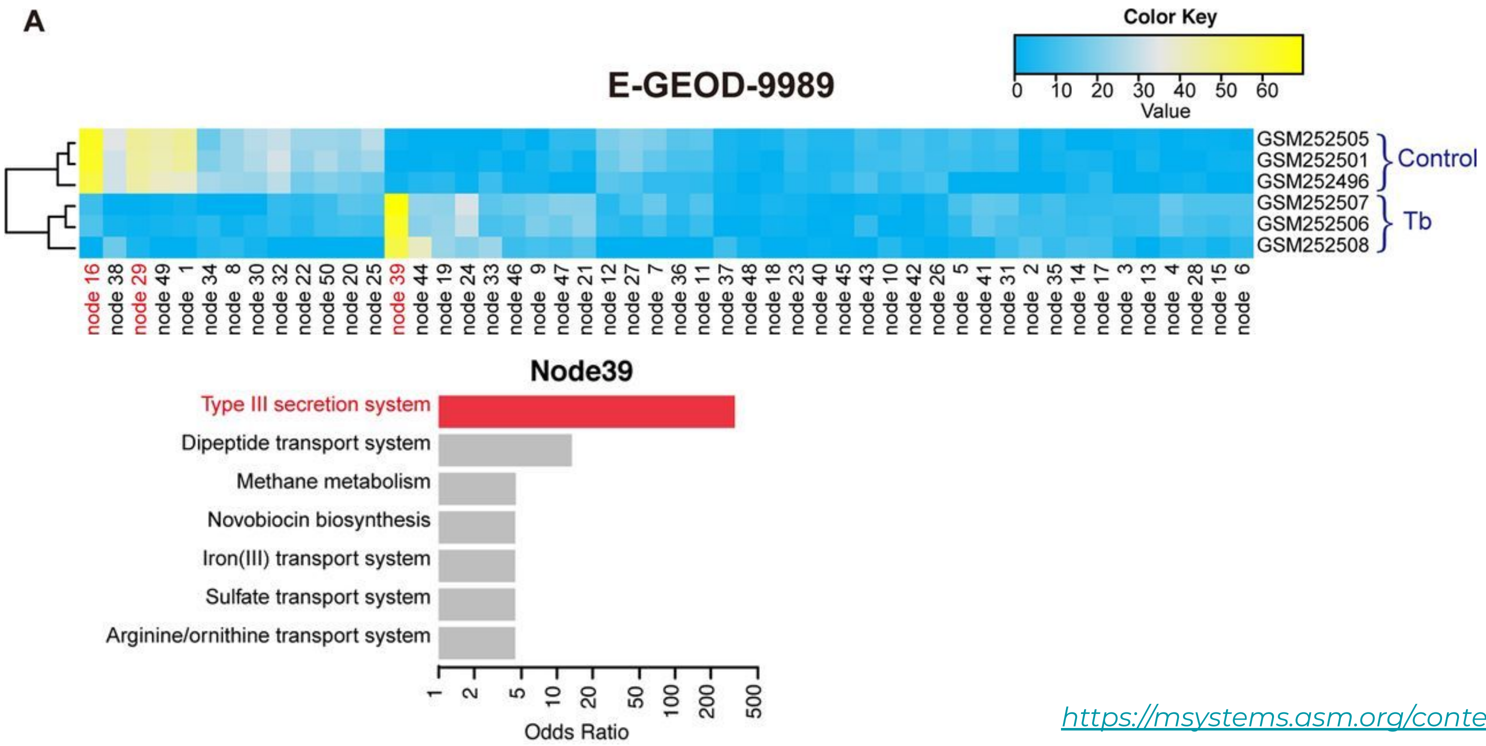
http://www.cse.chalmers.se/~richajo/dit866/lectures/l9/MNIST%20dimensionality%20reduction.html

Low dimensional features can differentiate between sex, SKCM tumors



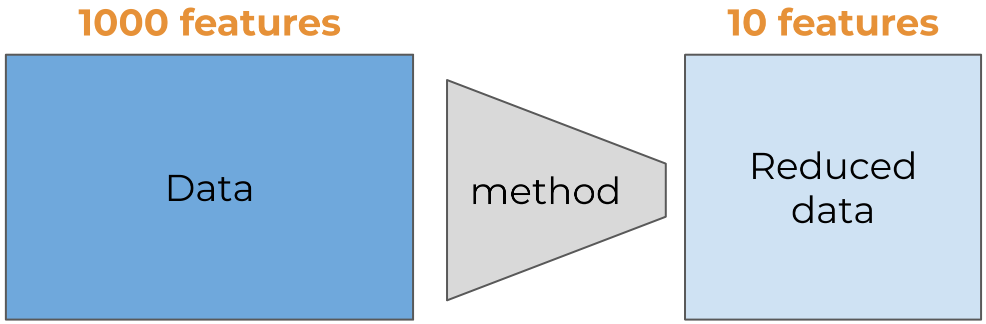
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5728678/

Low dimensional features can represent biological processes

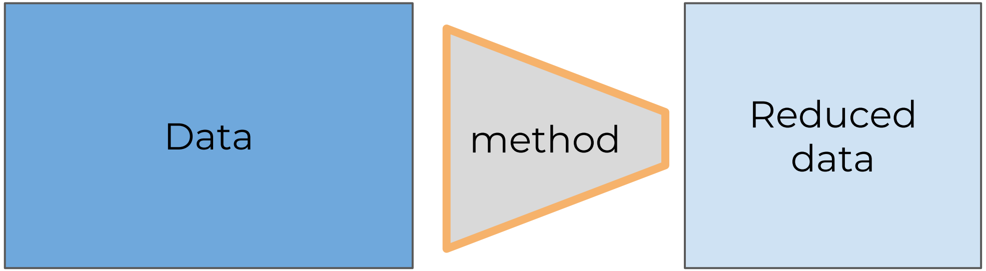


https://msystems.asm.org/content/1/1/e00025-15

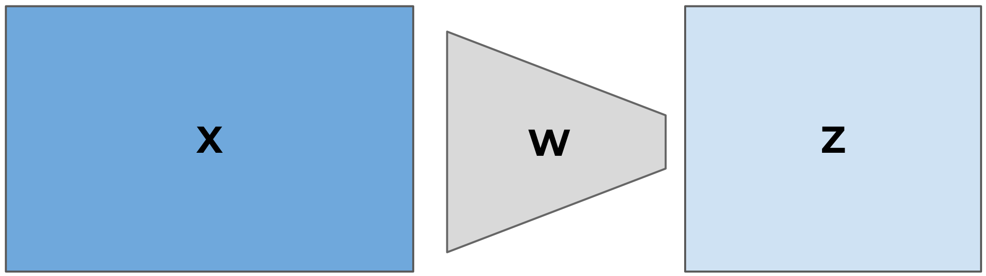
Dimensionality reduction methods learn a low dimensional representation of the data



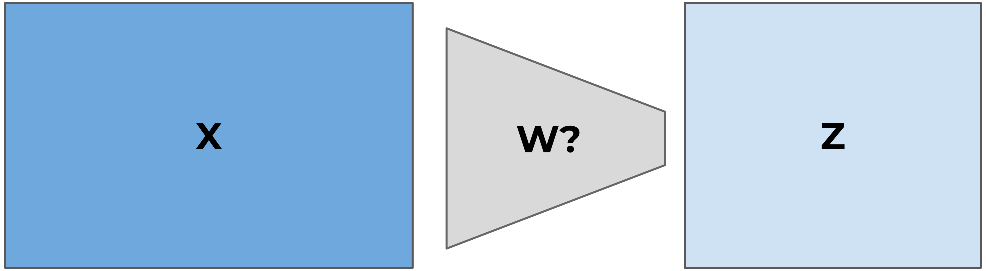
Dimensionality reduction methods learn a low dimensional representation of the data



Principal Component Analysis (PCA)

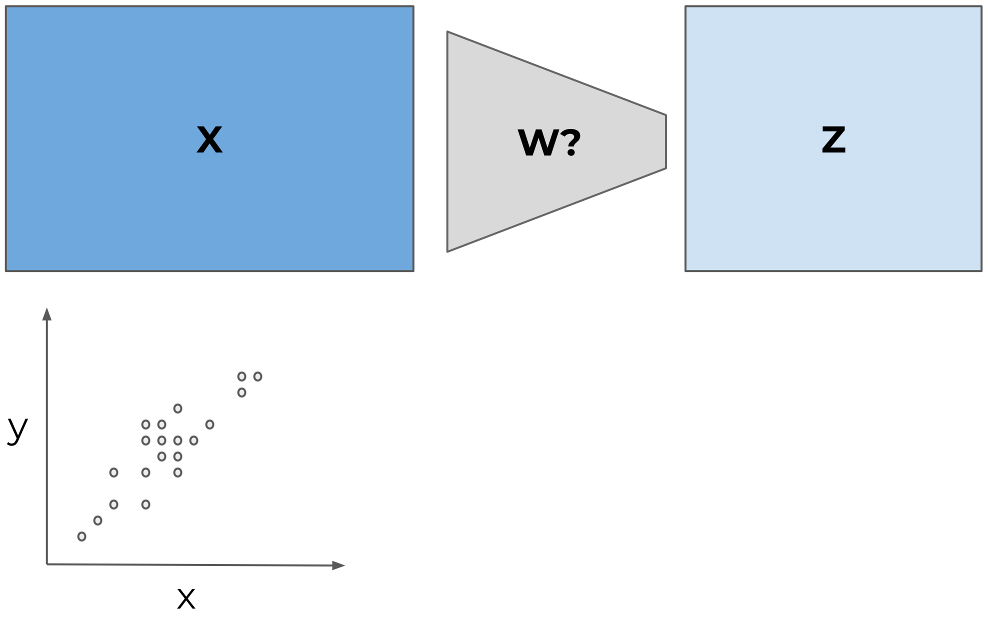


Principal Component Analysis (PCA)

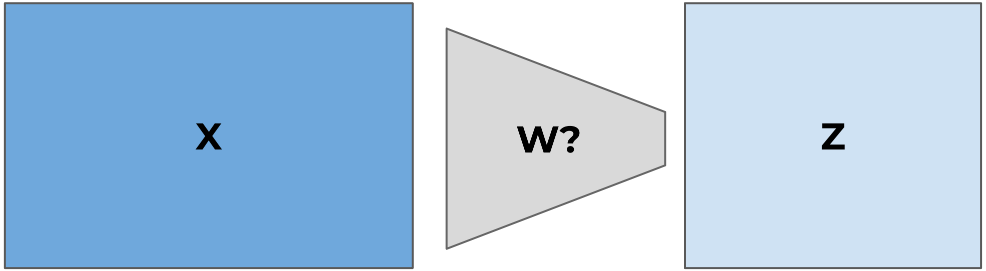


Goal: Find W that projects data on to a low dimensional space while preserving broad trends in the data

Principal Component Analysis (PCA)

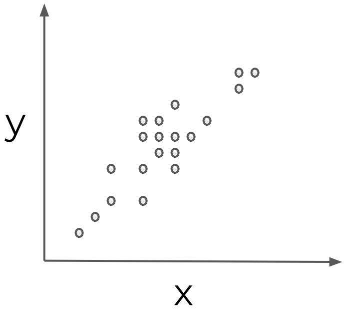


Principal Component Analysis (PCA)

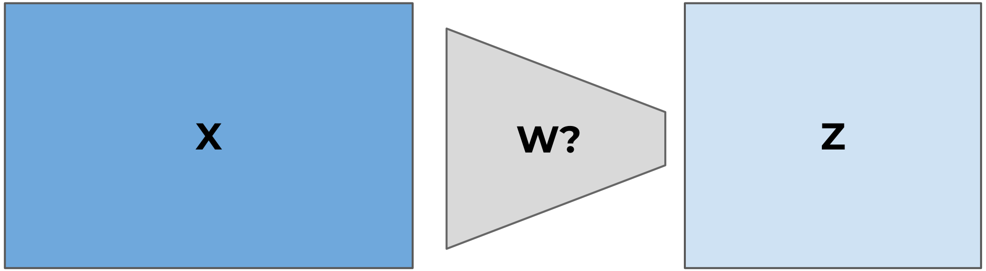


1. Calculate Covariance(**X**)

$$\begin{pmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{pmatrix}$$

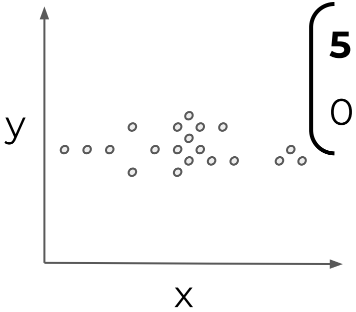


Principal Component Analysis (PCA)

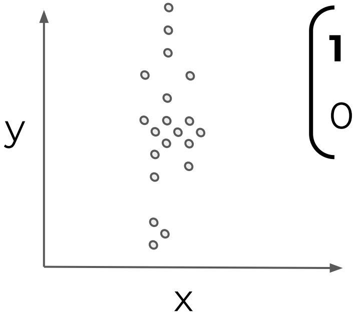


1. Calculate Covariance(**X**)

$$\begin{pmatrix} \mathbf{Var(x)} & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(x,y) & \mathbf{Var(y)} \end{pmatrix}$$

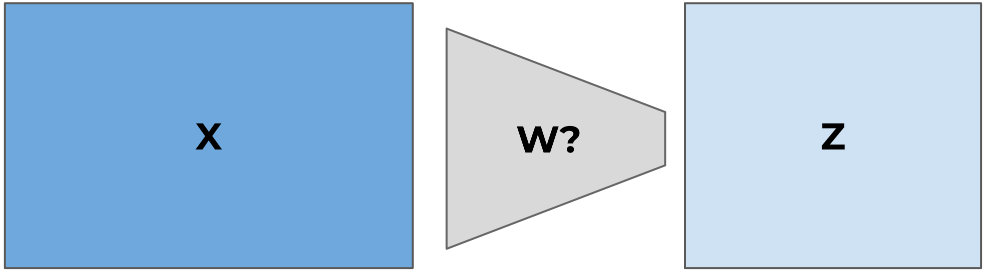


$$\begin{pmatrix} \mathbf{5} & 0 \\ 0 & \mathbf{1} \end{pmatrix}$$



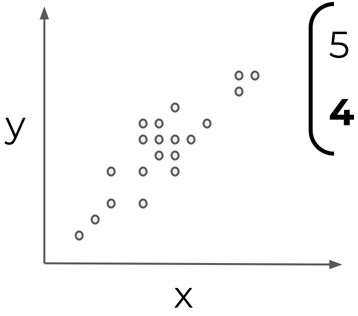
$$\begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{5} \end{pmatrix}$$

Principal Component Analysis (PCA)

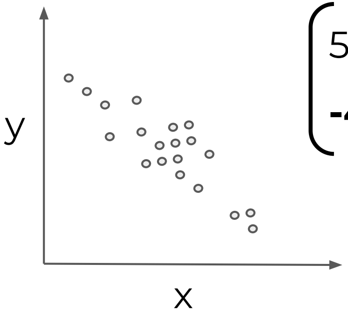


1. Calculate Covariance(**X**)

$$\begin{pmatrix} \mathrm{Var}(x) & \mathbf{Cov(x,y)} \\ \mathbf{Cov(x,y)} & \mathrm{Var}(y) \end{pmatrix}$$

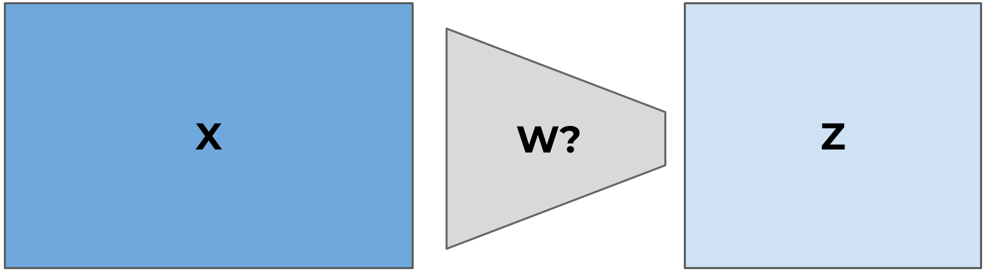


$$\begin{pmatrix} 5 & \mathbf{4} \\ \mathbf{4} & 6 \end{pmatrix}$$

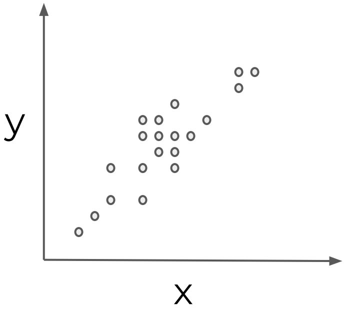


$$\begin{pmatrix} 5 & \mathbf{-4} \\ \mathbf{-4} & 6 \end{pmatrix}$$

Principal Component Analysis (PCA)



1. Calculate Covariance(**X**)

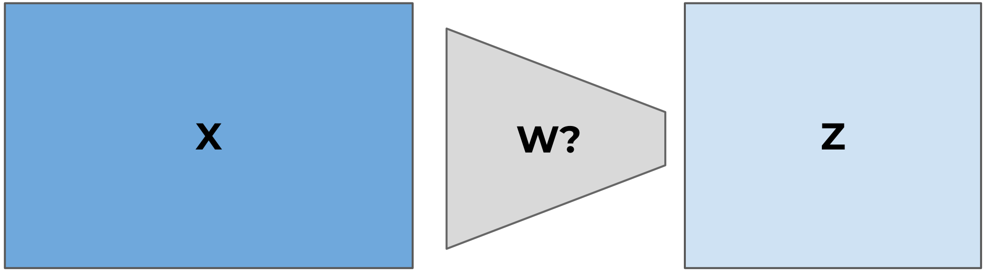


Orientation of the data

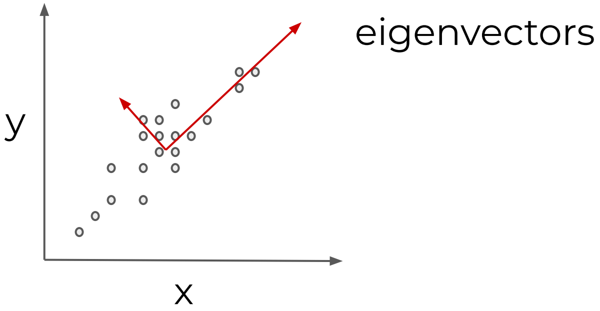
$$\begin{pmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{pmatrix}$$

Spread/variance of the data

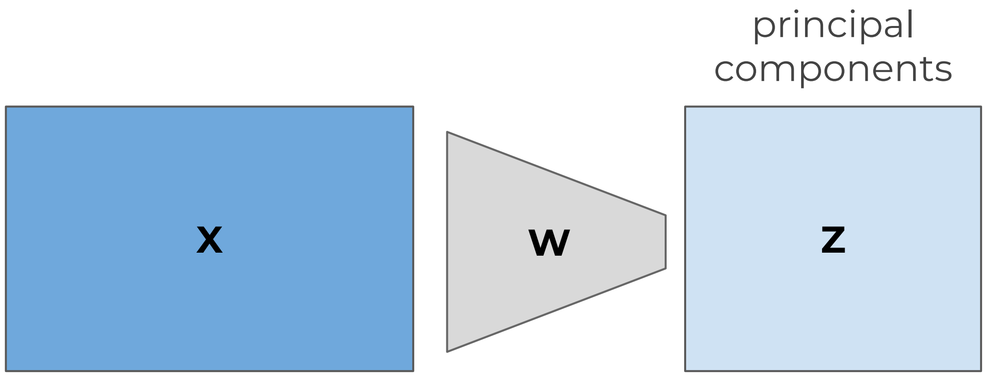
Principal Component Analysis (PCA)



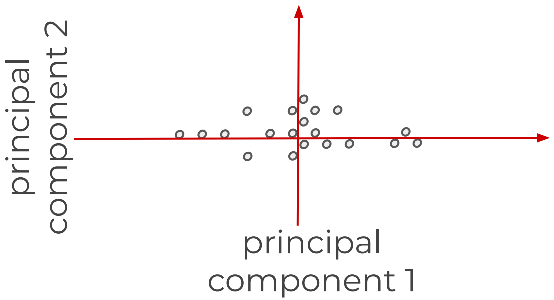
1. Calculate Covariance(**X**)
2. Factorize Covariance(**X**) = $\mathbf{VDV}^T$



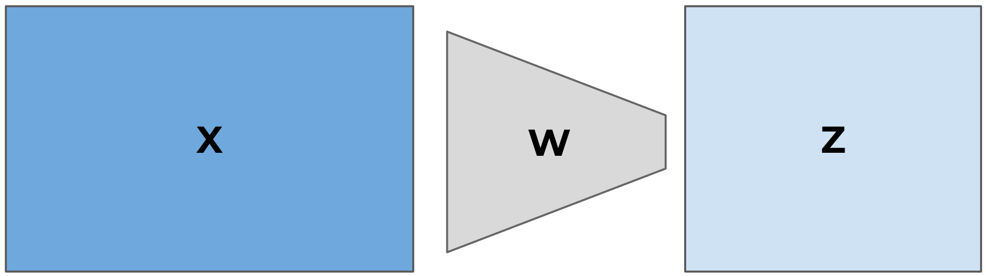
Principal Component Analysis (PCA)



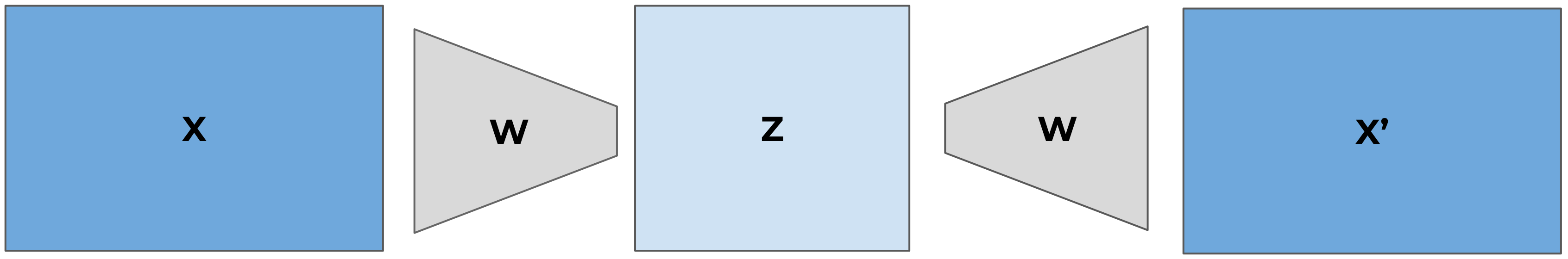
1. Calculate Covariance(**X**)
2. Factorize Covariance(**X**) = **VDV**^{T}
3. **W** contains principal components
4. **XW** = **Z**



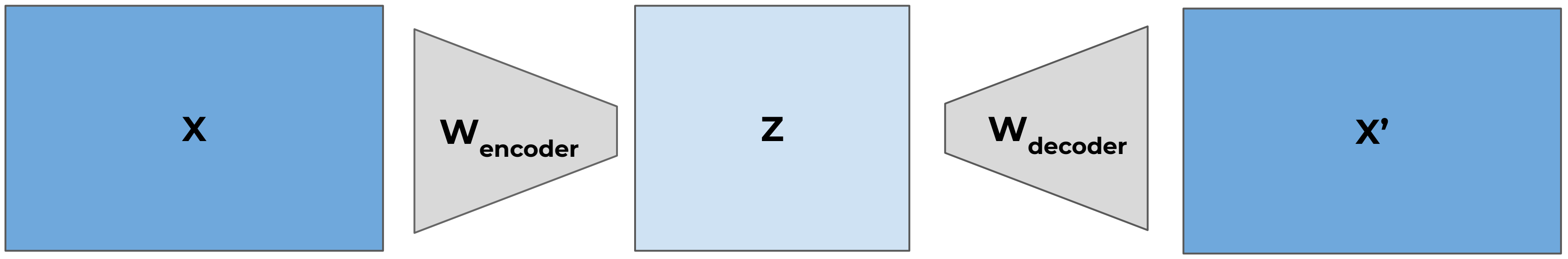
Autoencoder (AE)



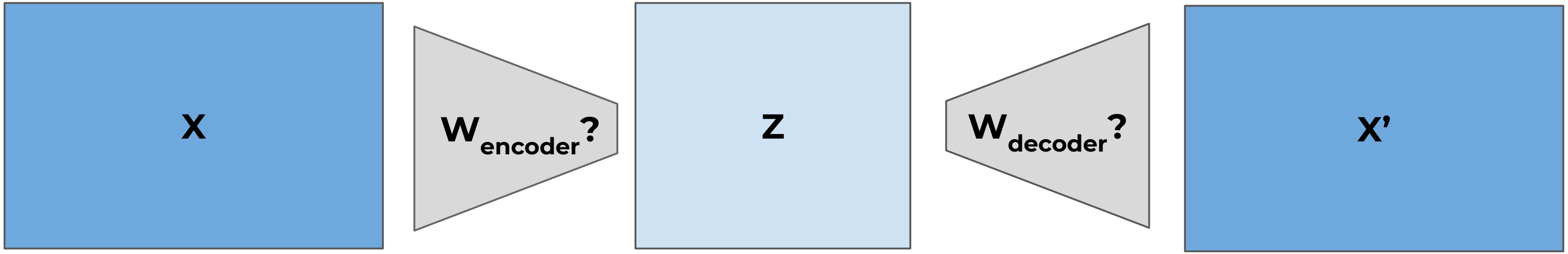
Autoencoder (AE)



Autoencoder (AE)

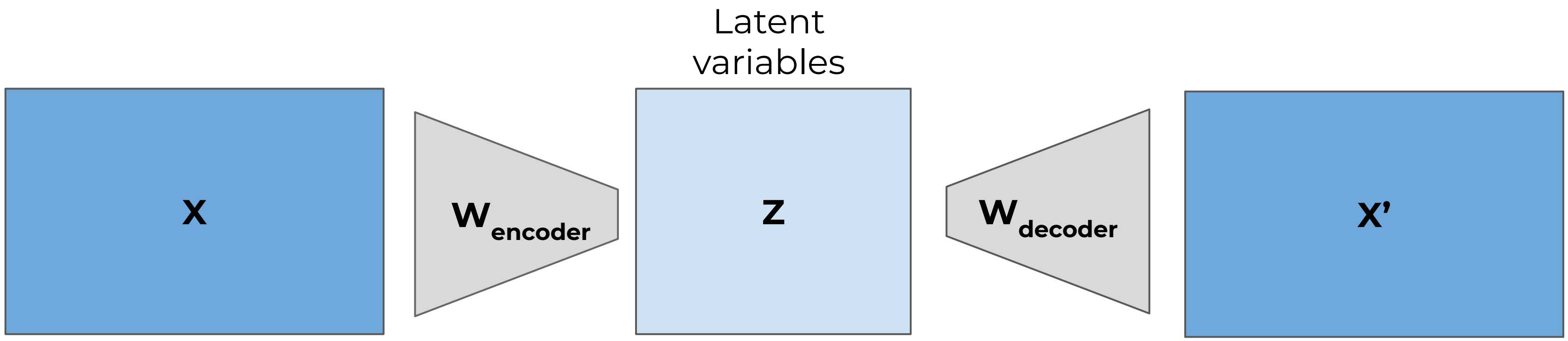


Autoencoder (AE)



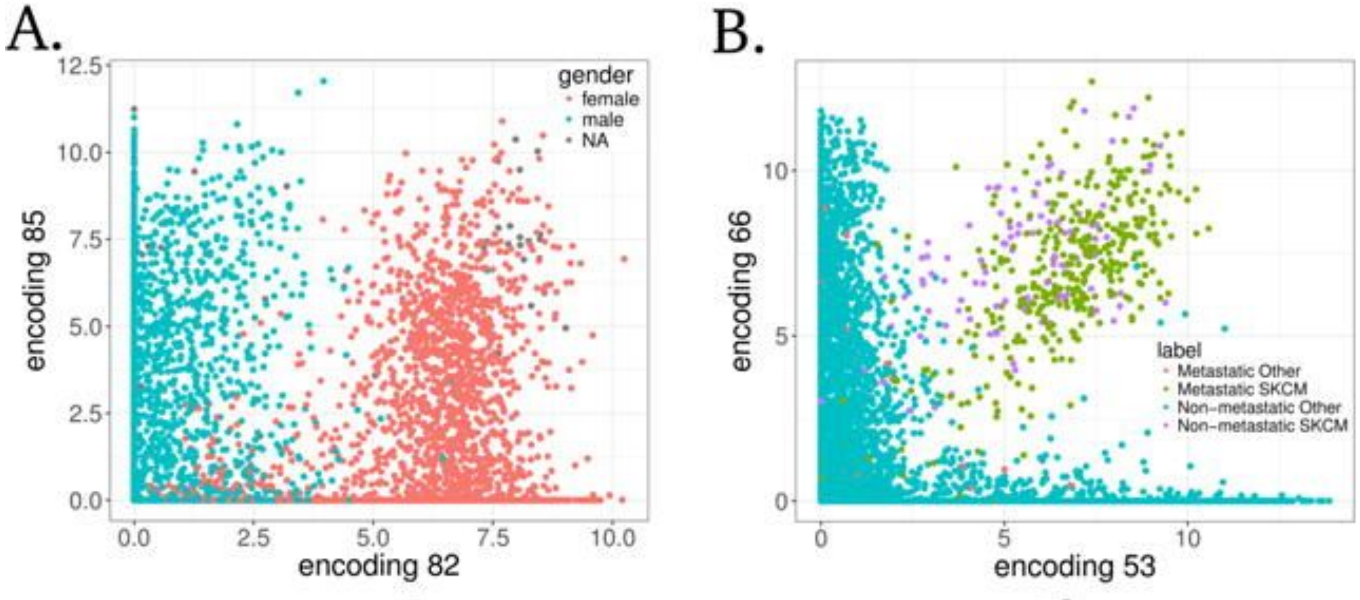
Goal: Find W that projects data on to a low dimensional space while preserving broad trends in the data

Autoencoder (AE)



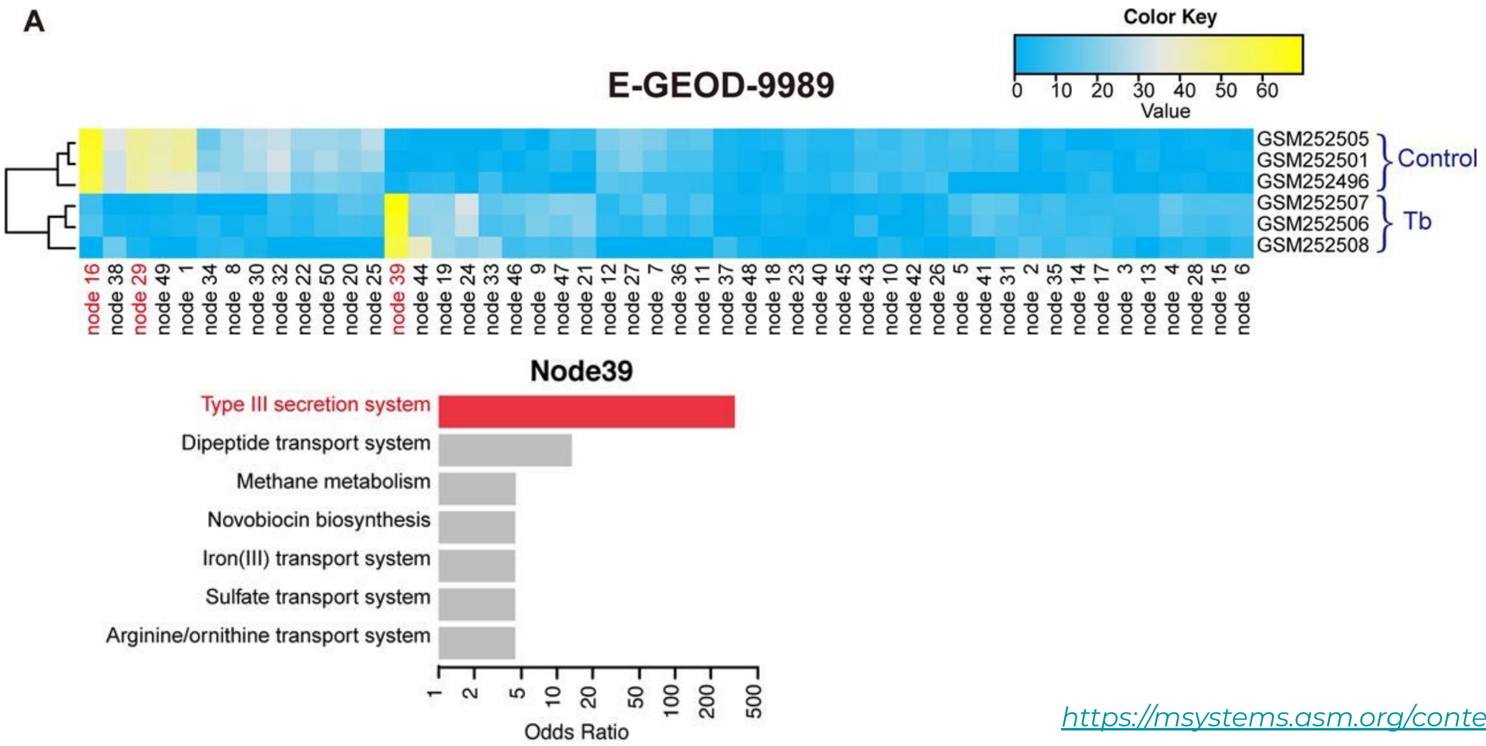
$\mathbf{W}_{\mathbf{encoder}}, \mathbf{W}_{\mathbf{decoder}} \leftarrow \min(\mathrm{error}(X, X'))$

Low dimensional features can differentiate between sex, SKCM tumors



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5728678/

Low dimensional features can represent biological processes



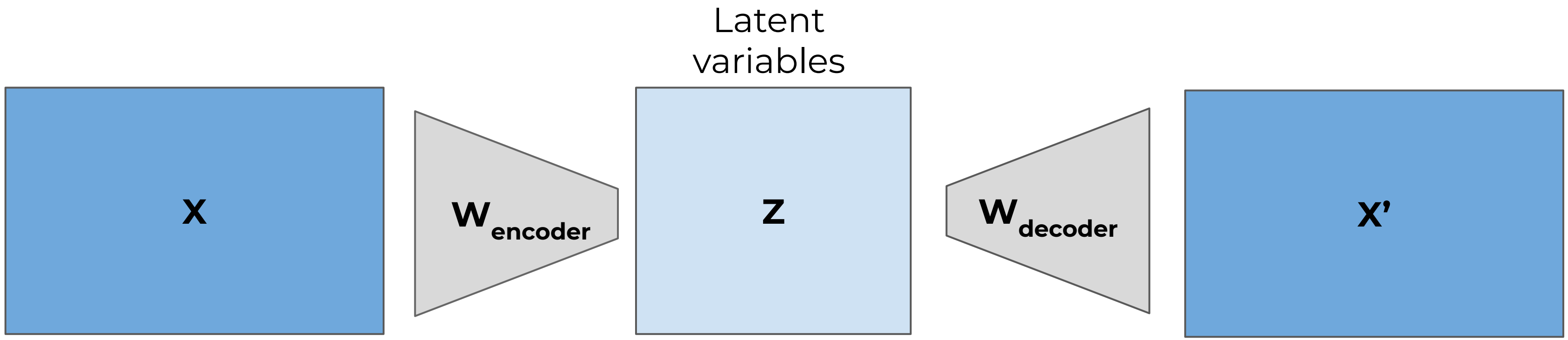
https://msystems.asm.org/content/1/1/e00025-15

Other methods you can google

- Principal Component Analysis (PCA)
- Nonnegative Matrix Factorization (NMF)
- Linear Decomposition Analysis (LDA)
- Autoencoders (AE): Denoising (DAE), Variational (VAE)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Uniform Manifold Approximation and Projection (UMAP)

Other flavors of autoencoders

Variational Autoencoder (VAE)



$\mathbf{W}_{\mathbf{encoder}}, \mathbf{W}_{\mathbf{decoder}} \leftarrow \min(\text{error}(X, X')) + \mathbf{Z} \sim \textit{Normal}(0,1)$

Takeaway

- Dimensionality reduction methods can help to denoise and find structure in your data
- Which method you use will depend on your problem
- Caveats:
    - Throwing away some information, rare signals

Conclusion

- Depending on your problem, you might consider some of these deep learning methods
    - Tips for using deep learning in biology: https://github.com/ajlee21/deep-rules

Conclusion

- Depending on your problem, you might consider some of these deep learning methods
  - Tips for using deep learning in biology: https://github.com/ajlee21/deep-rules
- Don't be afraid to just play around with some data
  - https://github.com/ben-heil/dl_workshop/tree/main/notebooks

Thank you

Greene Lab

Where data and biology meet


THANKS YOU FOR LISTENING
TO THIS PRESENTATION