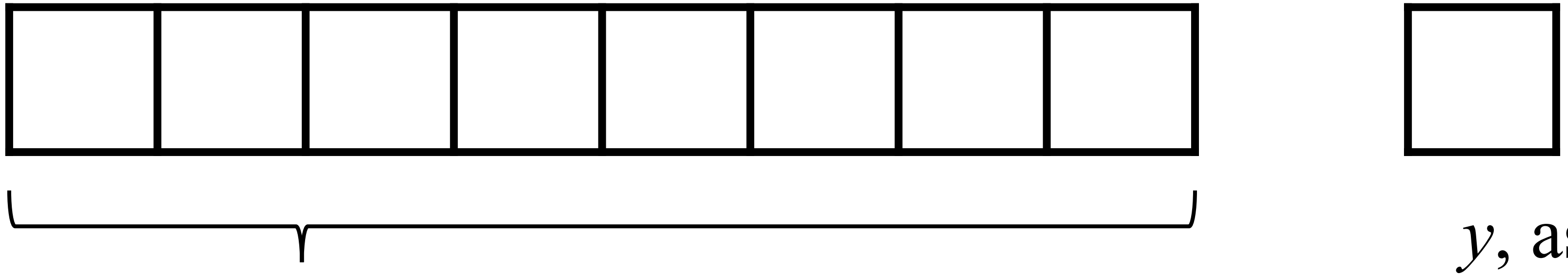
Beyond Supervised Learning

Matthew Engelhard

Supervised Learning

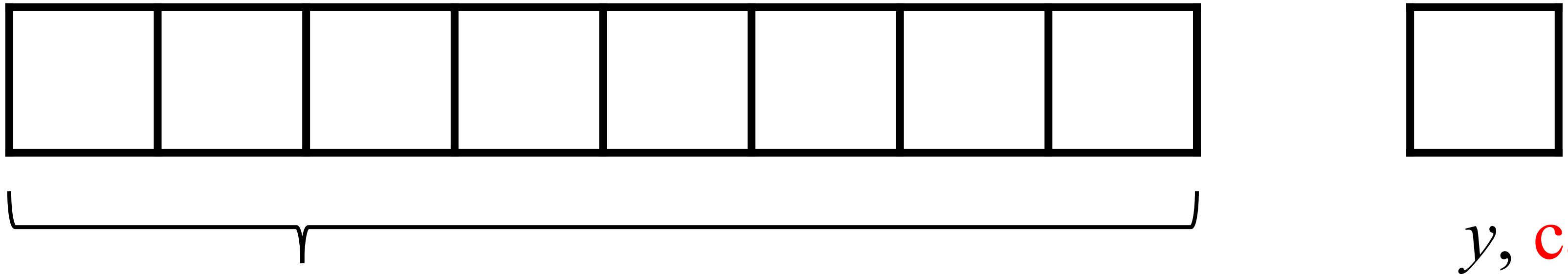


x, data/features for a subject or patient

y, associated value or label

The learning process: find the equation that best predicts y based on x

Supervised Learning: Classification



x, data/features for
a subject or patient

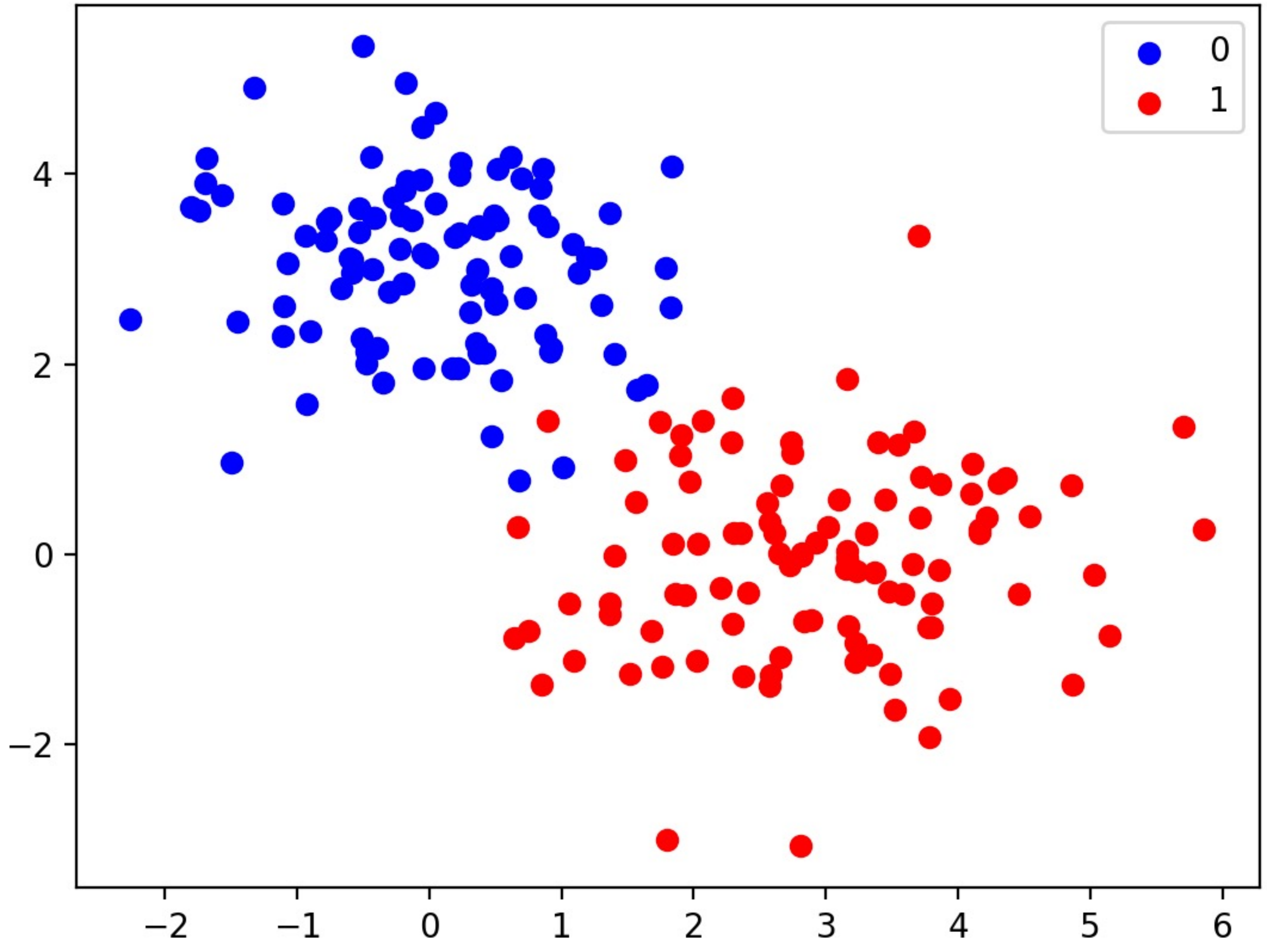
y, categorical
label

The learning process: find the equation that best predicts y based on x

Supervised Learning: Classification

<u>Goal</u>:

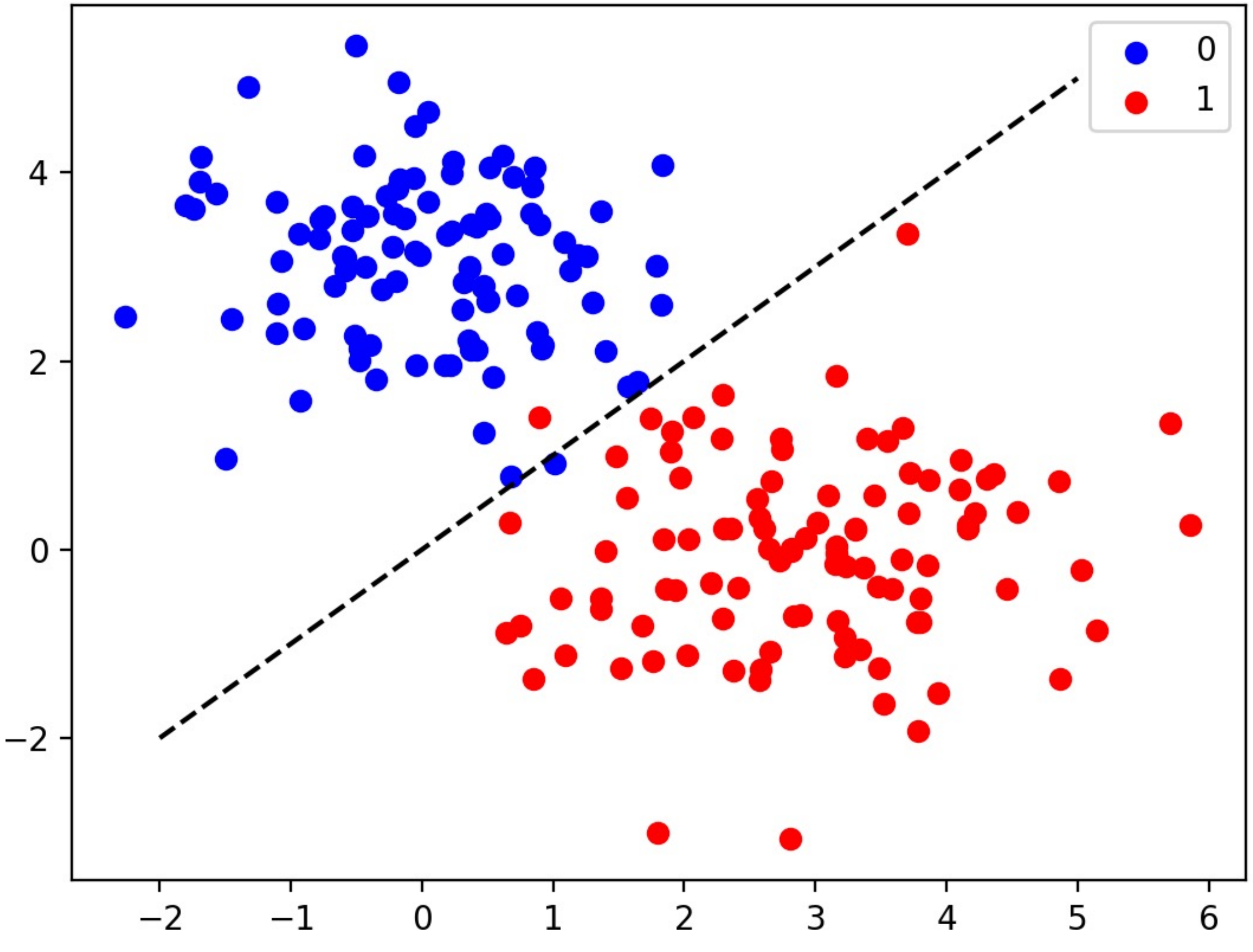
Learn a **decision boundary** that separates 0s from 1s



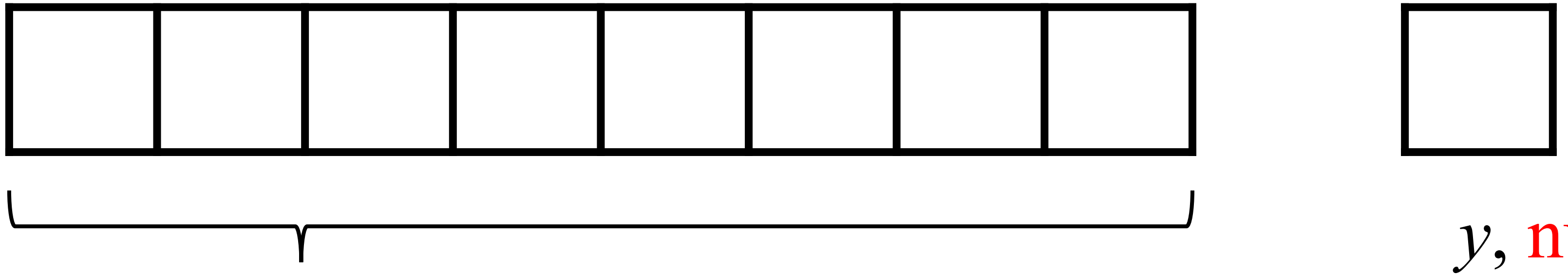
Supervised Learning: Classification

<u>Goal</u>:

Learn a **decision boundary** that separates 0s from 1s



Supervised Learning: Regression



x, data/features for
a subject or patient

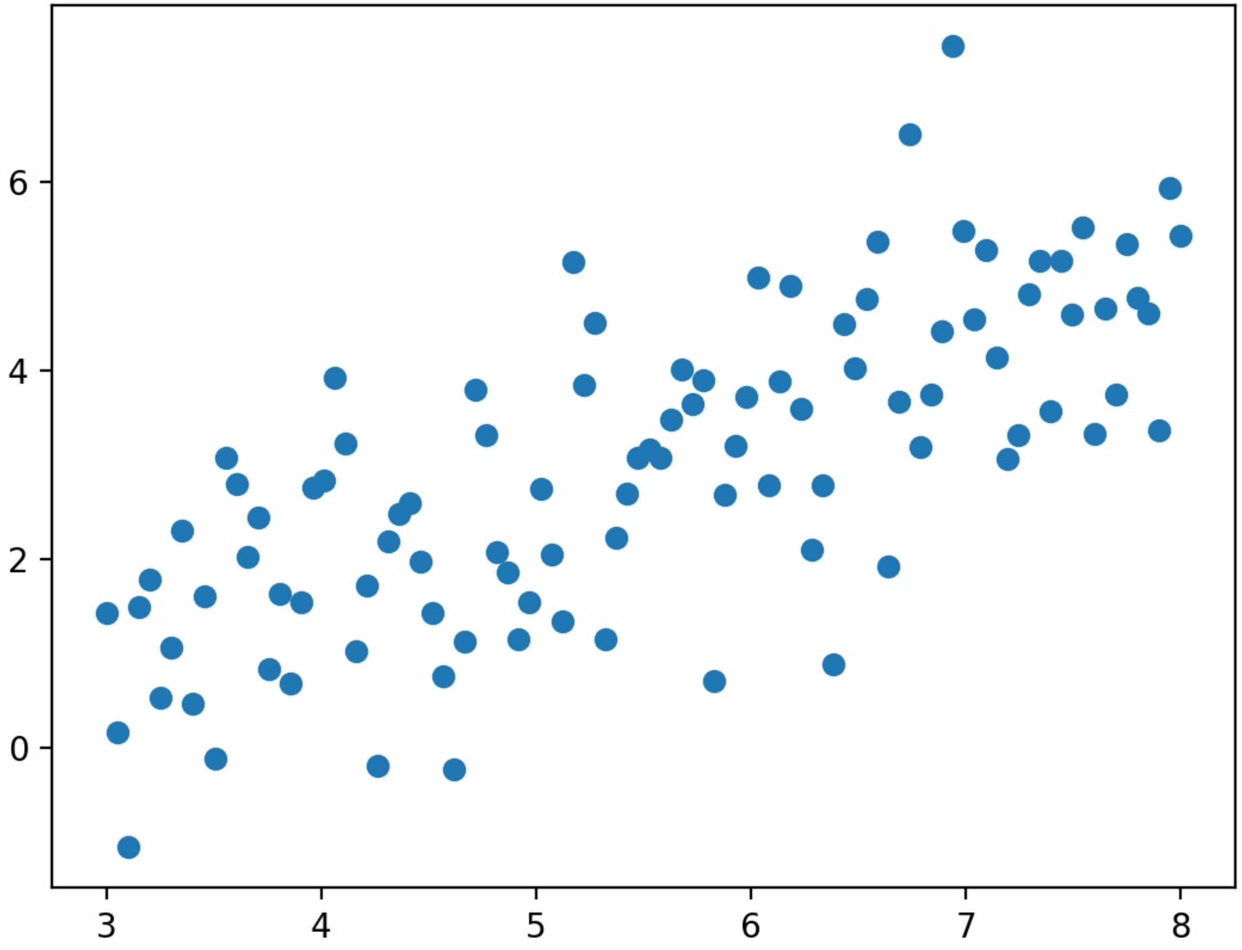
y, numeric
value

The learning process: find the equation that best predicts y based on x

Supervised Learning: Regression

<u>Goal</u>:

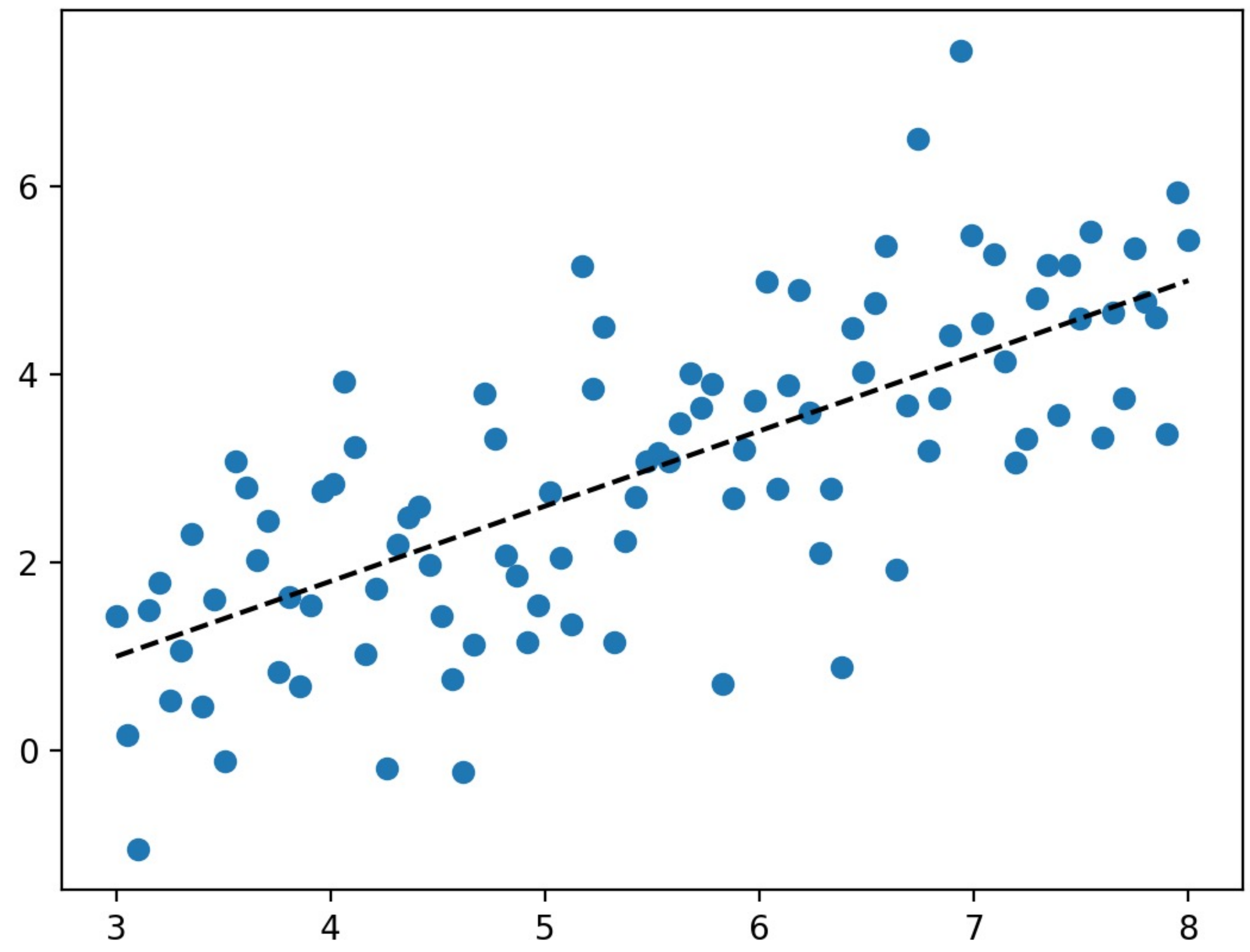
Learn a function that predicts y based on x



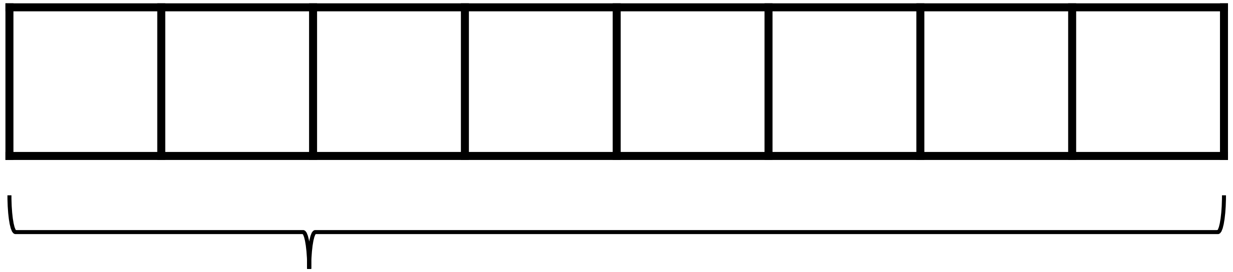
Supervised Learning: Regression

<u>Goal</u>:

Learn a function that predicts y based on x



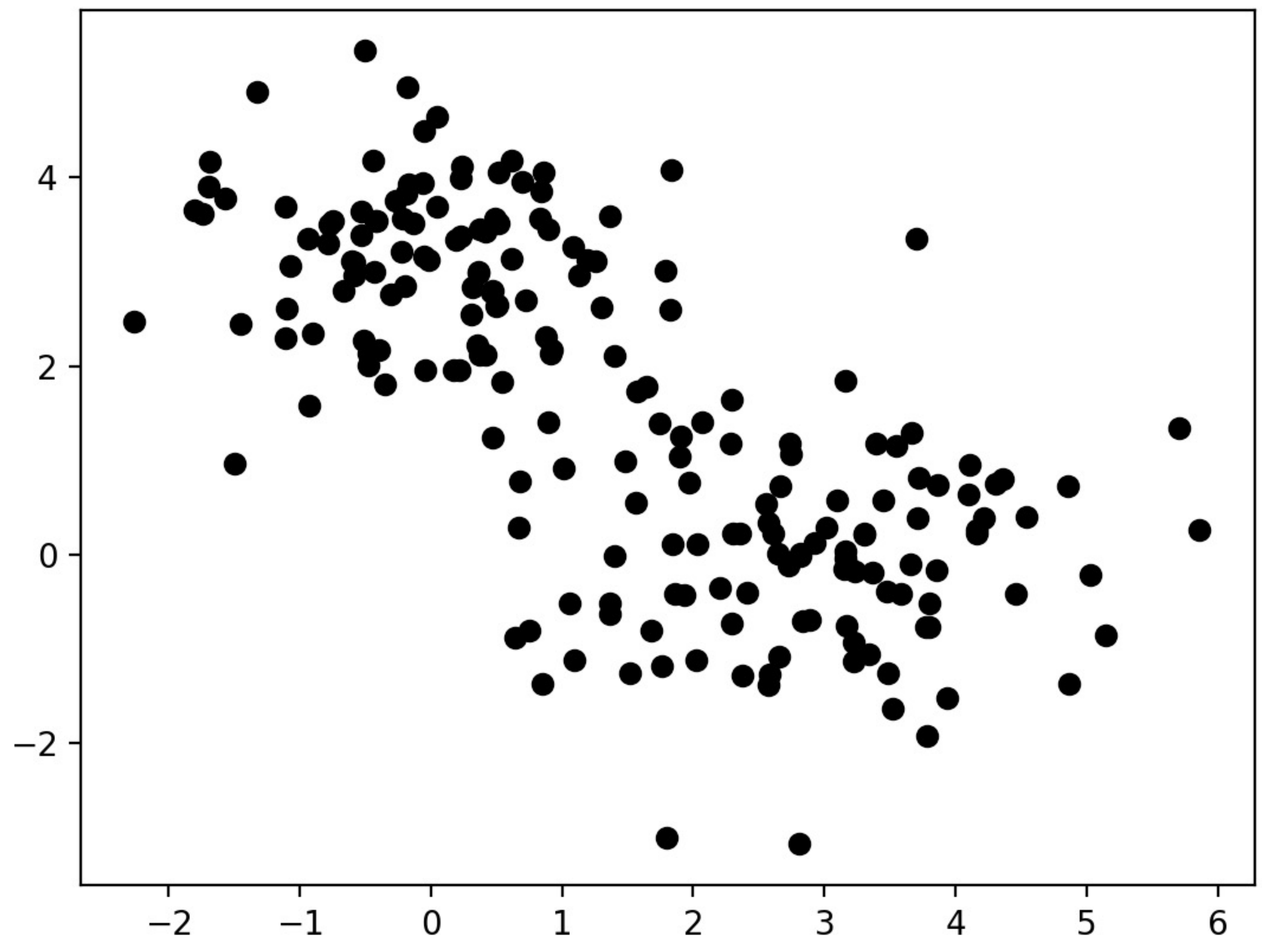
Unsupervised Learning



x, data/features for
a subject or patient

The learning process: - find structure or patterns in the data
- describe the data or create new, similar data

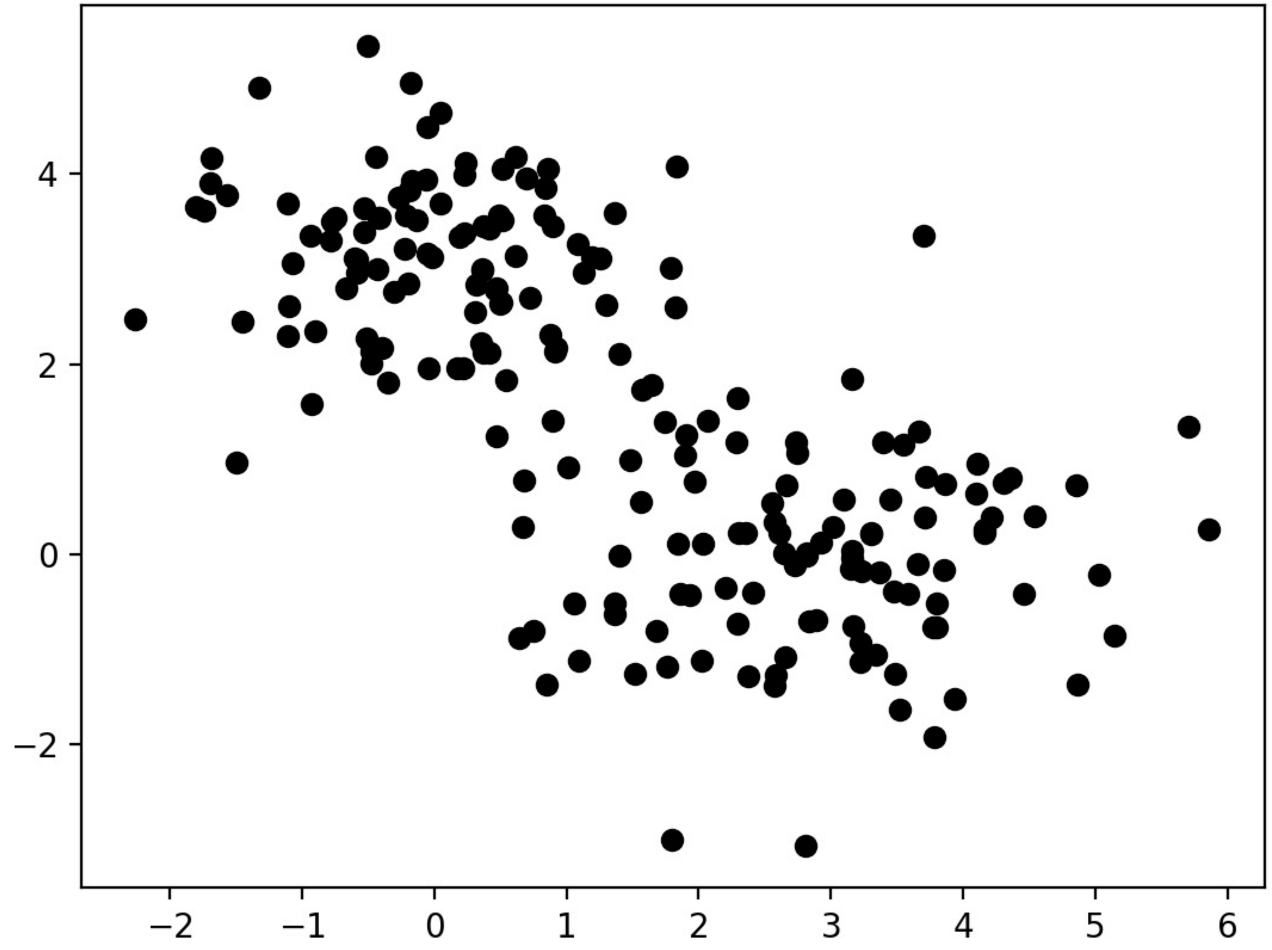
Unsupervised Learning



Unsupervised Learning: Clustering

<u>Goal</u>:

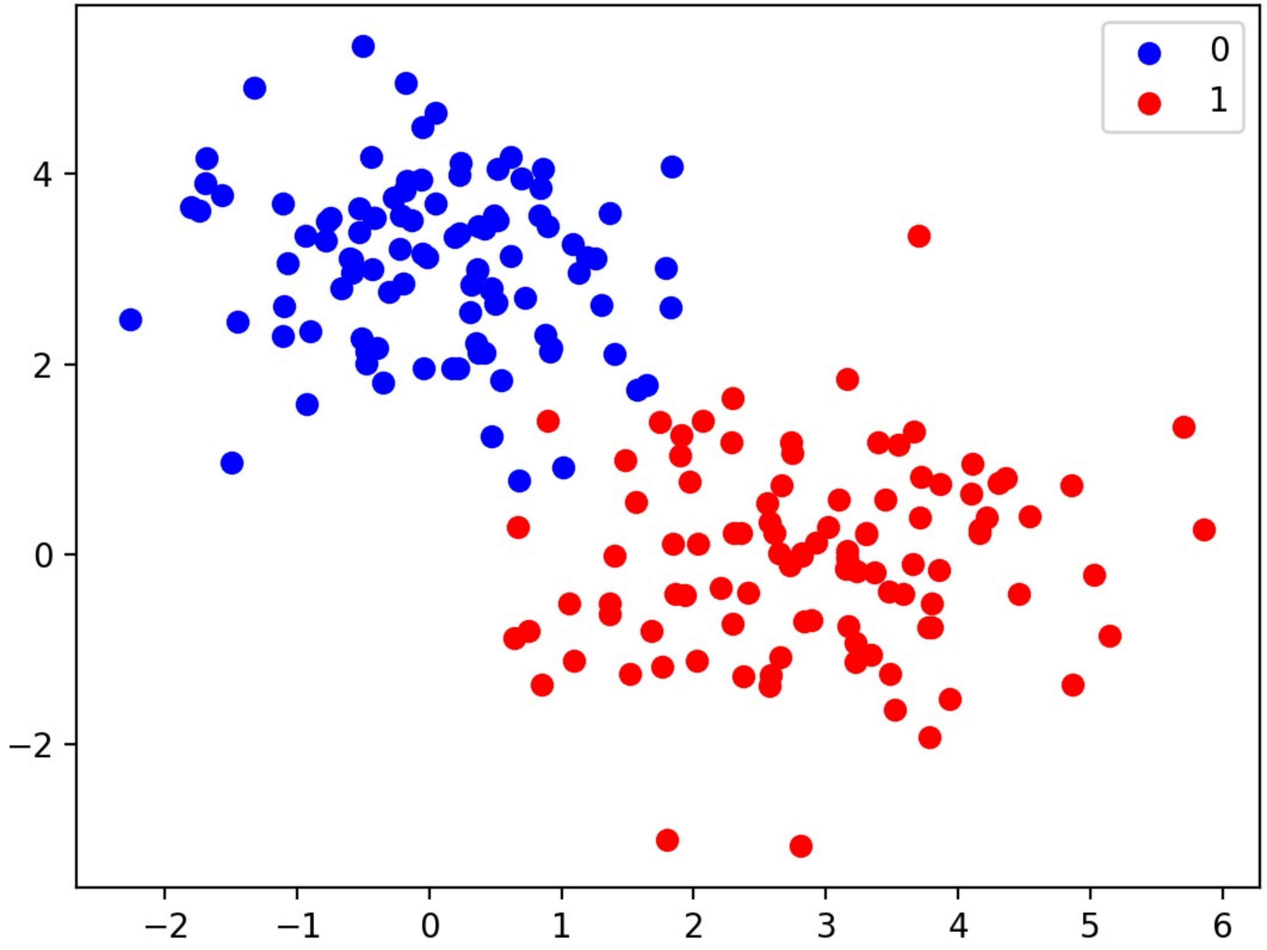
Assign points to distinct groups with shared characteristics



Unsupervised Learning: Clustering

<u>Goal</u>:

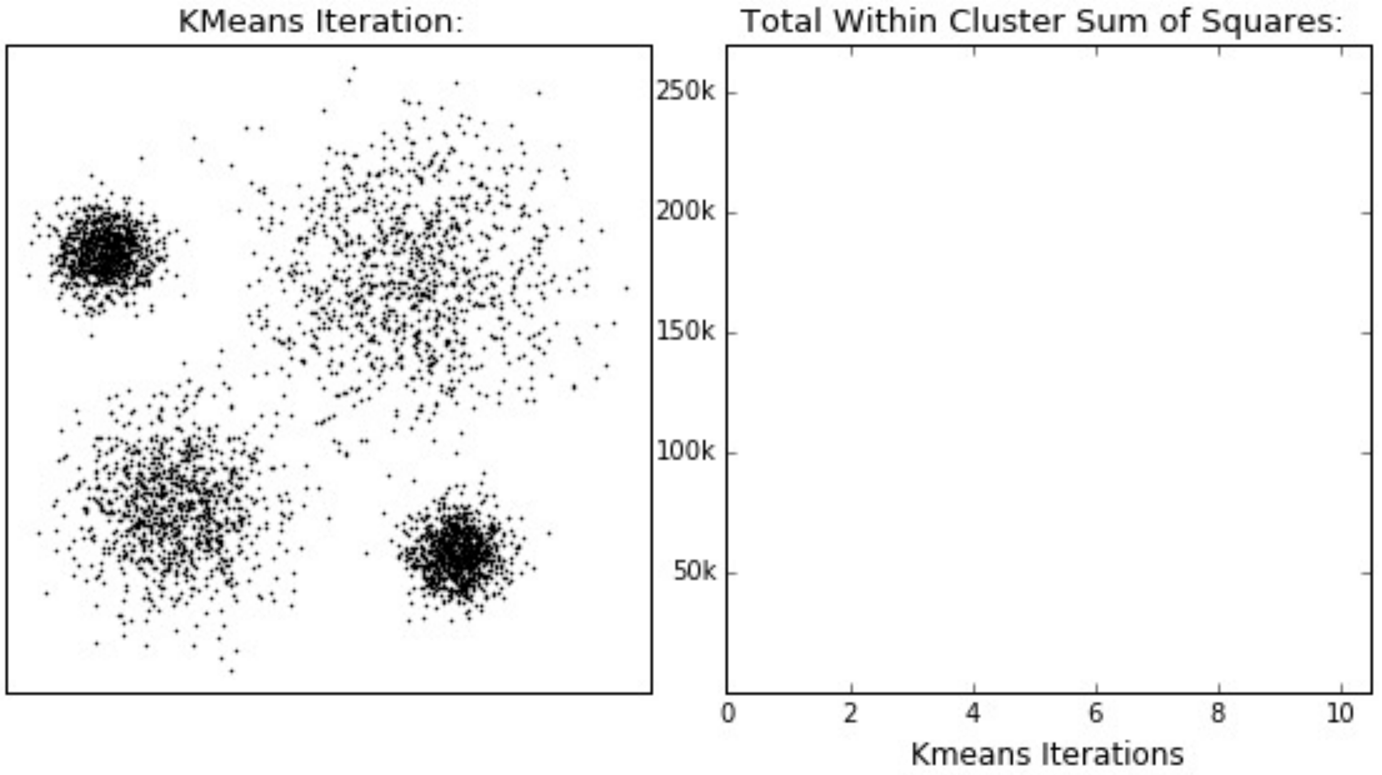
Assign points to distinct groups with shared characteristics



Example of K-Means Clustering

<u>Goal</u>:

Assign points to distinct groups with shared characteristics

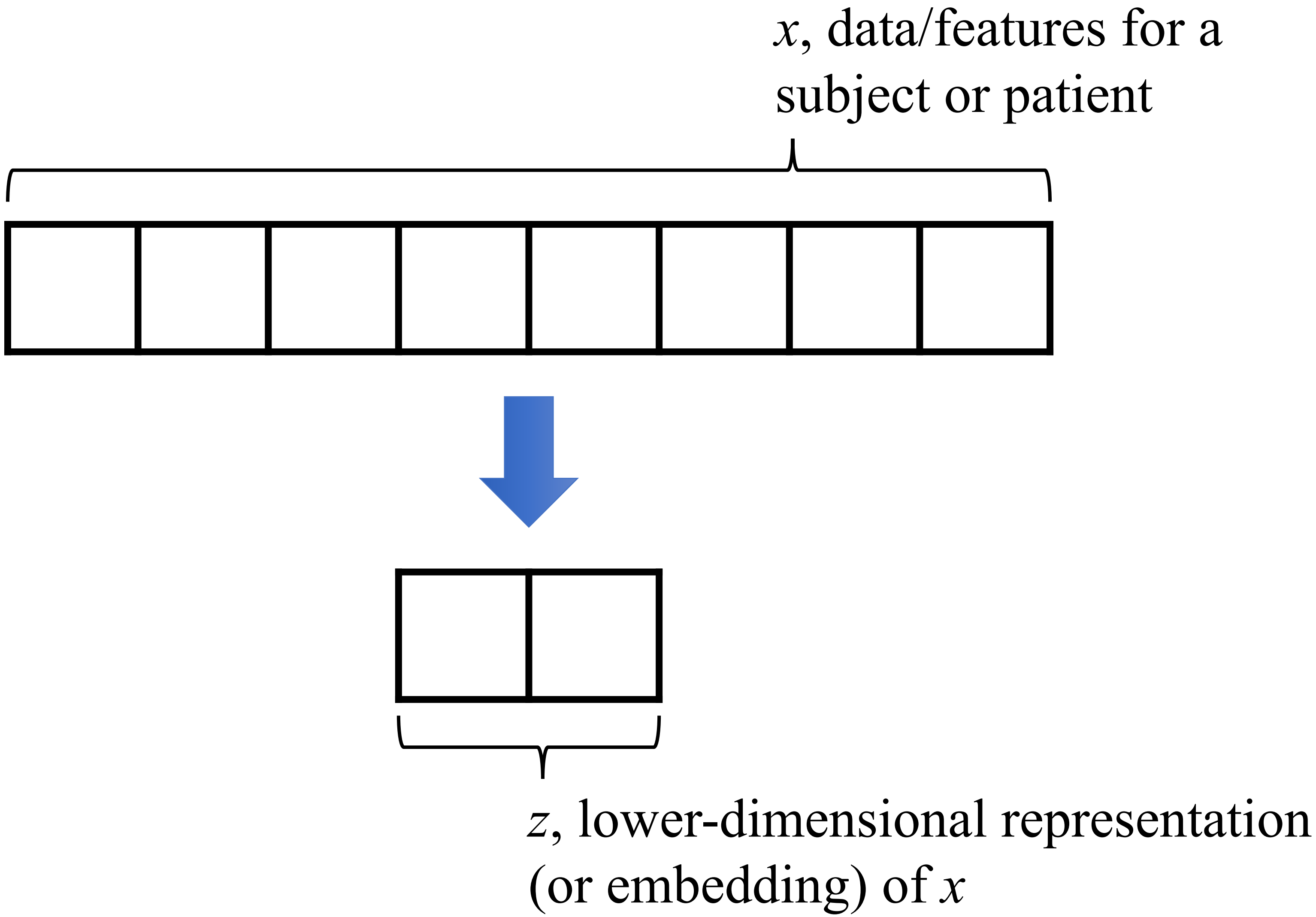


- sensitive to initialization
- minimize $\boldsymbol{\Sigma}$distance^2 from points to centers

Unsupervised Learning: Dimensionality Reduction

<u>Goal</u>:

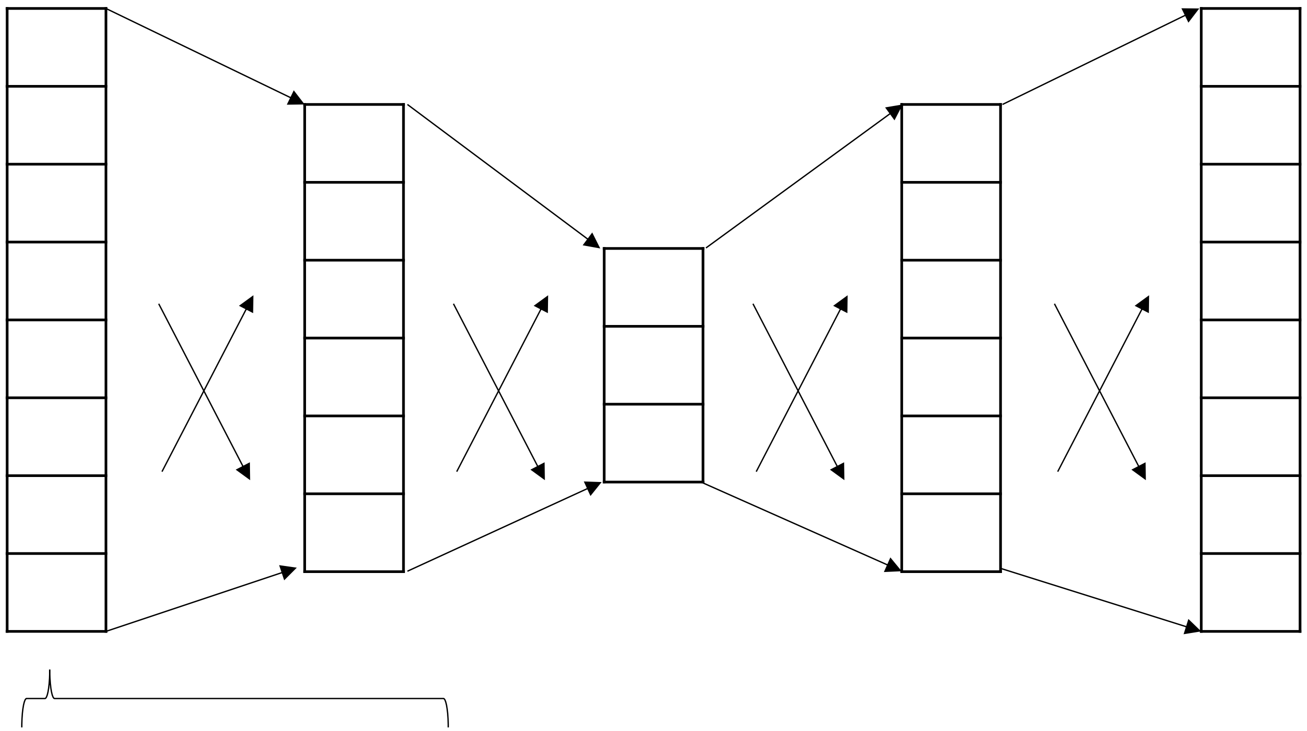
Describe a large number of features in terms of a smaller number of features



Dimensionality Reduction Example: Autoencoder

<u>Goal</u>:

- Describe a large number of features in terms of a smaller number of features
- Train to minimize **reconstruction loss**



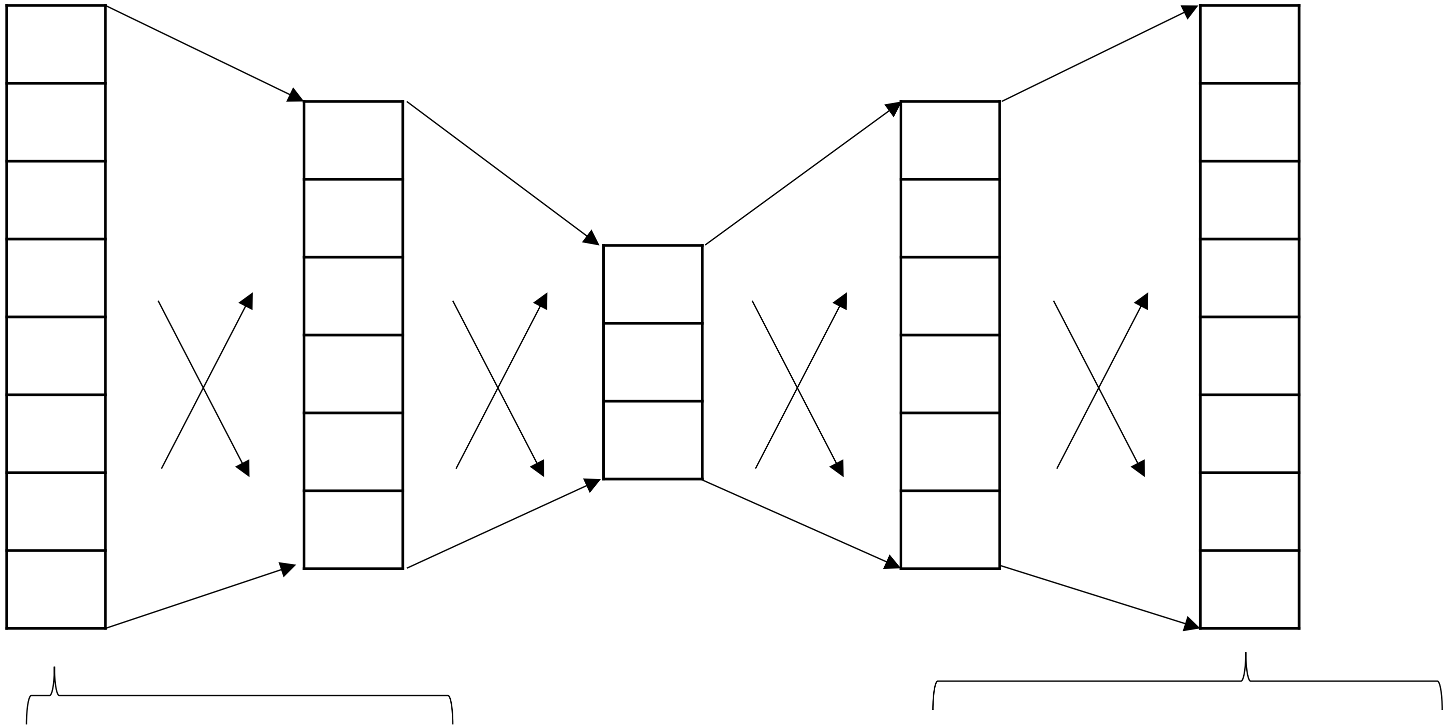
x, data/features for a subject or patient

Dimensionality Reduction Example: Autoencoder

<u>Goal</u>:

- Describe a large number of features in terms of a smaller number of features

- Train to minimize **reconstruction loss**



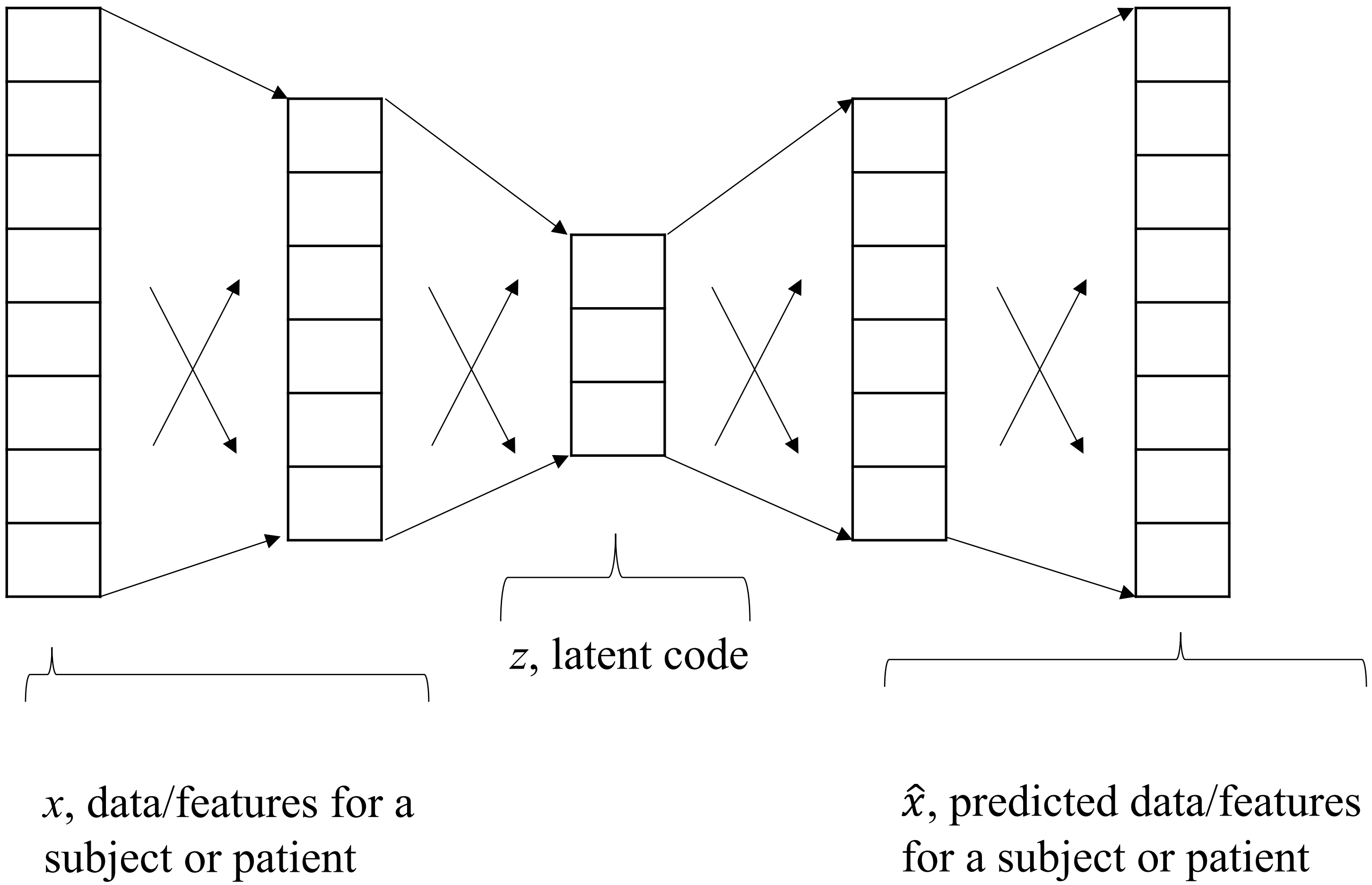
x, data/features for a subject or patient

$\hat{x}$, predicted data/features for a subject or patient

Dimensionality Reduction Example: Autoencoder

<u>Goal</u>:

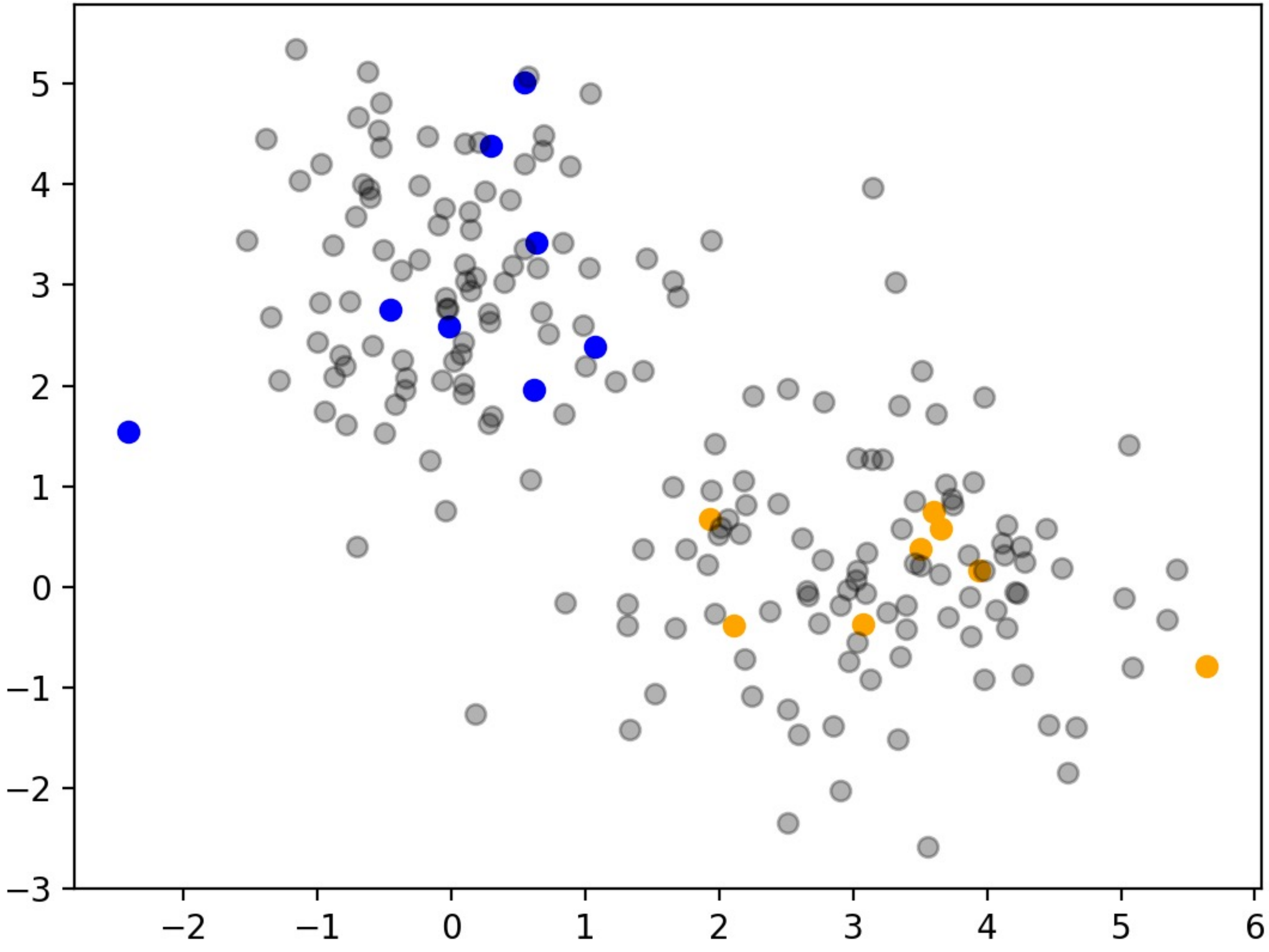
- Describe a large number of features in terms of a smaller number of features
- Train to minimize **reconstruction loss**



Semi-Supervised Learning

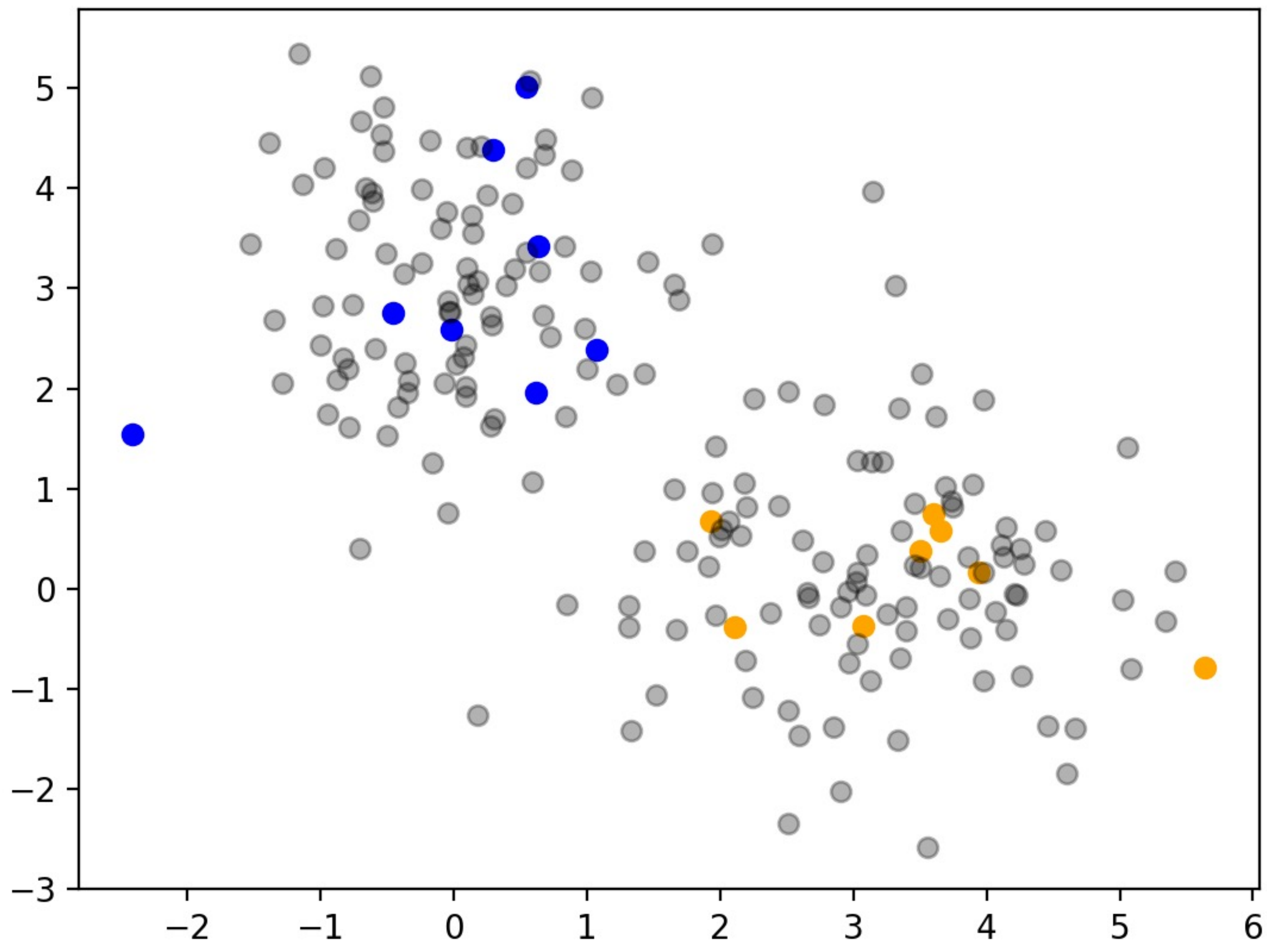
- Some points are labeled, some are not

- Try to use the unlabeled data to help us reason about the labeled data



Active Learning

- Some points are labeled, some are not
- We can get additional labels, but at a cost
- Request labels we believe will improve our classifier the most

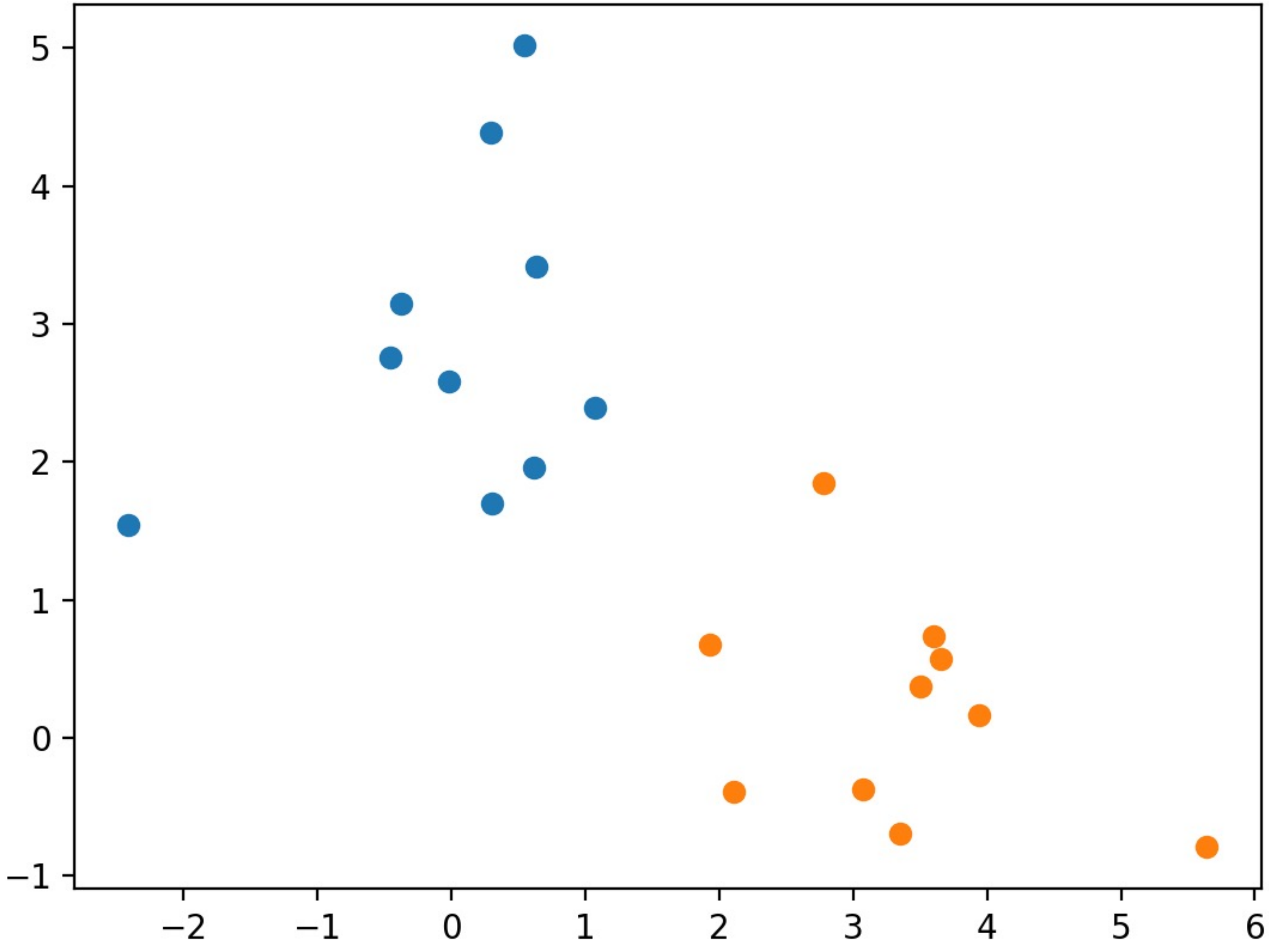


Online Learning

- Data arrives one point at a time, or in batches

- Continually improve our classifier without having to retrain from scratch with each arrival

- Uses a learning rate, much like RL

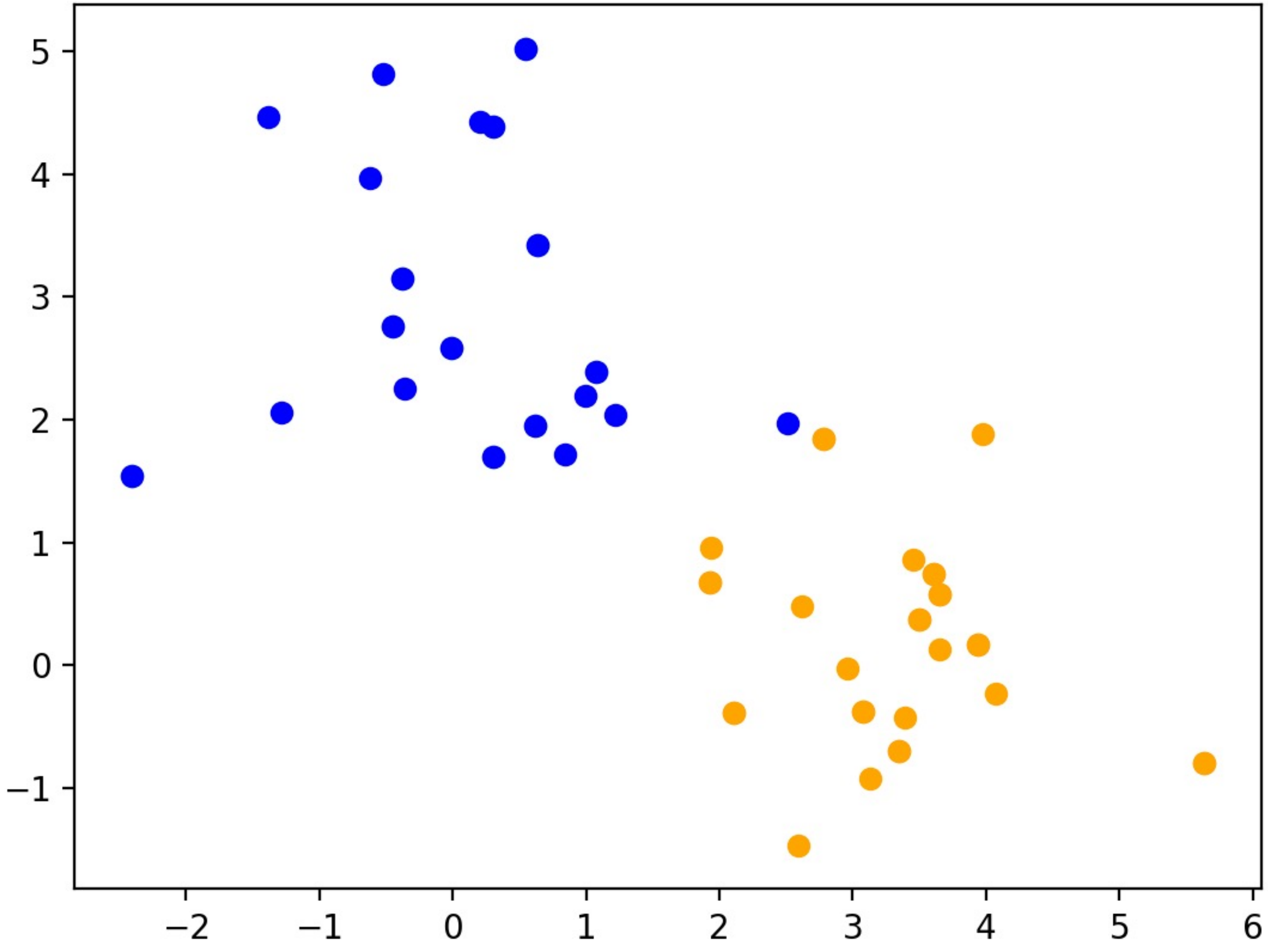


Online Learning

- Data arrives one point at a time, or in batches

- Continually improve our classifier without having to retrain from scratch with each arrival

- Uses a learning rate, much like RL

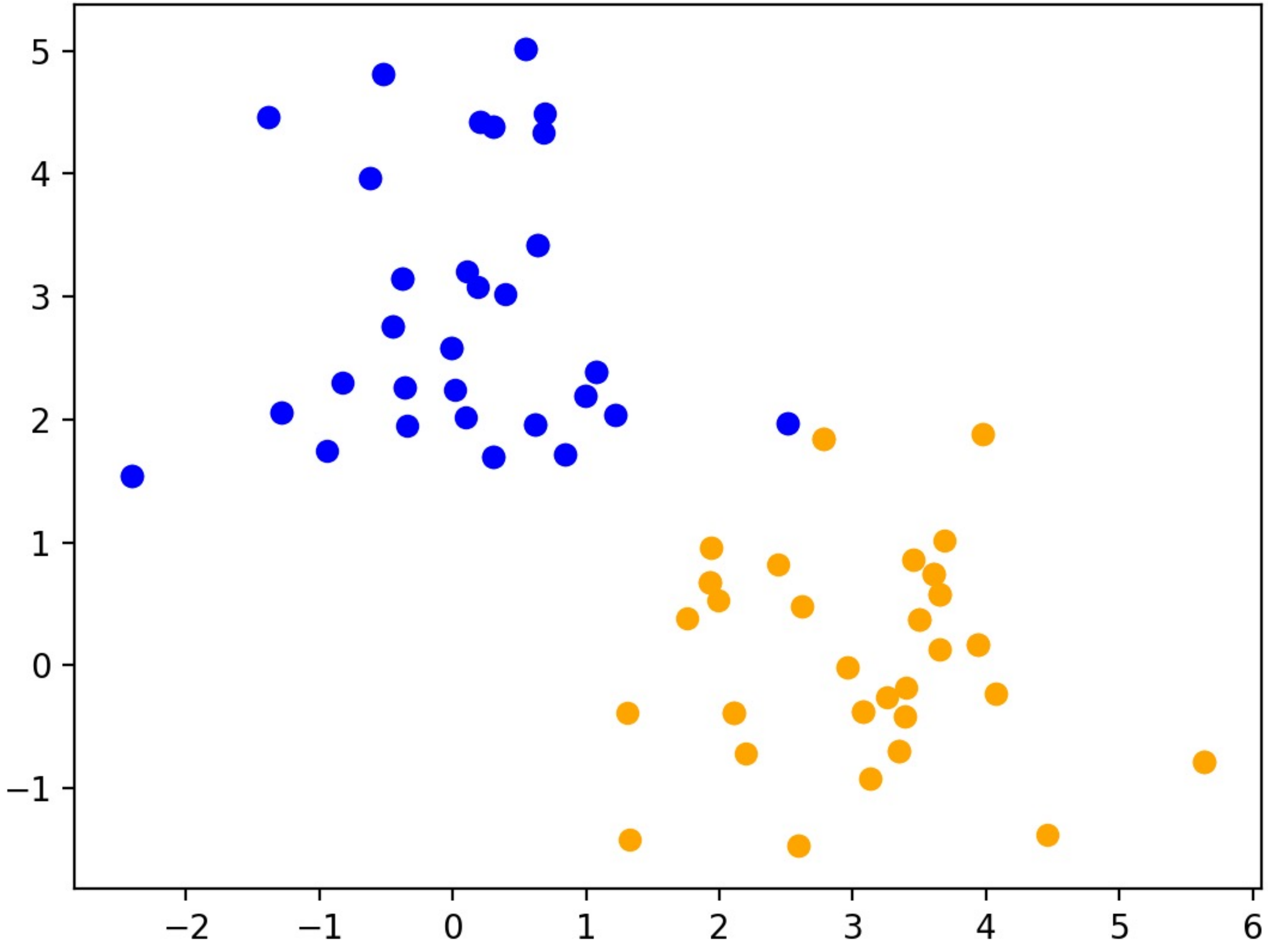


Online Learning

- Data arrives one point at a time, or in batches

- Continually improve our classifier without having to retrain from scratch with each arrival

- Uses a learning rate, much like RL

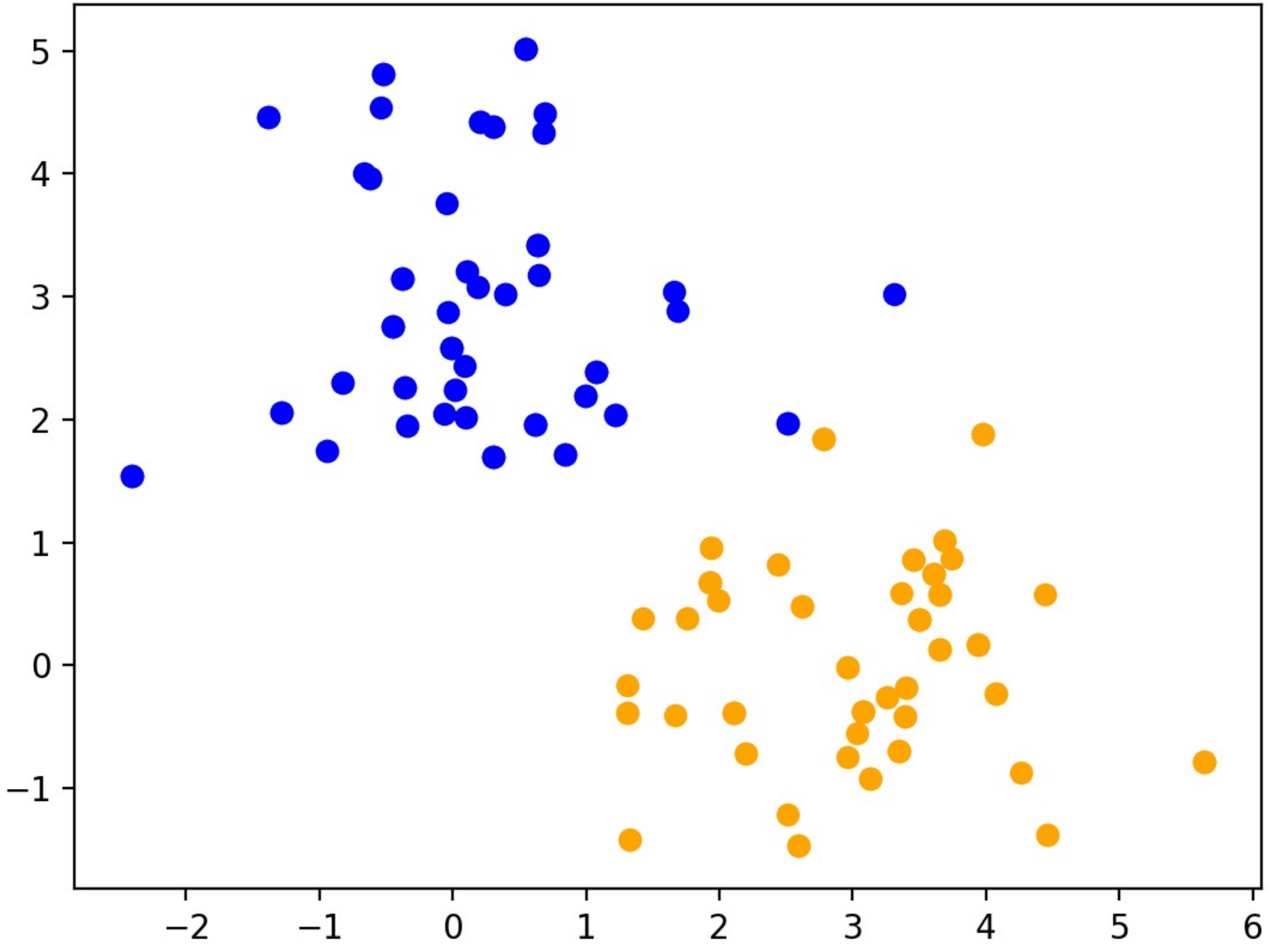


Online Learning

- Data arrives one point at a time, or in batches

- Continually improve our classifier without having to retrain from scratch with each arrival

- Uses a learning rate, much like RL

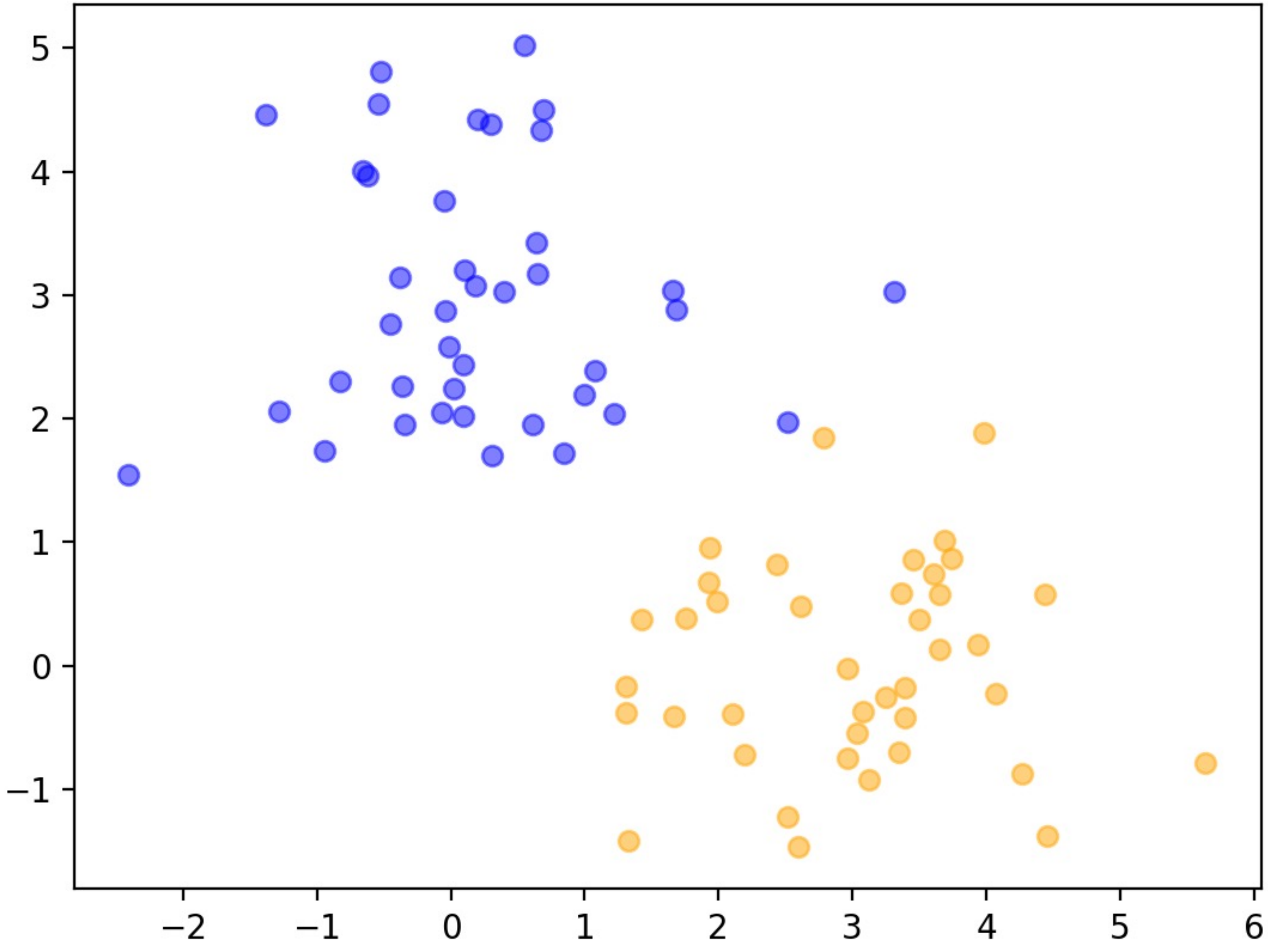


Lifelong Learning

- Similar to lifelong learning; also uses a learning rate

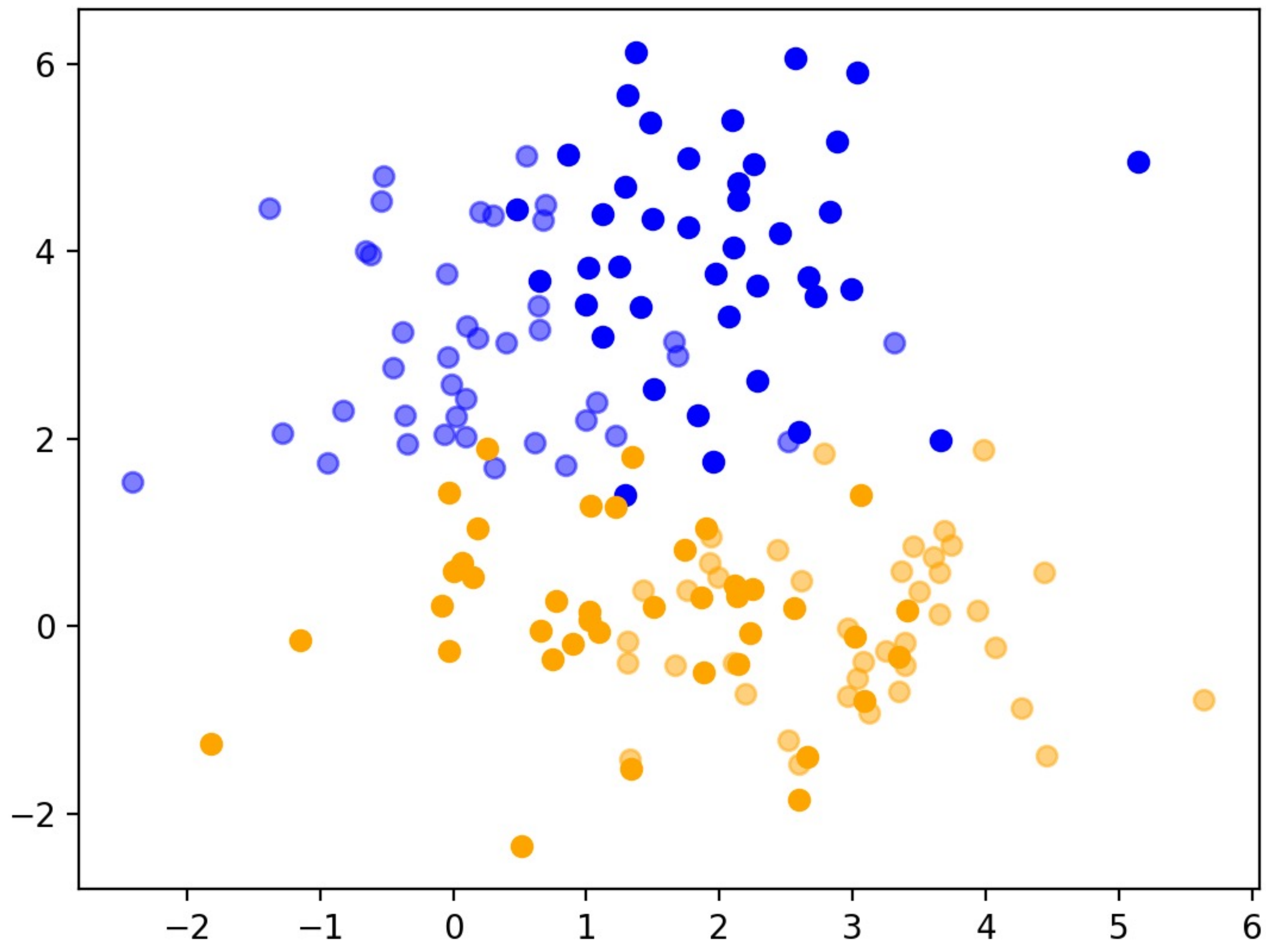
- Data characteristics change over time

- Continually refine our classifier to adjust to these changing characteristics



Lifelong Learning

- Similar to lifelong learning; also uses a learning rate
- Data characteristics change over time
- Continually refine our classifier to adjust to these changing characteristics



Summary:

- <u>Supervised Learning</u>: learn to predict labels from features
- <u>Unsupervised Learning</u>: make sense of the features

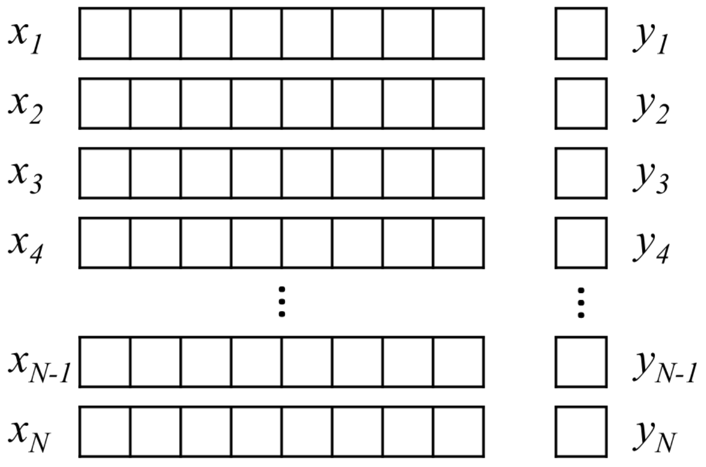
- <u>Semi-Supervised Learning</u>: learn to predict labels from features, but you have a limited number of labels
- <u>Active Learning</u>: learn to predict labels from features by selectively requesting or buying labels

- <u>Lifelong Learning</u>: continually re-learn how to predict labels from features

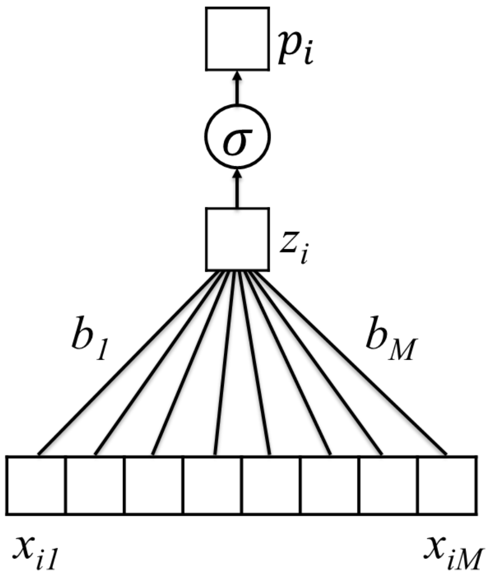
We have covered a lot!

(a quick victory lap)

Learning Model Parameters

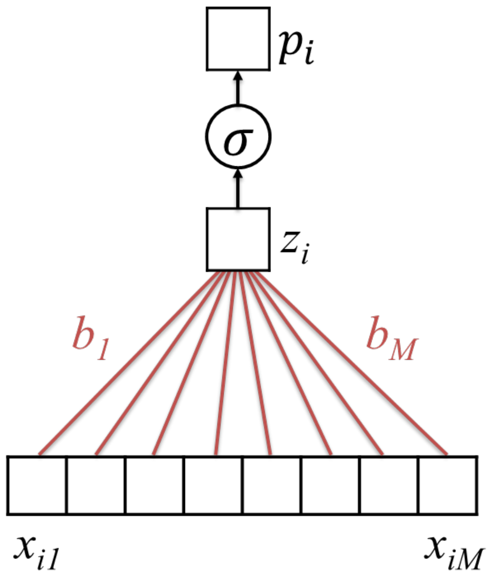


Training Set



$$p_i = \sigma(b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_M x_{iM})$$

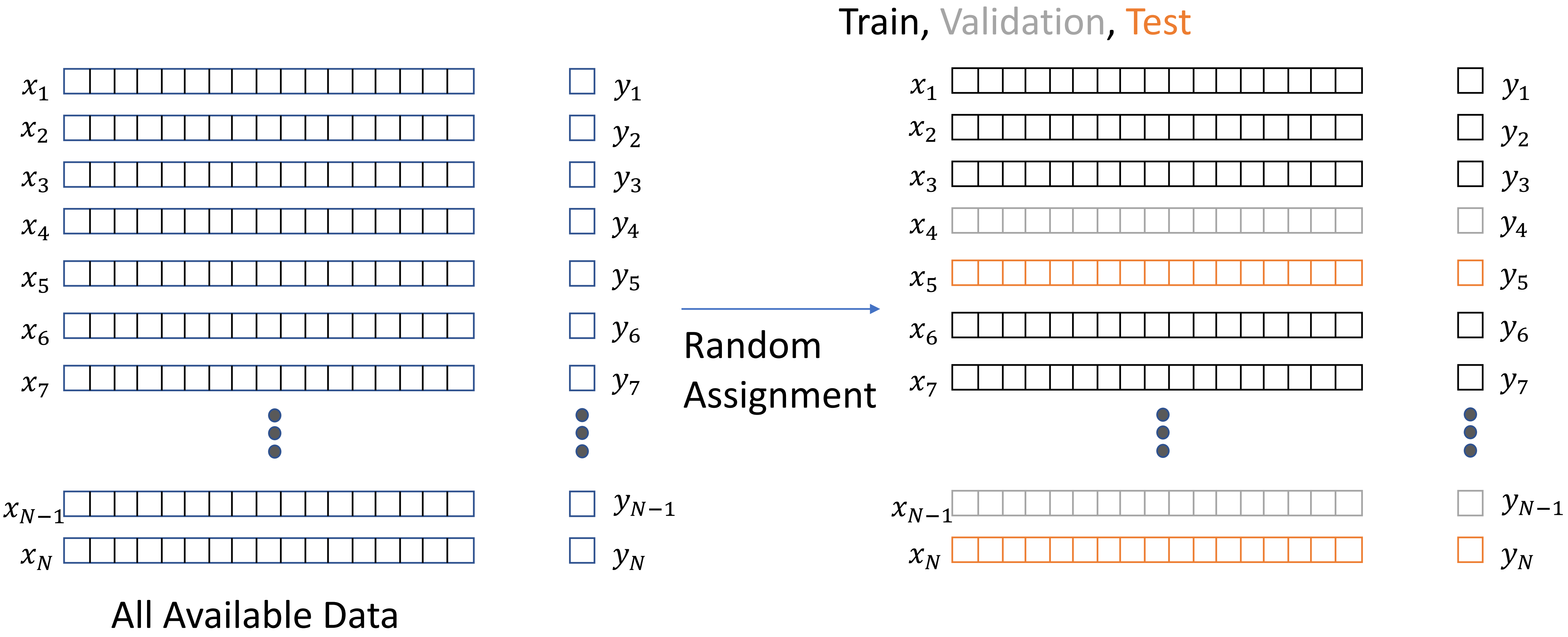
Untrained Logistic Regression Model (or "Network")



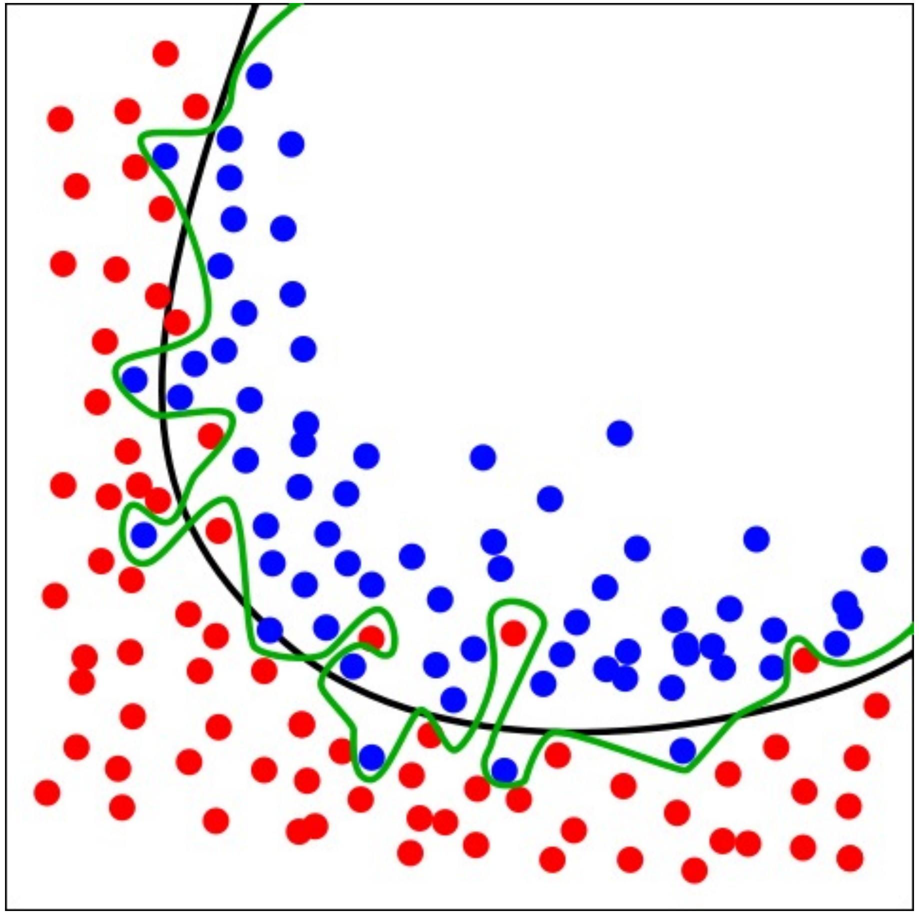
$$b = (b_0, \ldots b_M)$$

Trained Model (with learned parameters)

Split Data into Separate Groups



But some models can be *too* flexible.



By Chabacano - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=3610704

Green boundary:

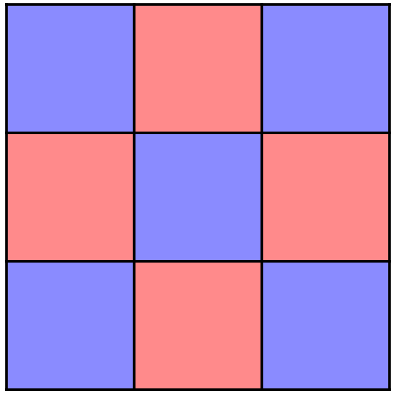
- This is overfitting

Black boundary:

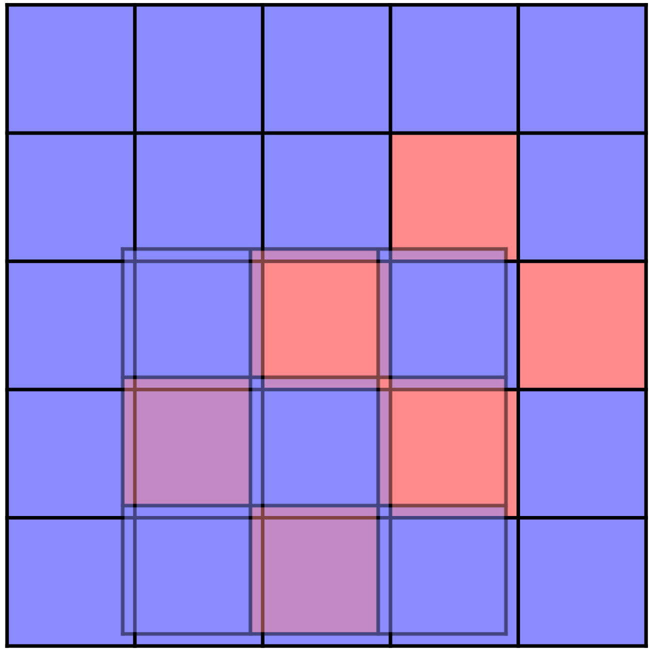
- Balance between fit and model complexity

-> The black boundary is likely to perform better on new data

An Example…



filter

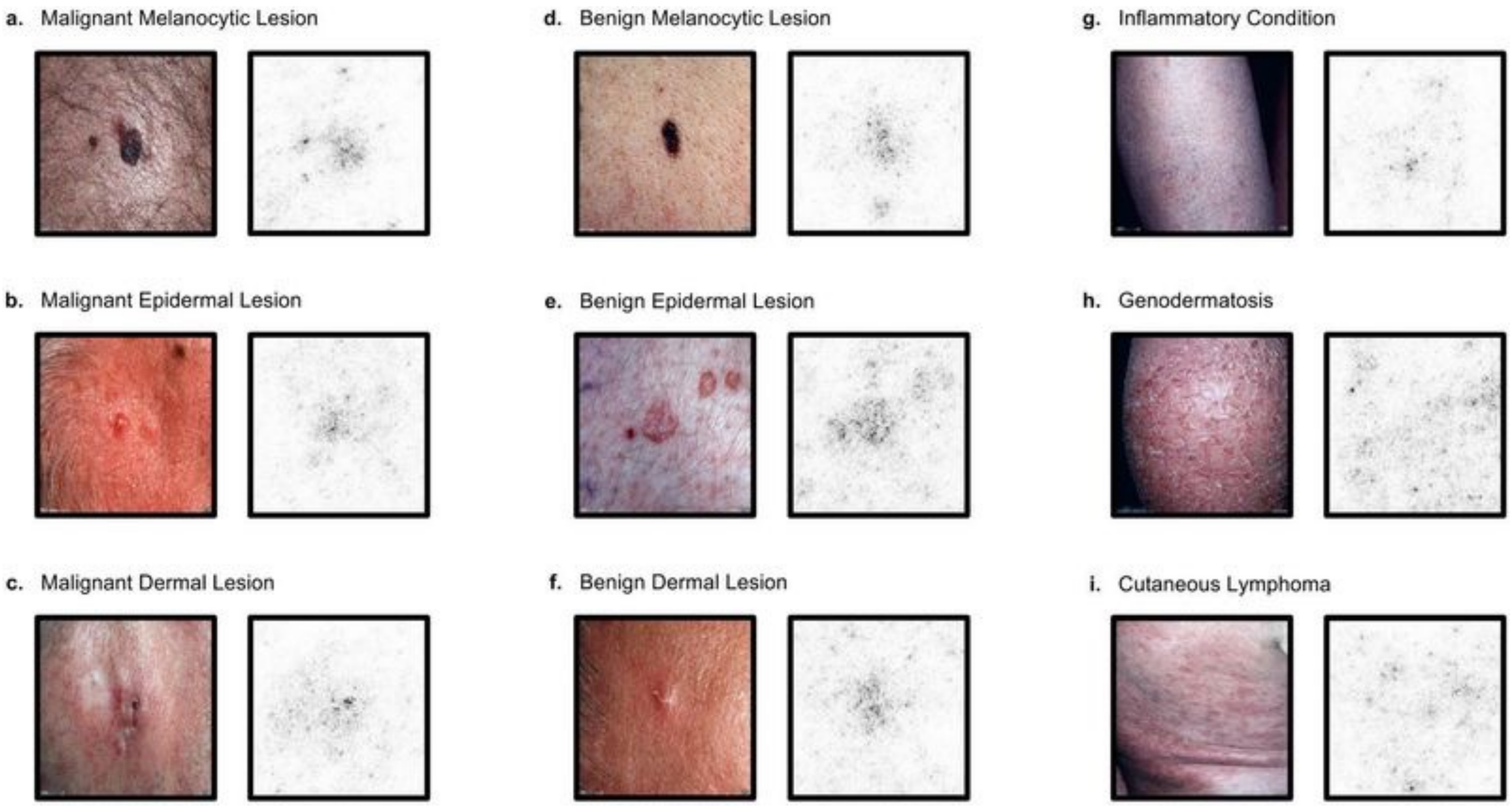


image

-1	5	-5
3	-5	9
-1	5	

$x_i^R \odot b$

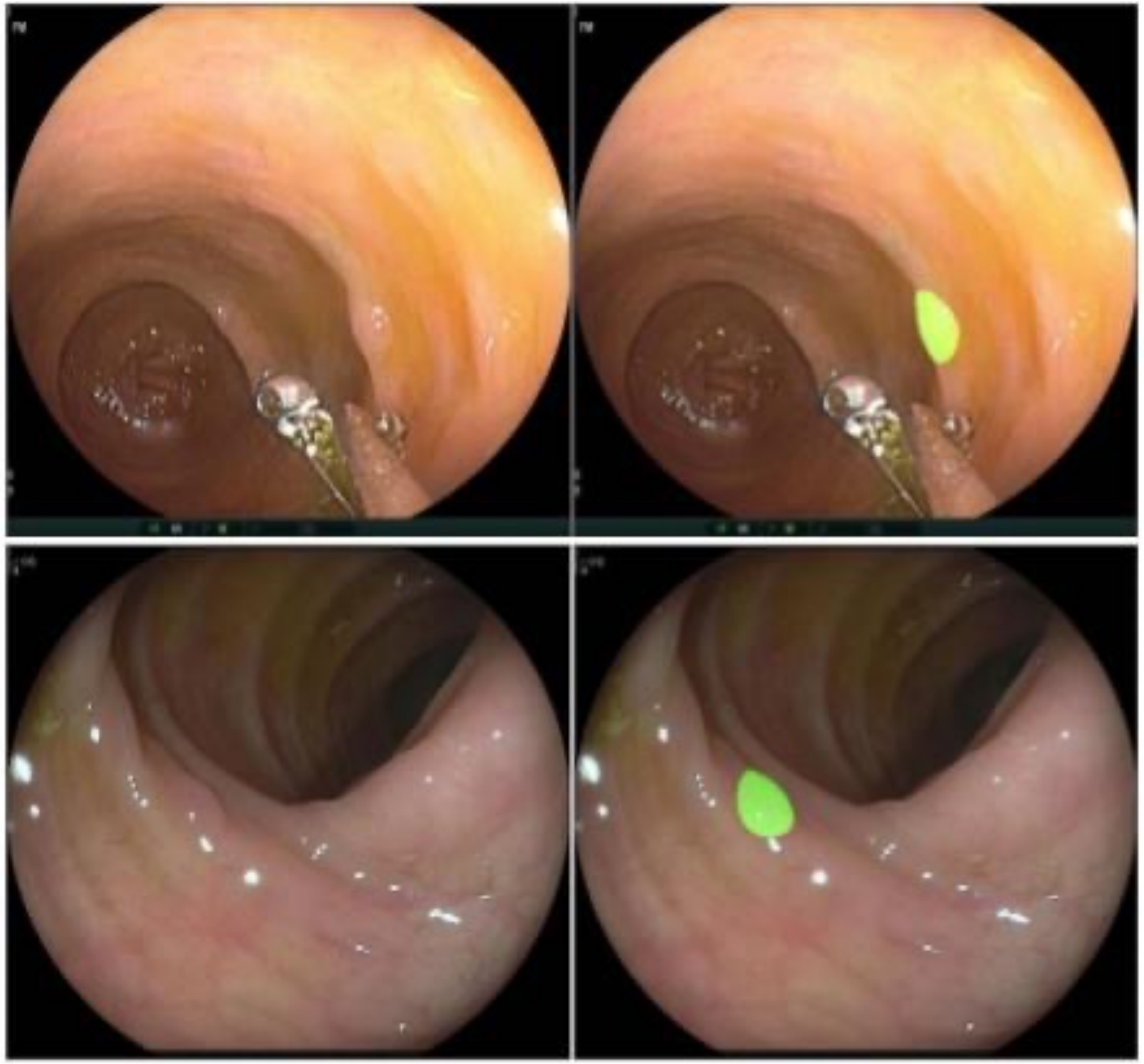
Saliency maps for example images



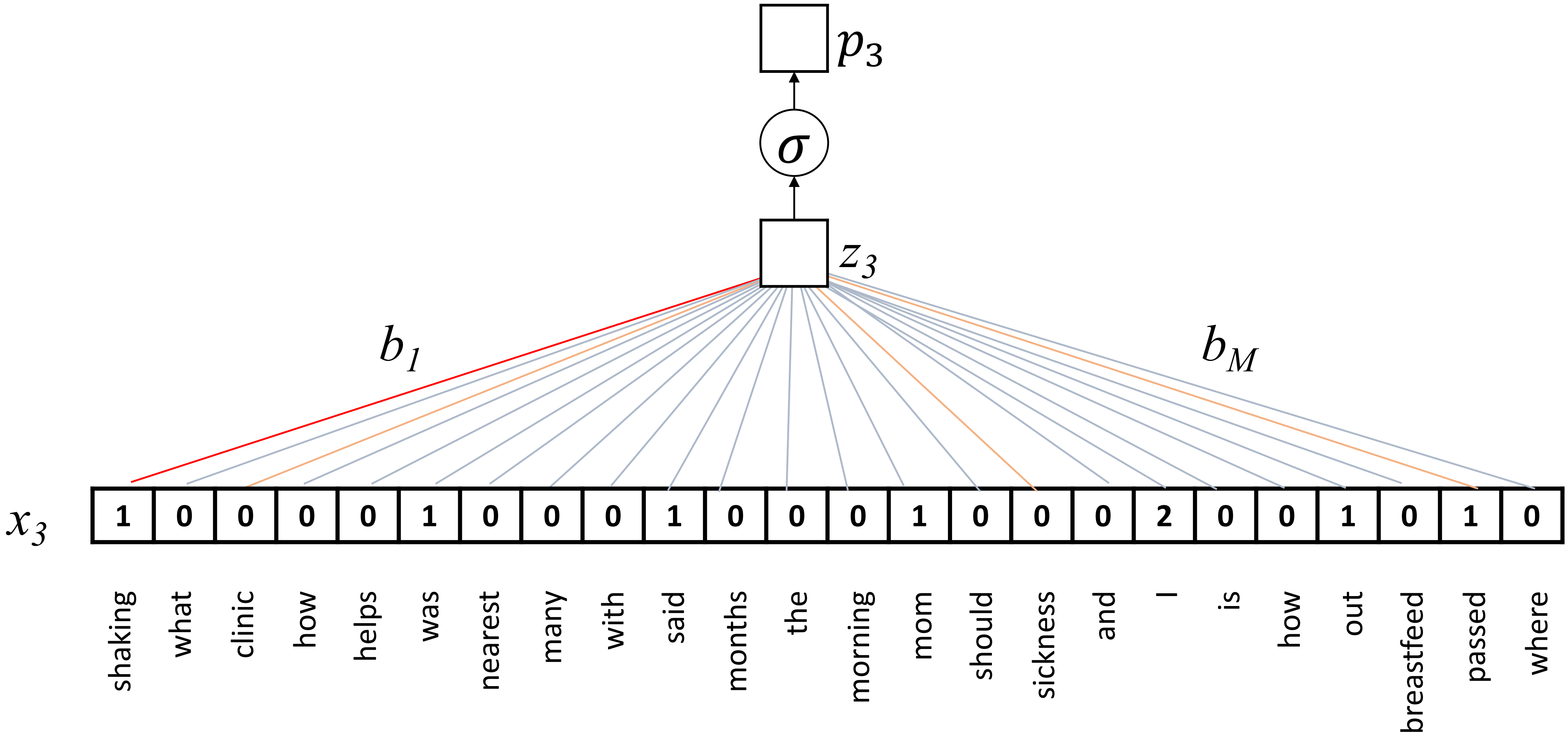
Saliency maps show gradients for each pixel with respect to the CNN's loss function. Darker pixels represent those with more influence.

Q: How much does this visualization help us understand the model?

Precisely Identify Boundaries

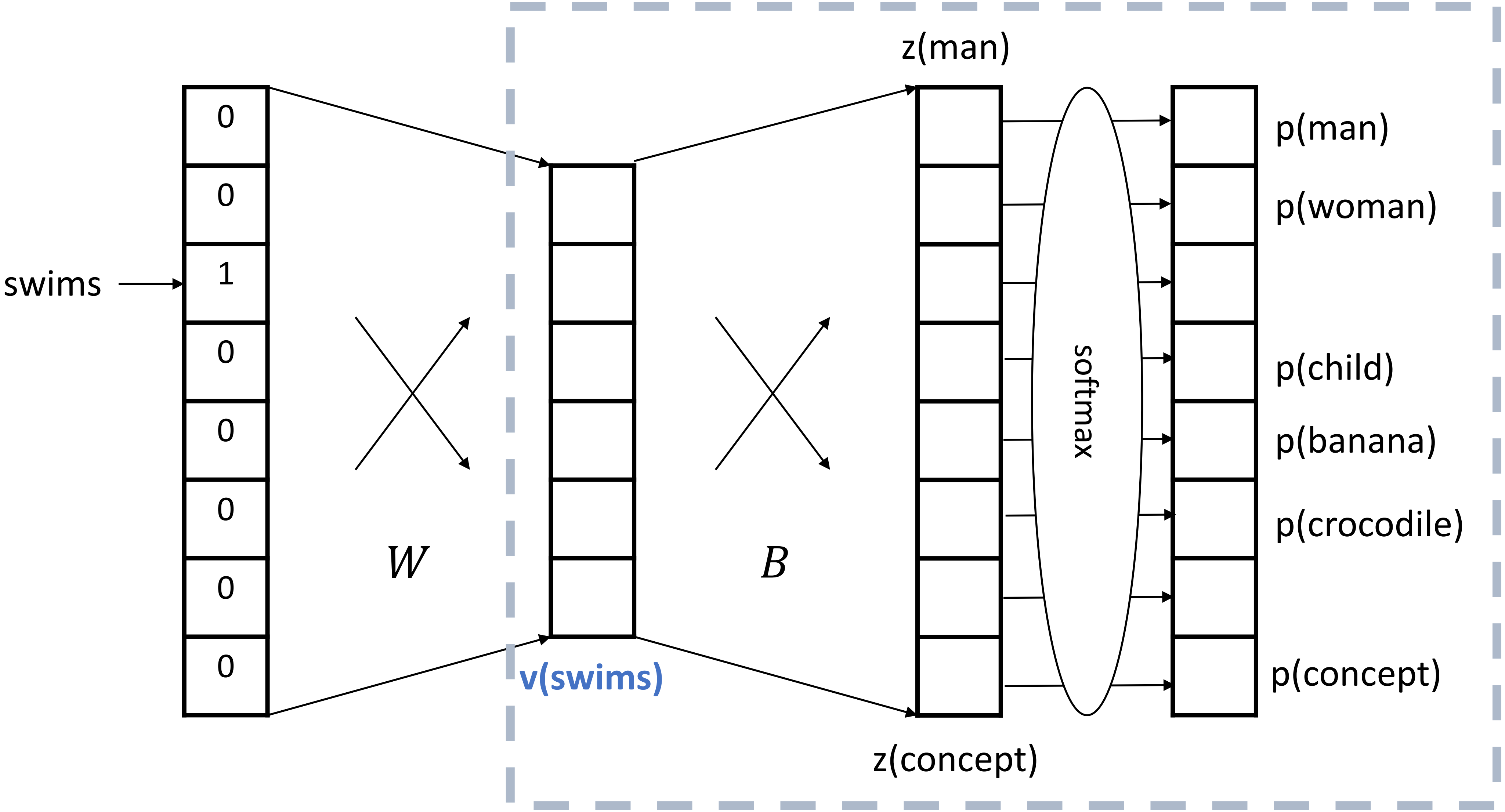


Logistic Regression for Text Classification

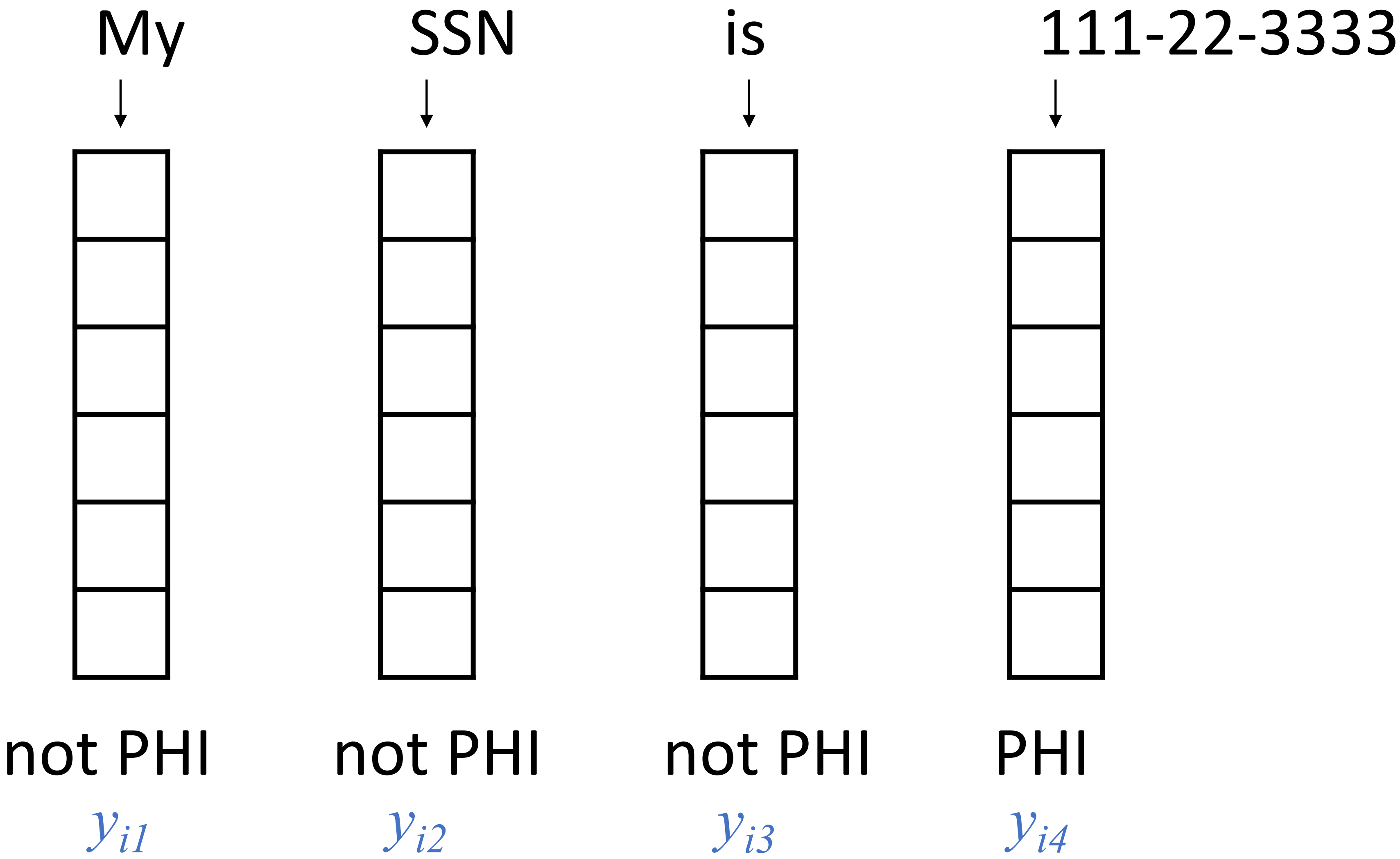


I passed out and Mom said I was shaking

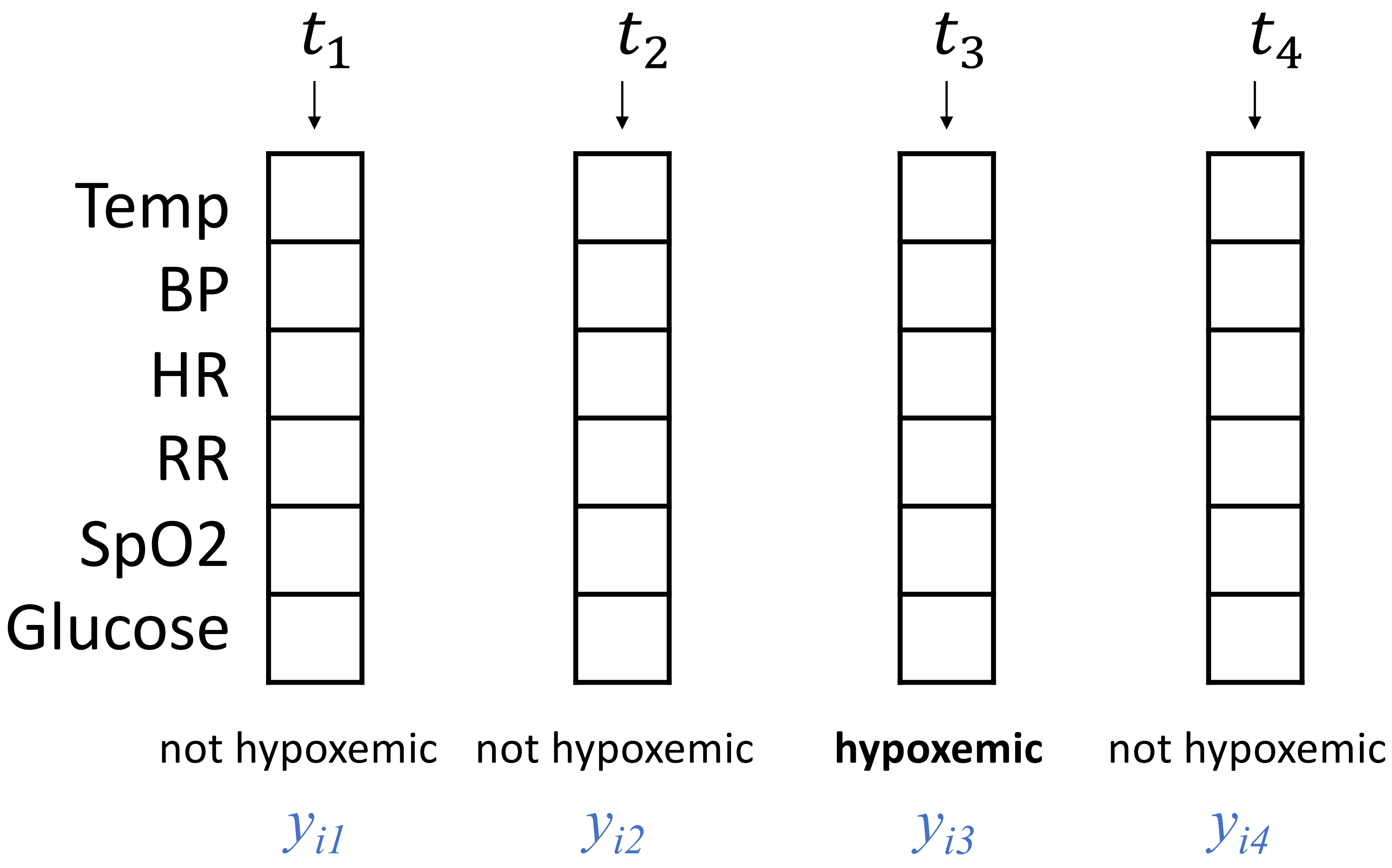
We now have a distributed representation of word *meaning* based on *context*



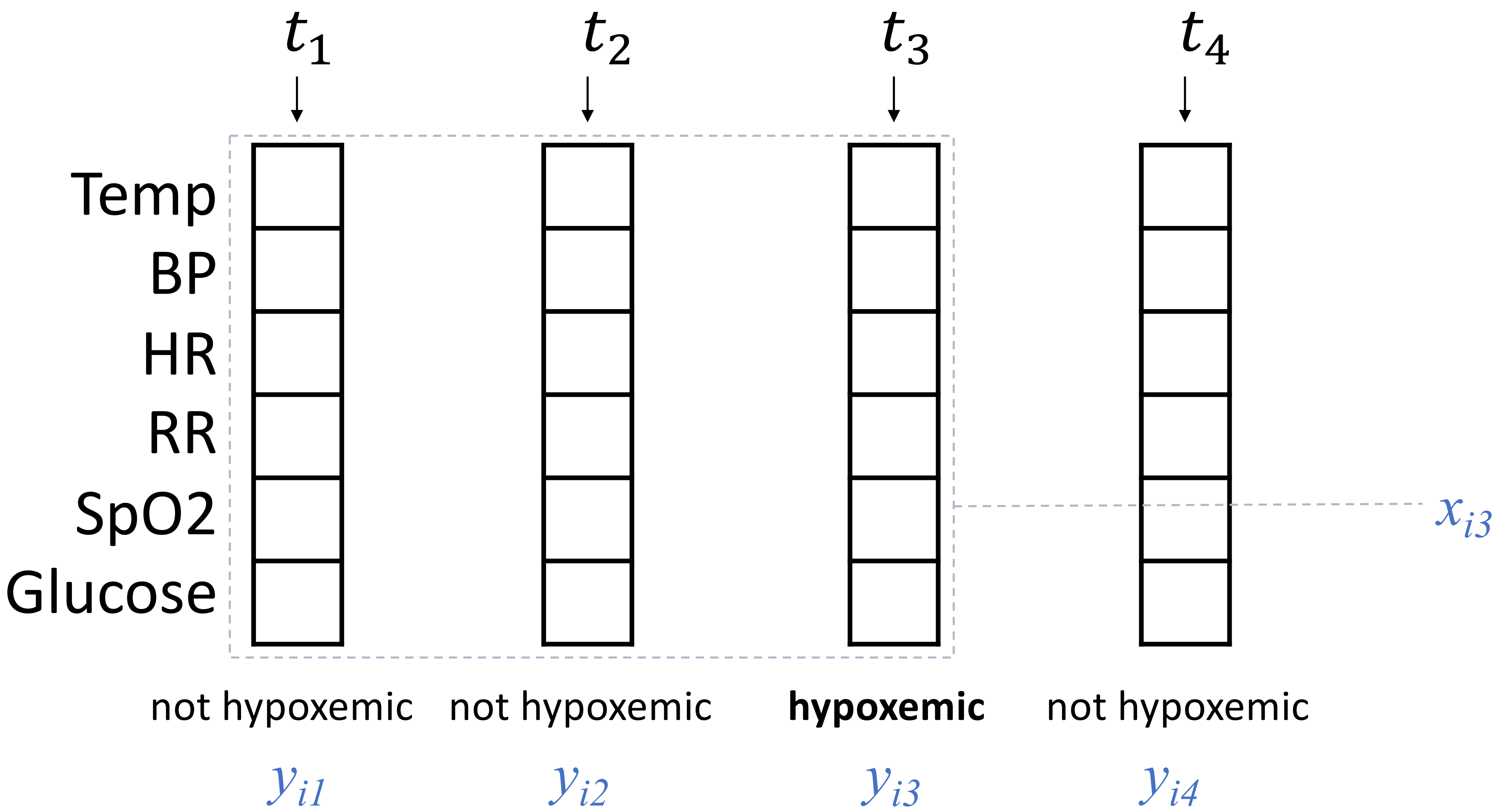
Task 2: Predict a label associated with each word



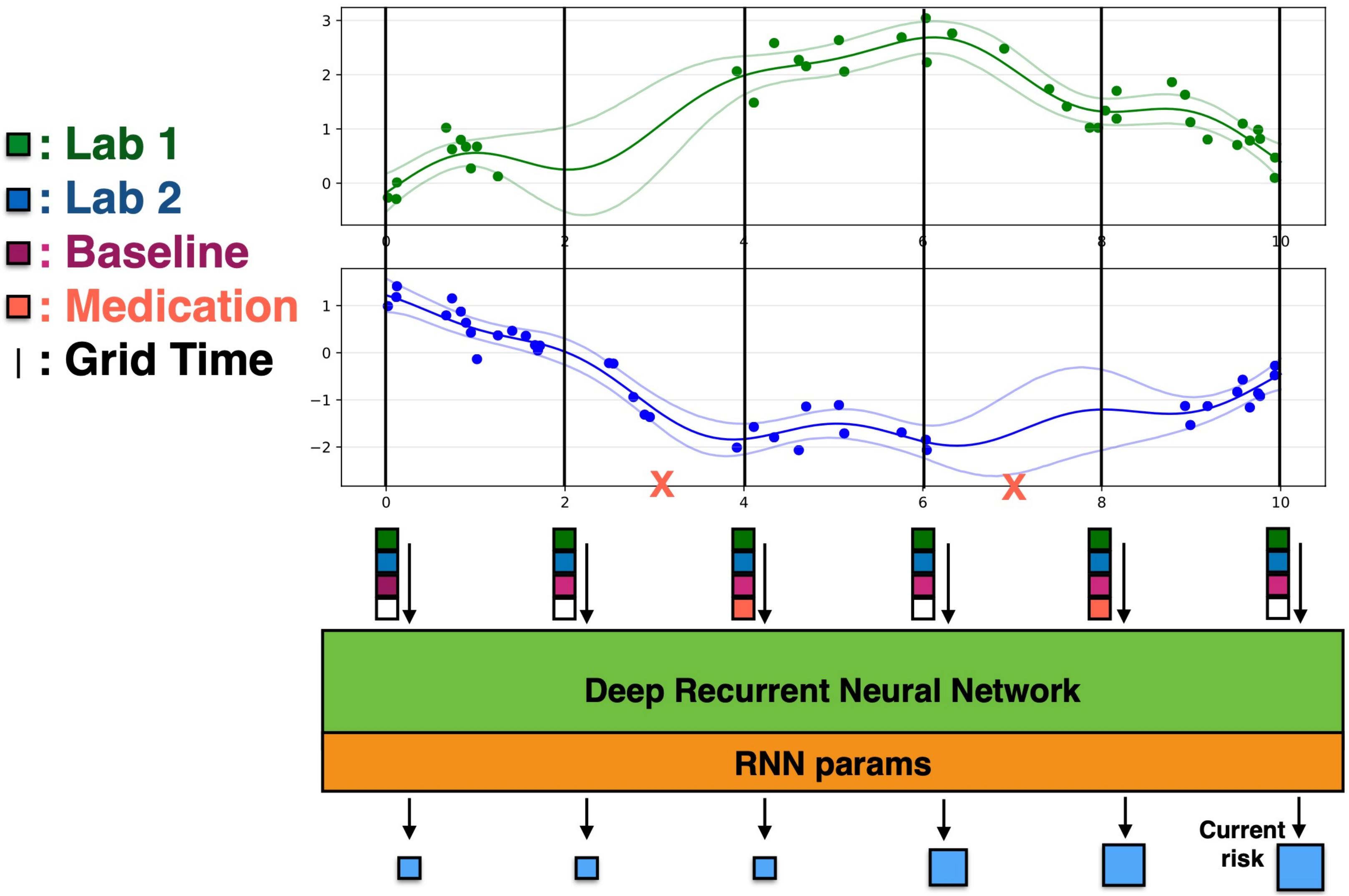
Task 2: Predict label assoc. with each time point



Similarly, we can aggregate measurements in a time-series



DIHI Sepsis Watch



<- Use GP regression to predict measurements at regular intervals

<- Predict sepsis risk using an RNN

Sequential Decision-Making

Make a series of decisions

based on a set of features (state)

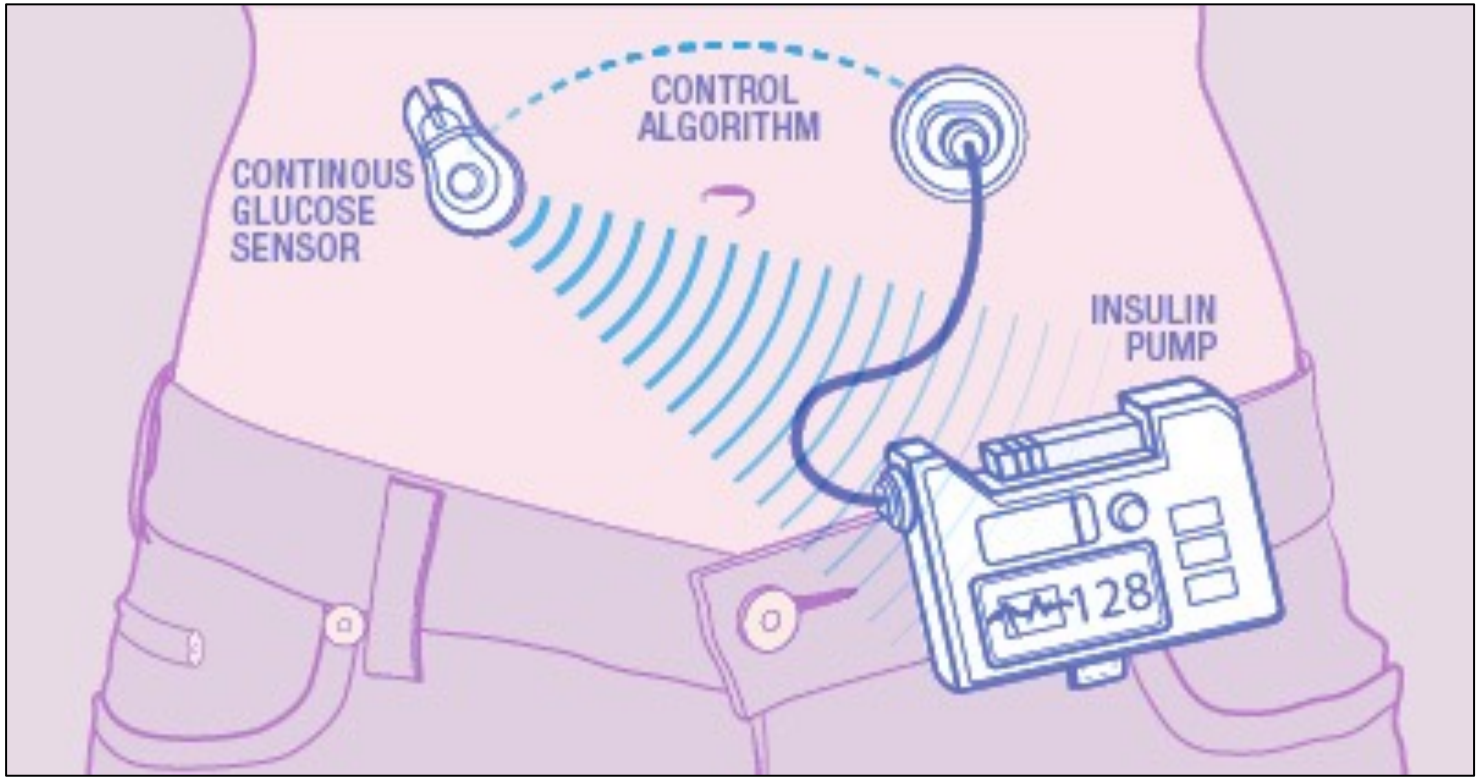
to maximize reward over time

Sequential Medical Decision-Making

Make a series of decisions

based on a set of features (state)

to maximize reward over time



Be in touch: m.engelhard@duke.edu

Thank you!