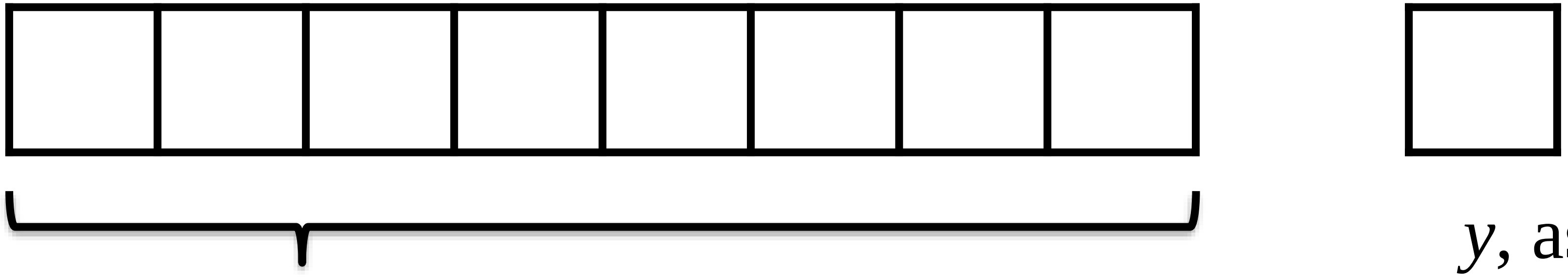
# Beyond Supervised Learning

## August 3, 2019

MMCi Applied Data Science
Block 5, Lecture 3

Matthew Engelhard

# Supervised Learning

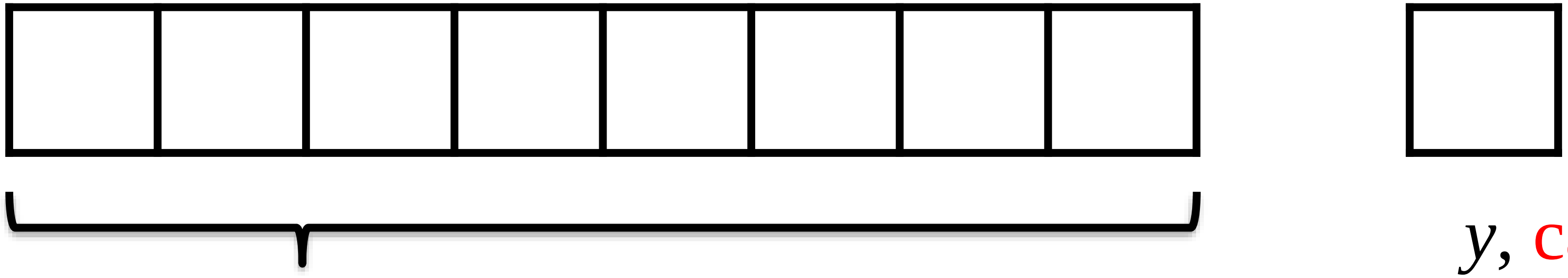


*y*, associated value or label

*x*, data/features for a subject or patient

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

# Supervised Learning: Classification



*x*, data/features for
a subject or patient

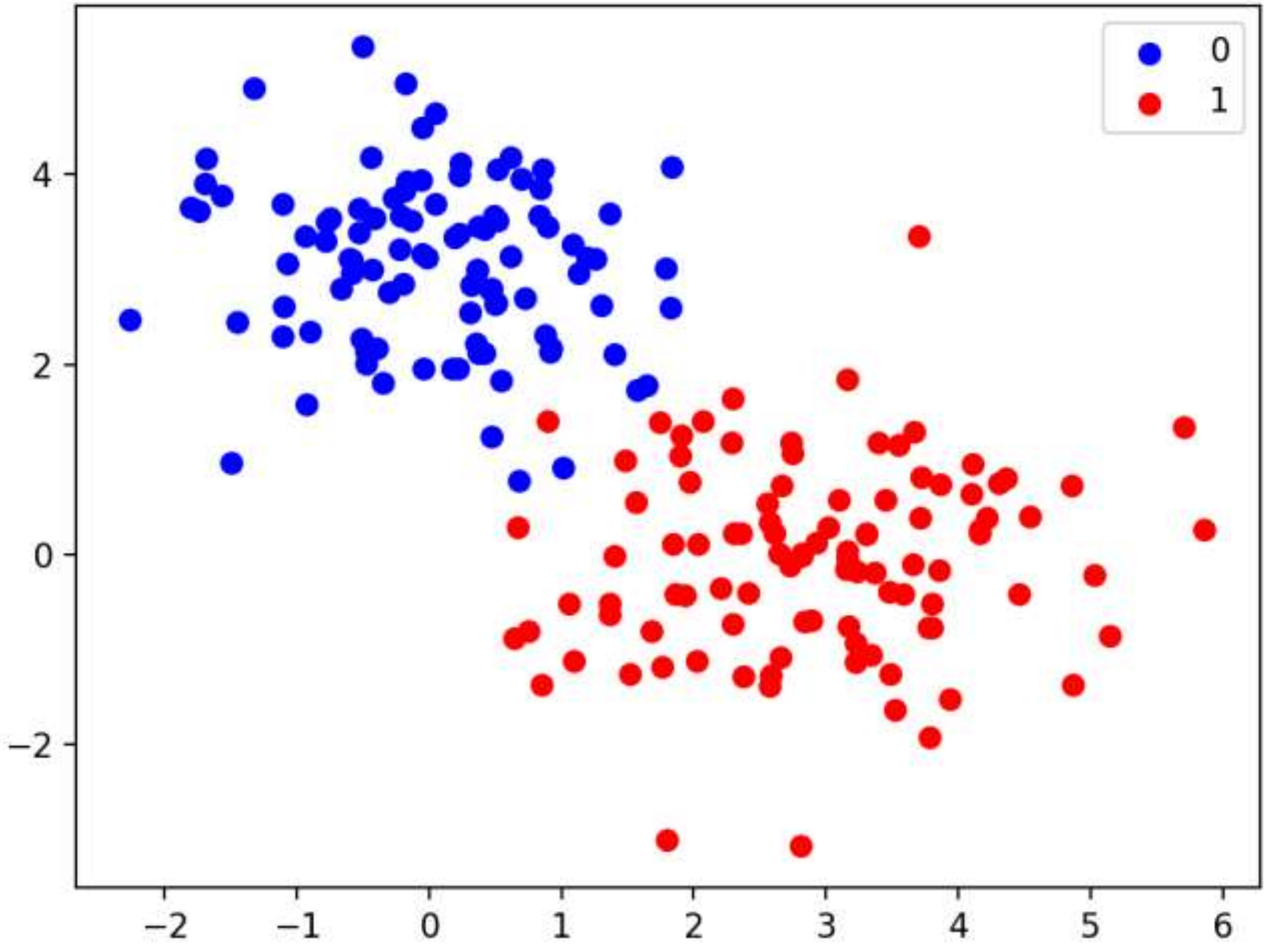
*y*, categorical
label

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

# Supervised Learning: Classification

<u>Goal</u>:

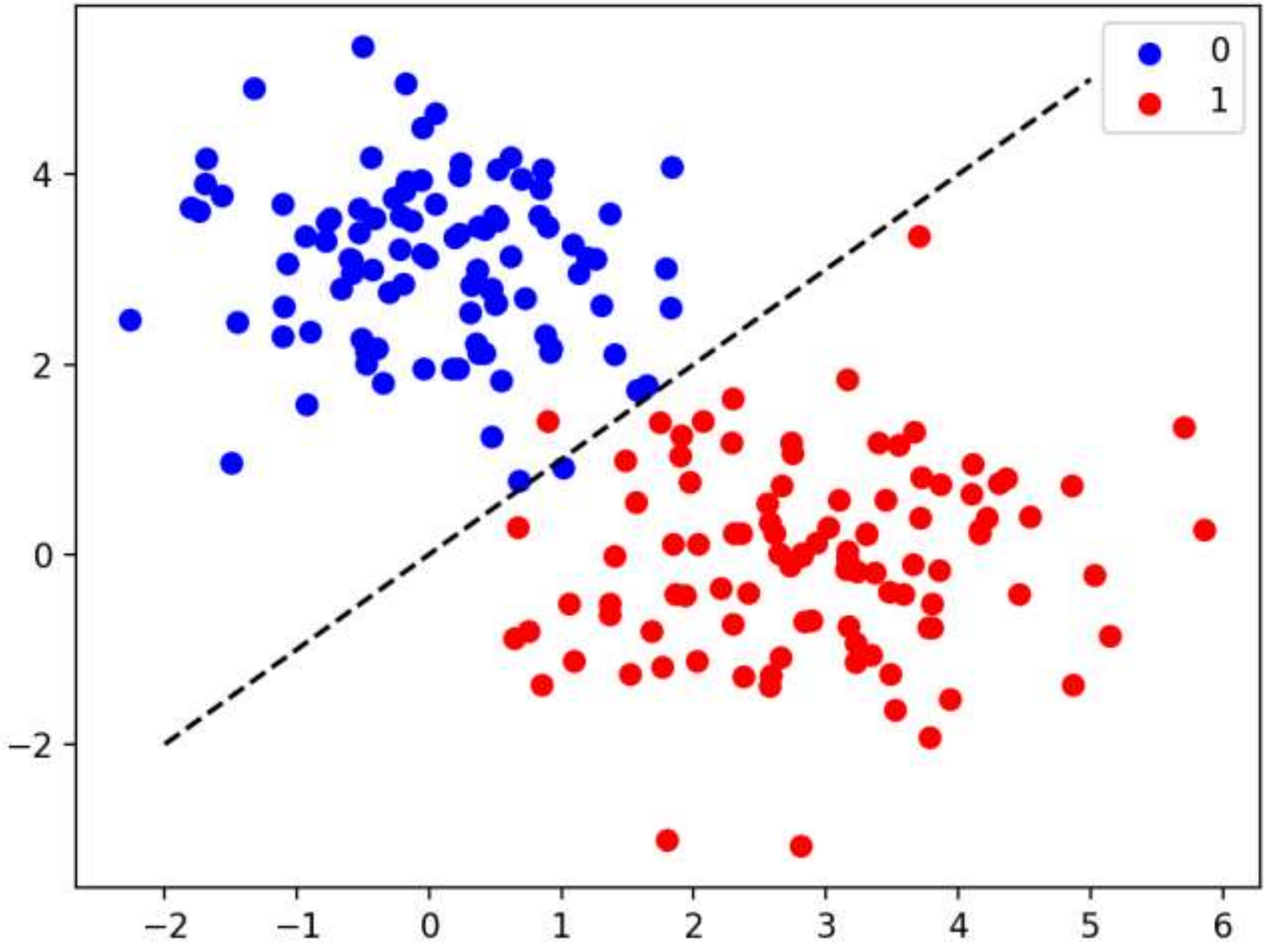
Learn a **decision boundary** that separates 0s from 1s



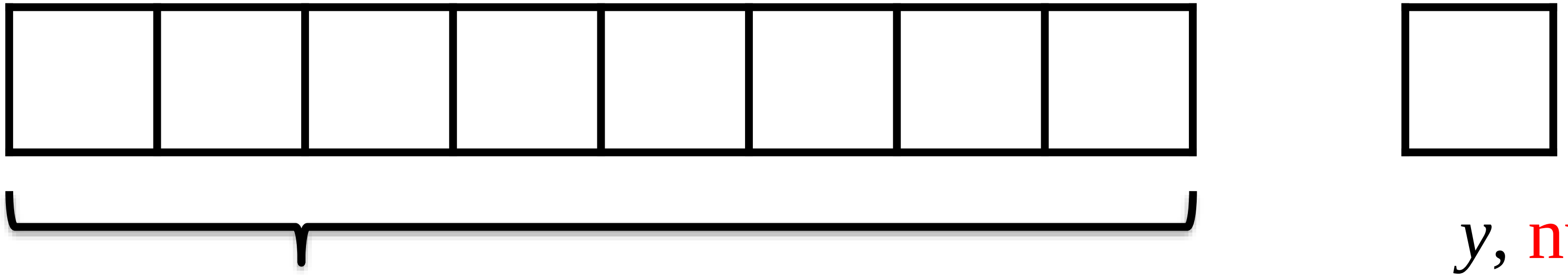
# Supervised Learning: Classification

<u>Goal</u>:

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



# Supervised Learning: Regression



*y*, numeric value

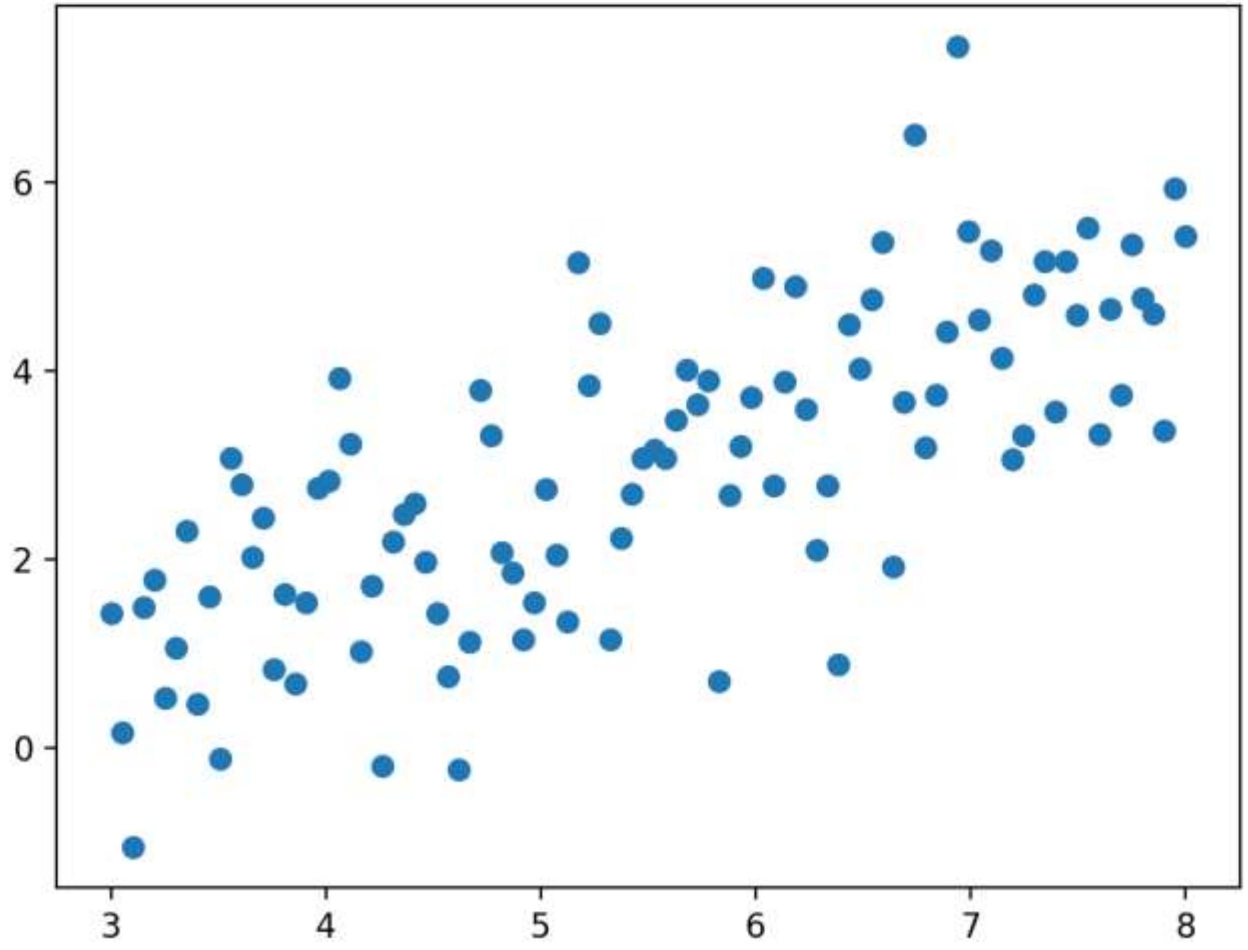
*x*, data/features for a subject or patient

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

# Supervised Learning: Regression

Goal:

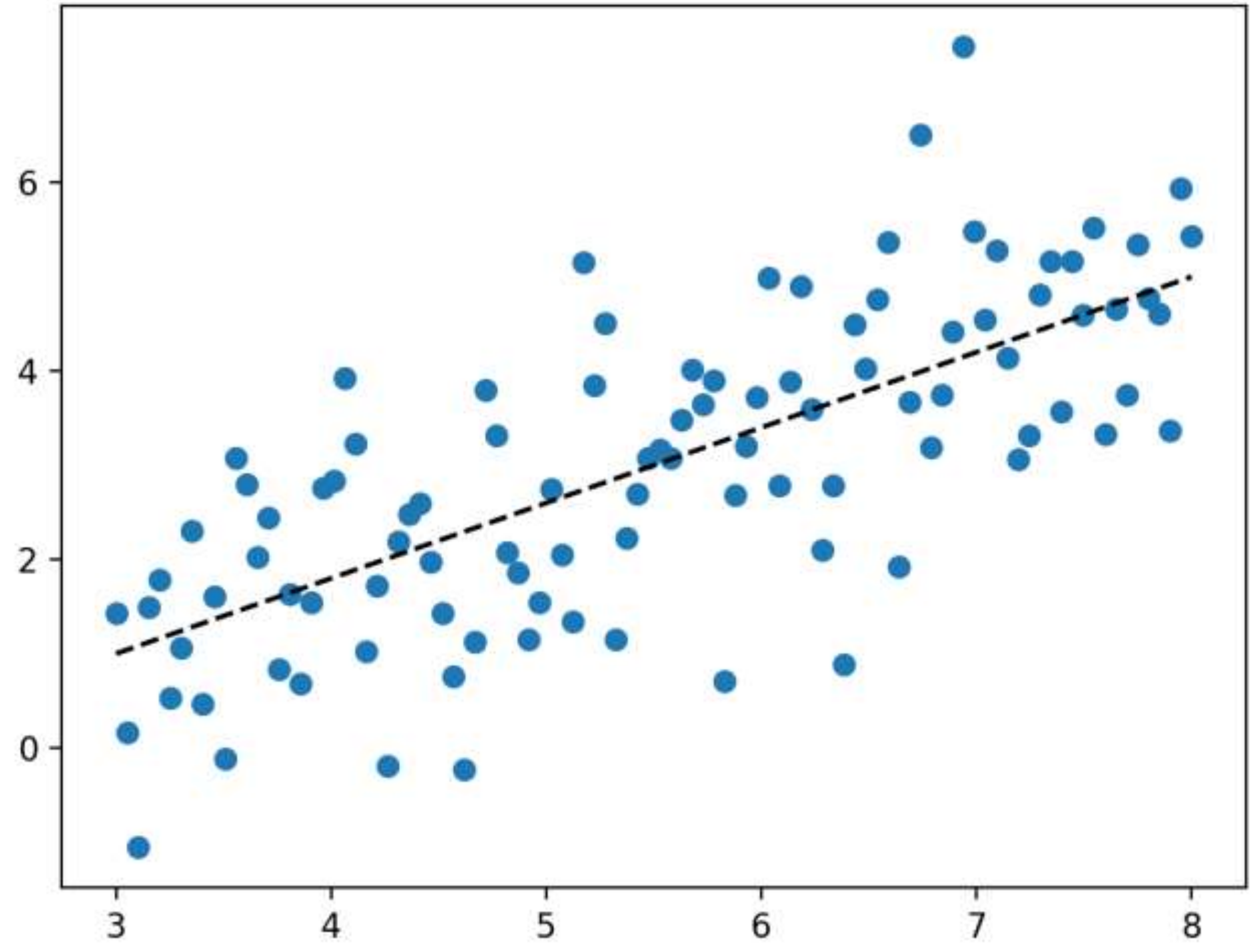
Learn a function that predicts $y$ based on $x$



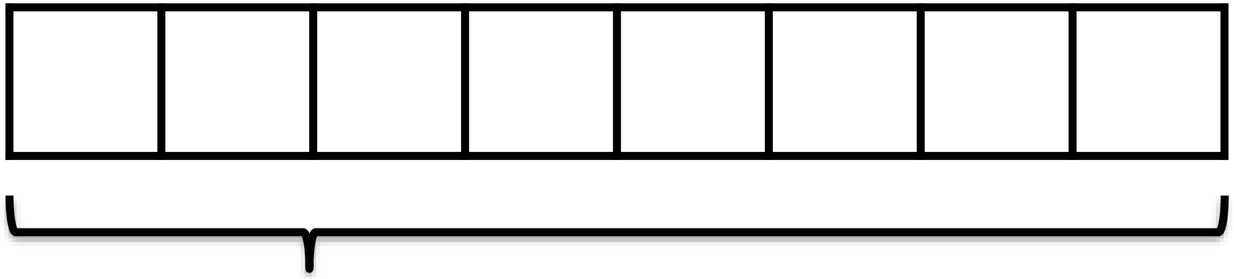
# Supervised Learning: Regression

Goal:

Learn a function that predicts $y$ based on $x$



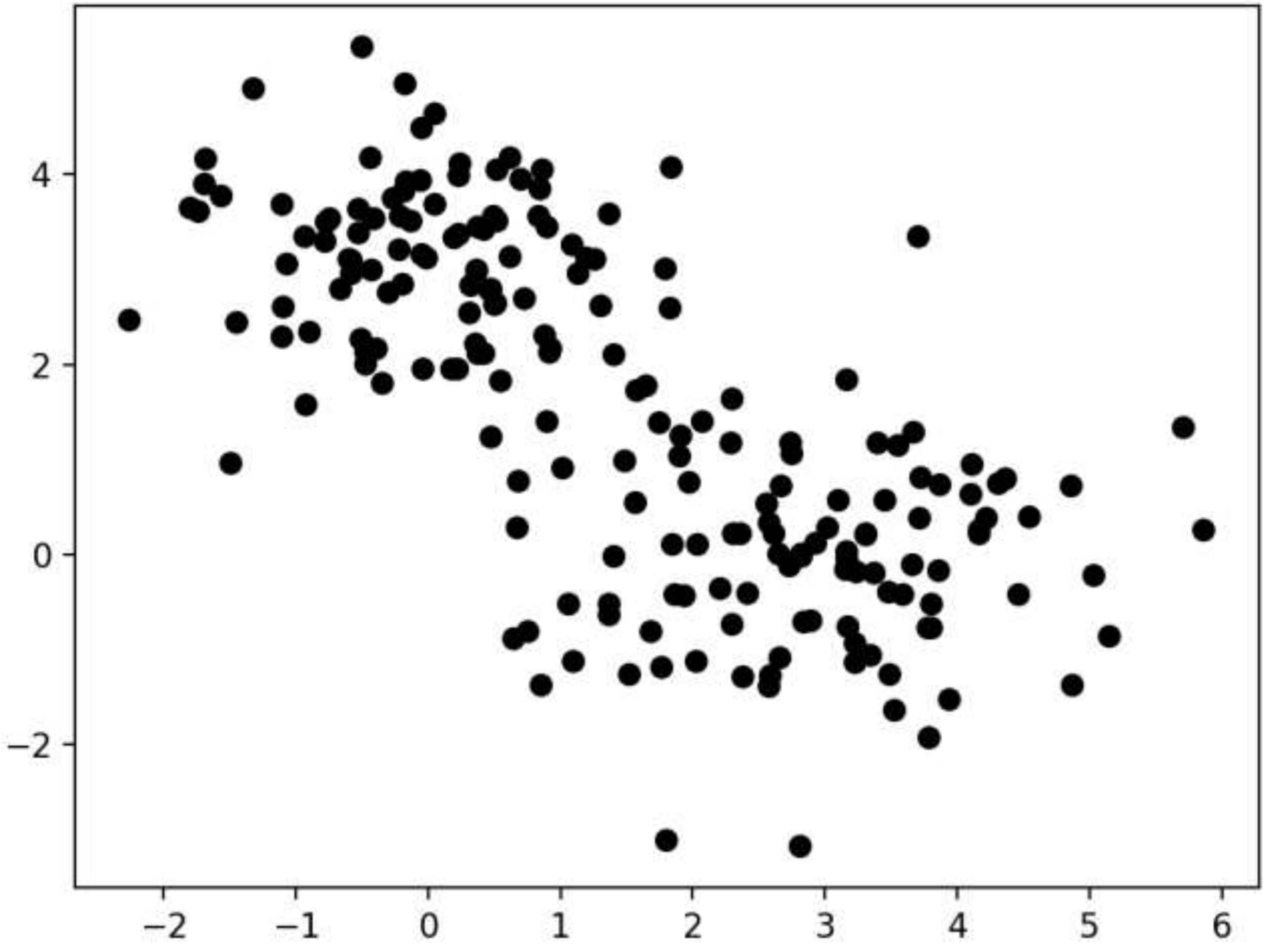
# Unsupervised Learning



*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

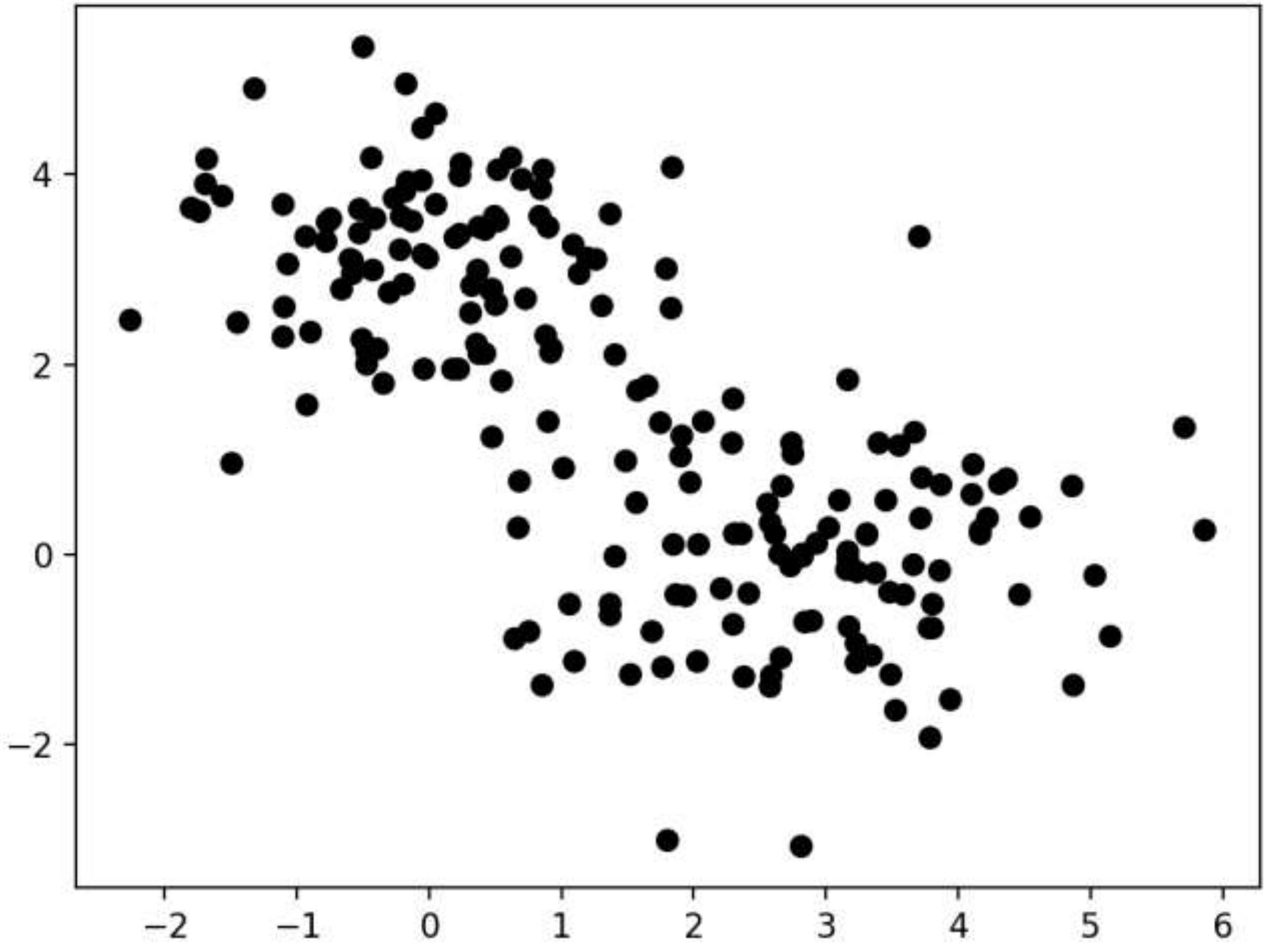
# Unsupervised Learning



# Unsupervised Learning: Clustering

<u>Goal</u>:

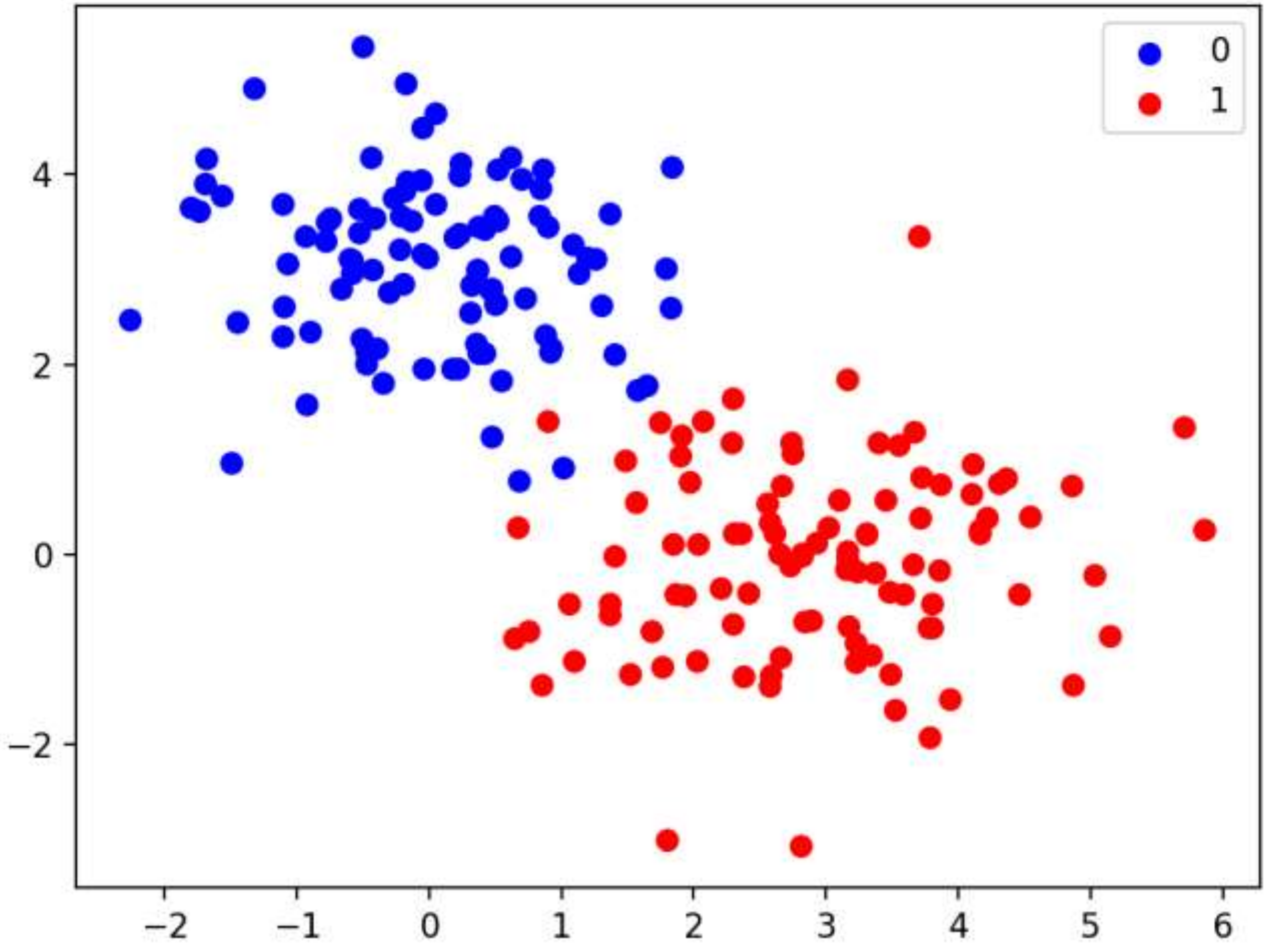
**Assign points to distinct groups** with shared characteristics



# Unsupervised Learning: Clustering

Goal:

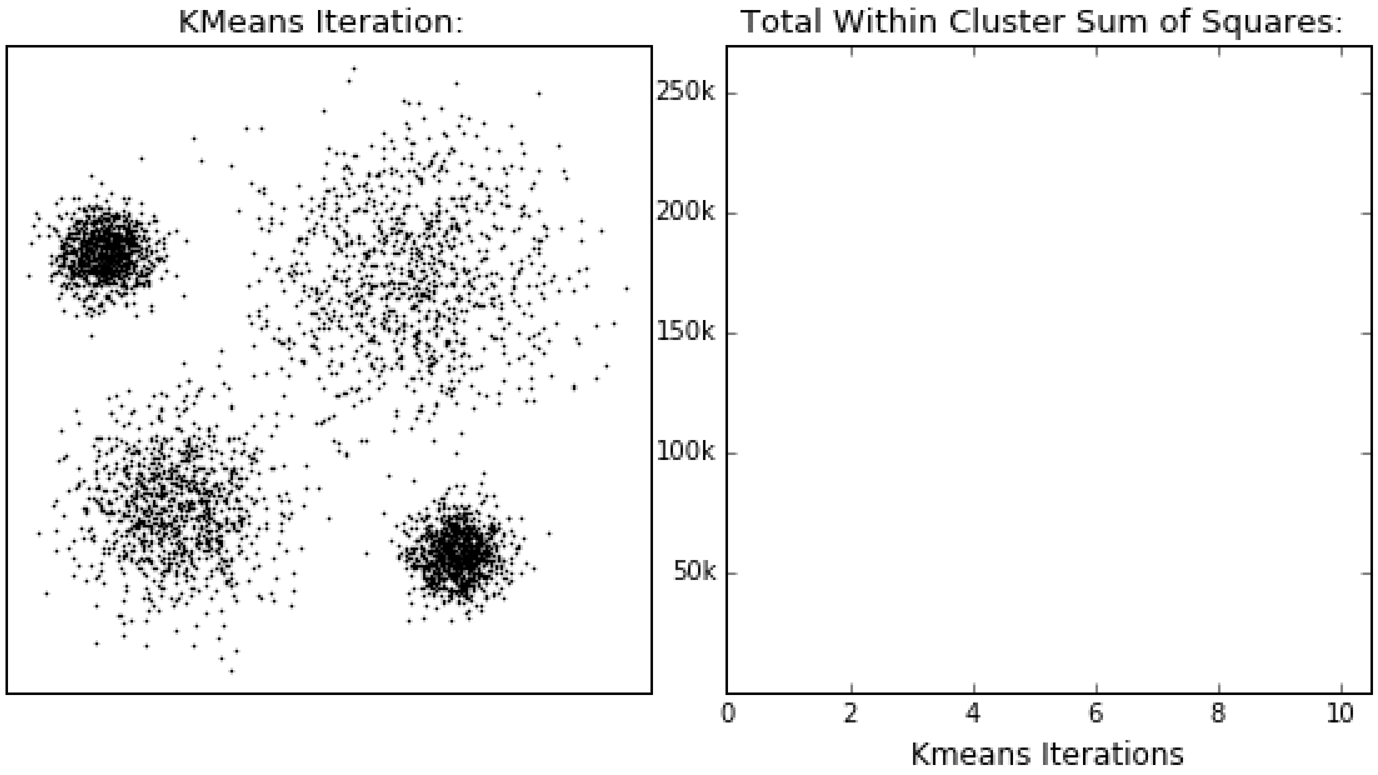
**Assign points to distinct groups** with shared characteristics



# Example of K-Means Clustering

<u>Goal</u>:

**Assign points to distinct groups** with shared characteristics

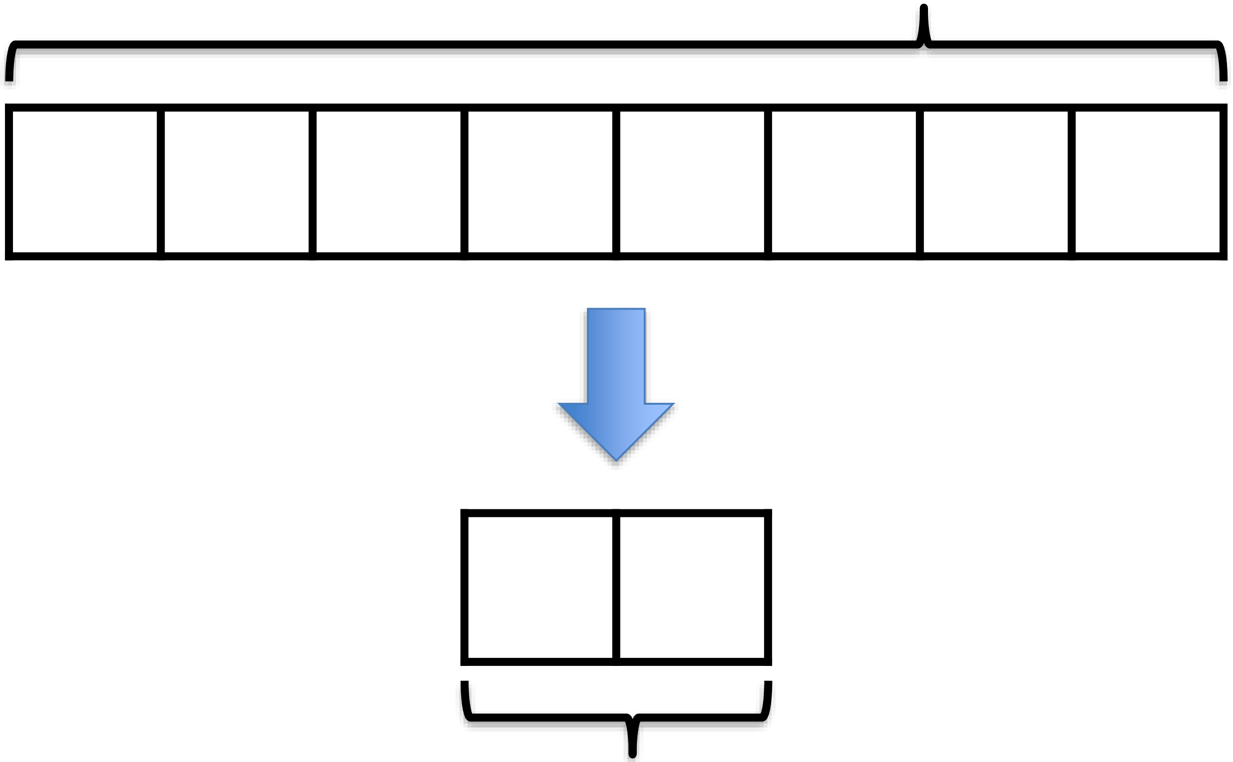



- sensitive to initialization
- minimize ***Σ***distance$^2$ from points to centers

# Unsupervised Learning: Dimensionality Reduction

<u>Goal</u>:

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

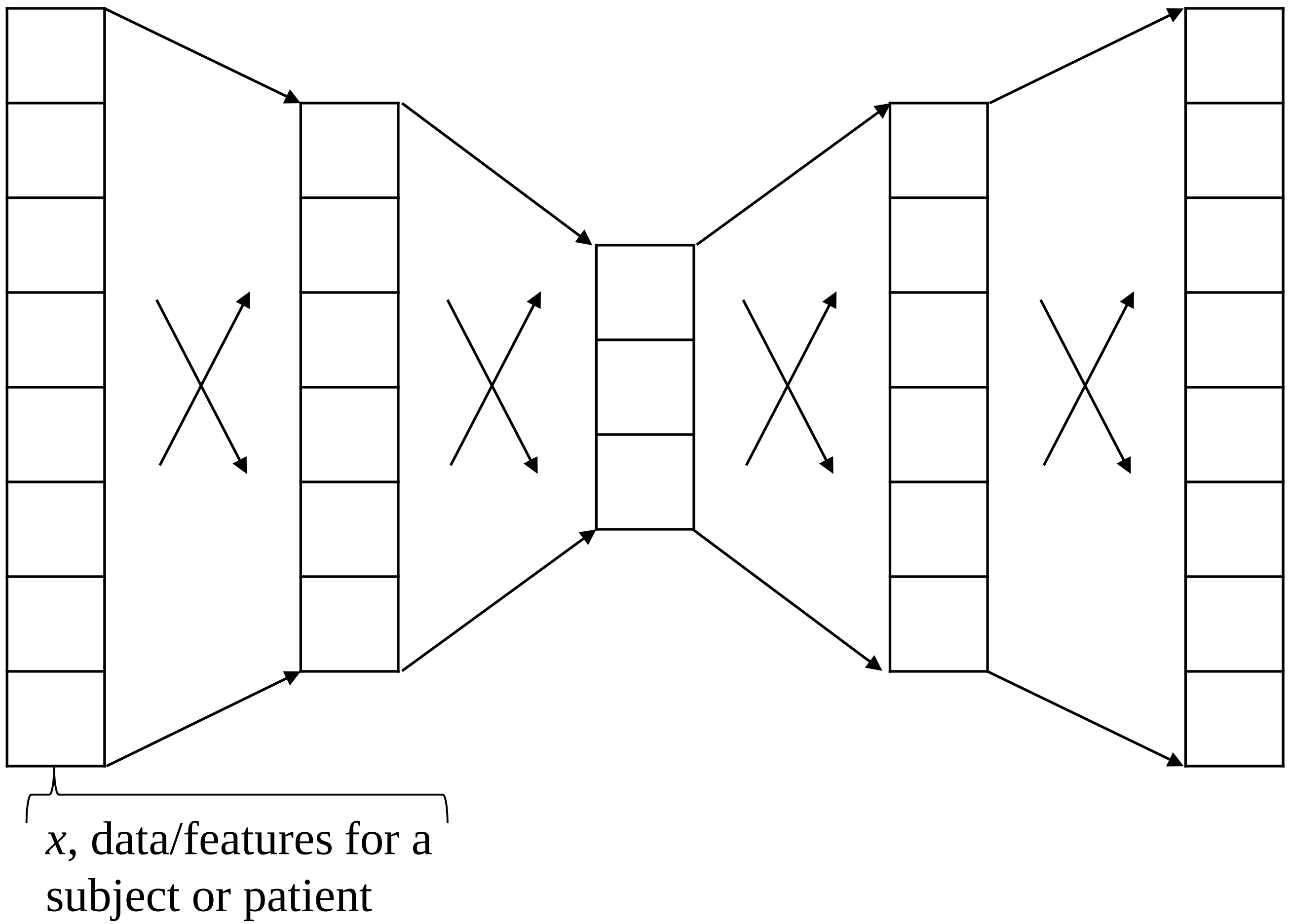


# Dimensionality Reduction Example: Autoencoder

<u>Goal</u>:

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

- Train to minimize **reconstruction loss**

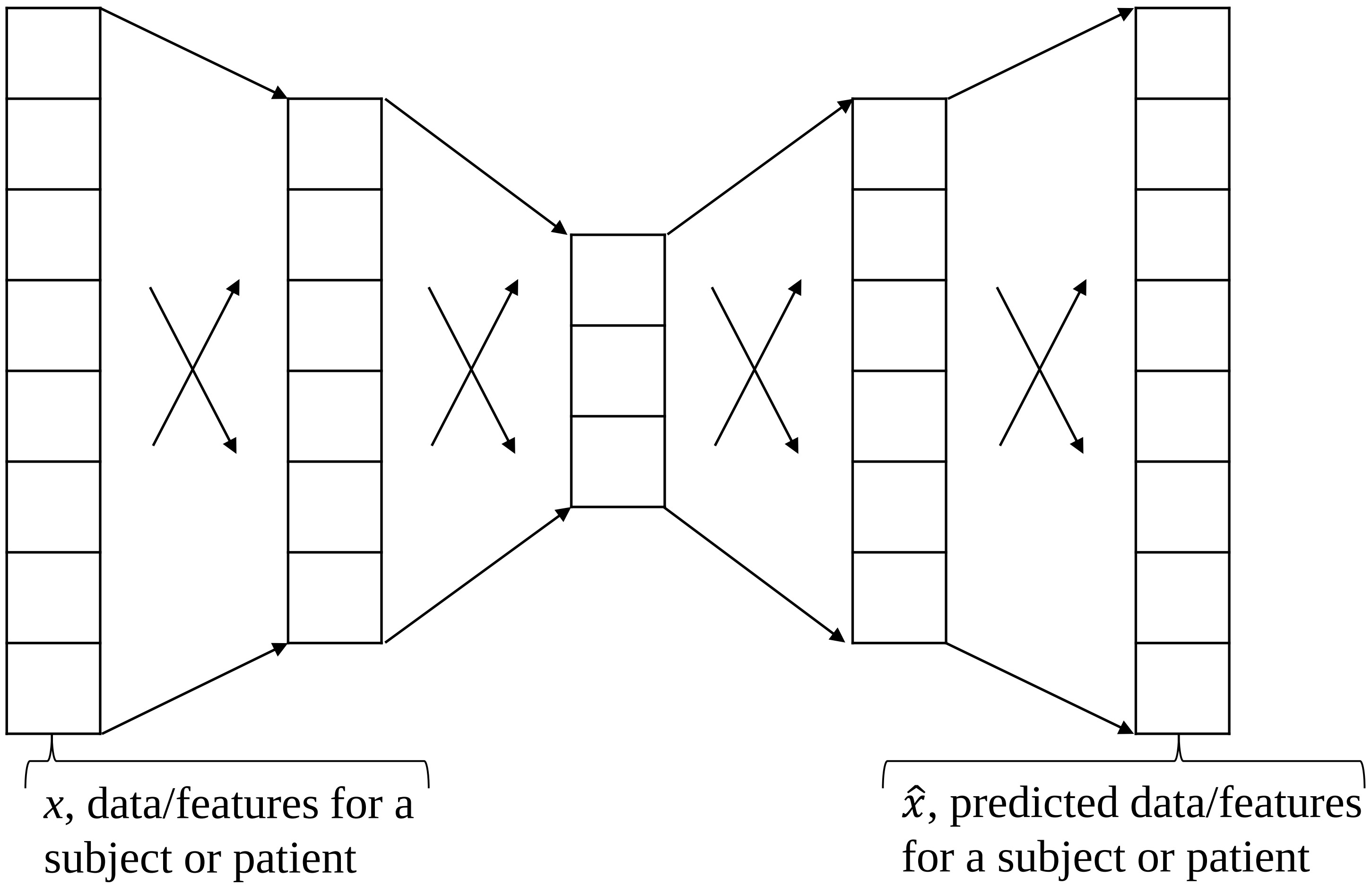


# Dimensionality Reduction Example: Autoencoder

<u>Goal</u>:

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

- Train to minimize **reconstruction loss**

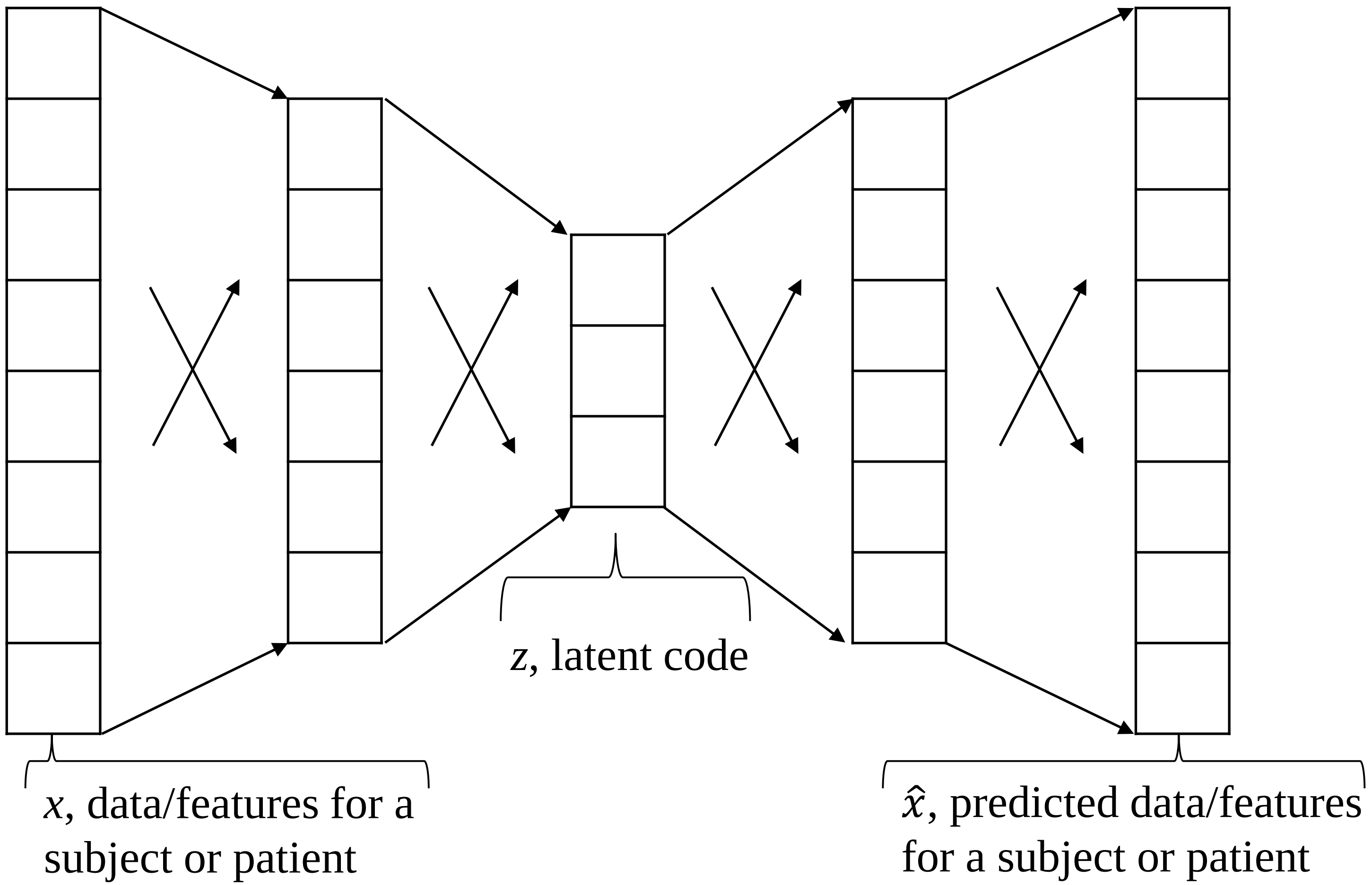


# Dimensionality Reduction Example: Autoencoder

Goal:

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

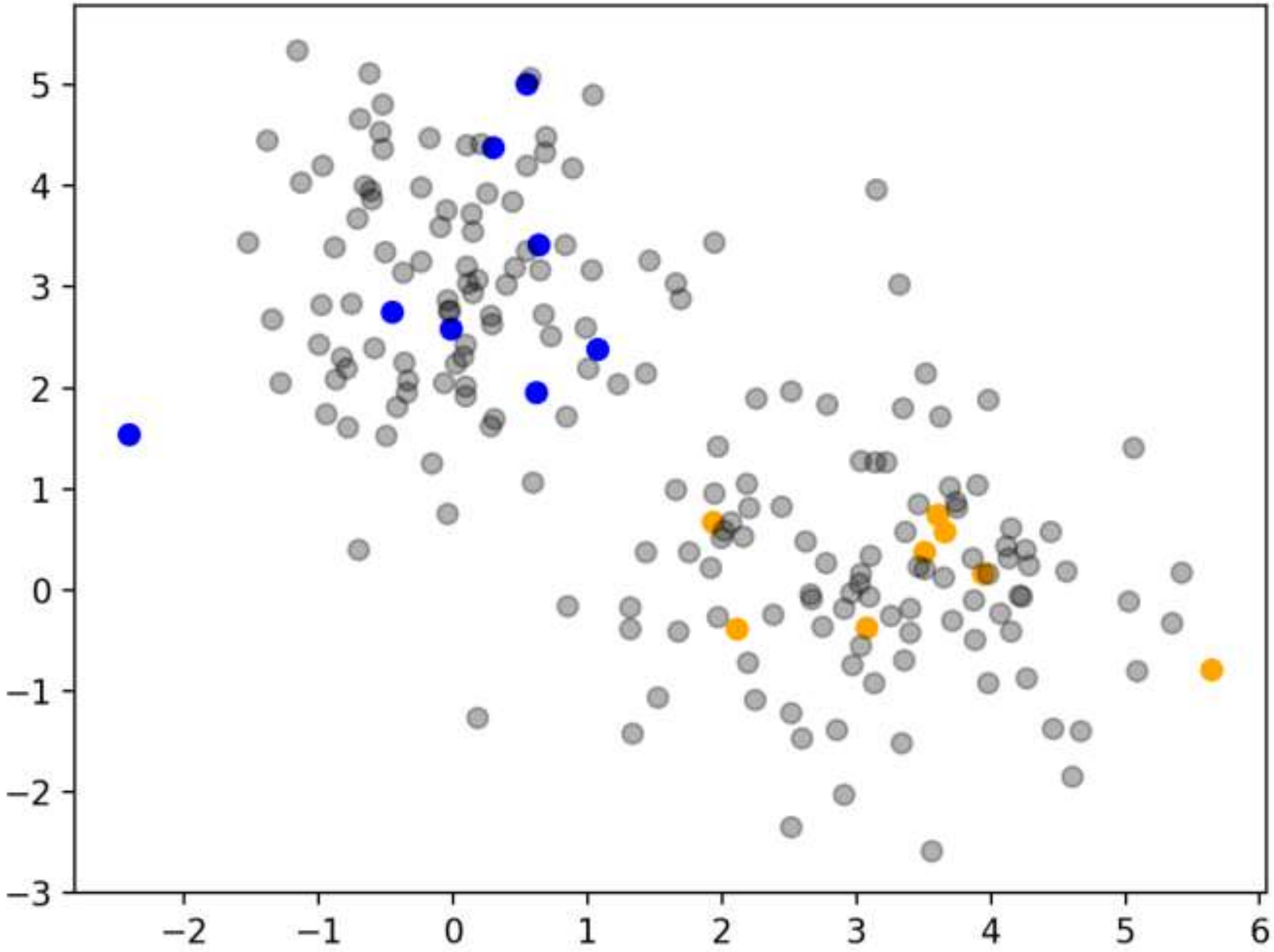
- Train to minimize **reconstruction loss**



# Semi-Supervised Learning

- Some points are labeled, some are not

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

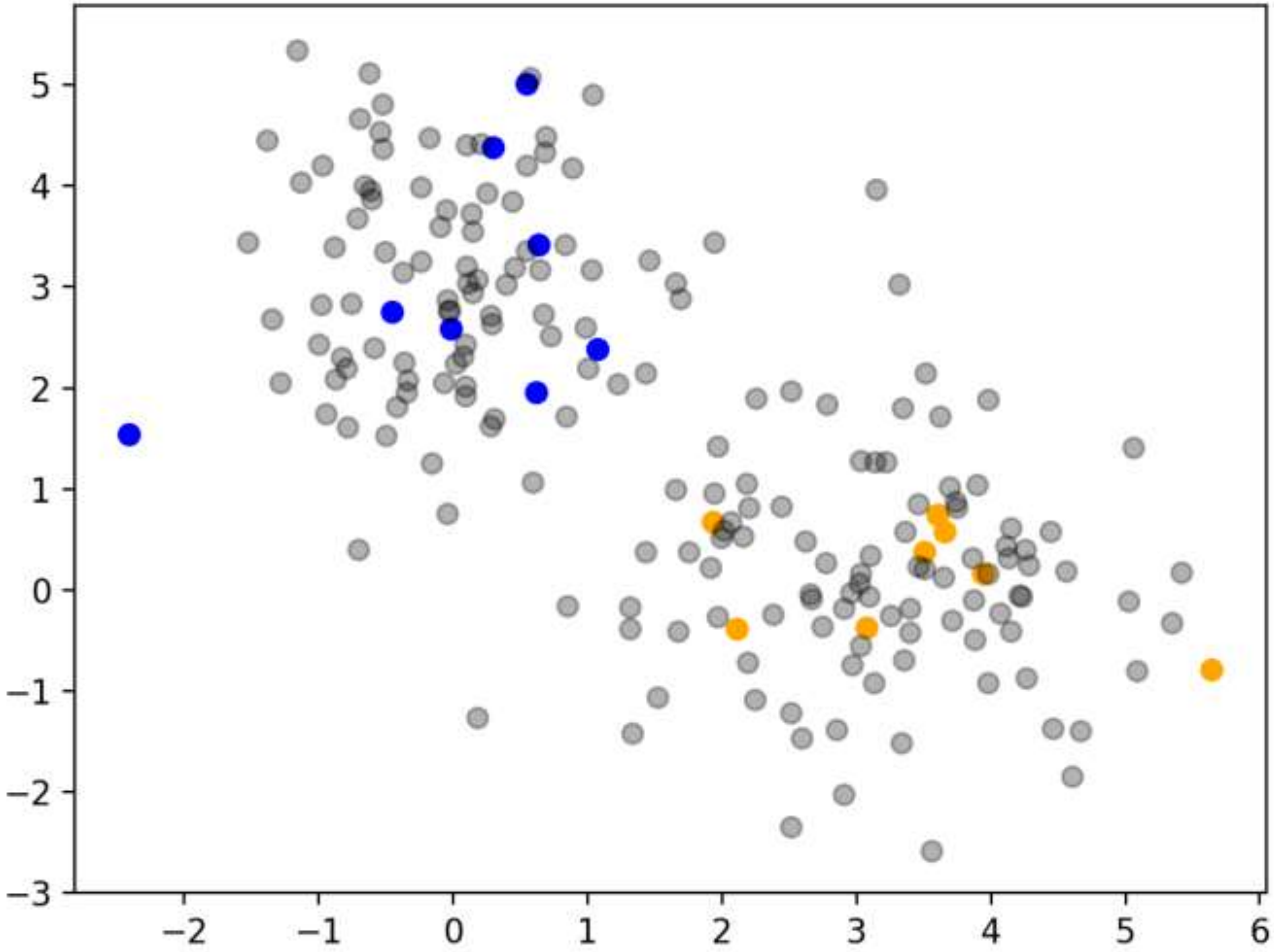


# Active Learning

- 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

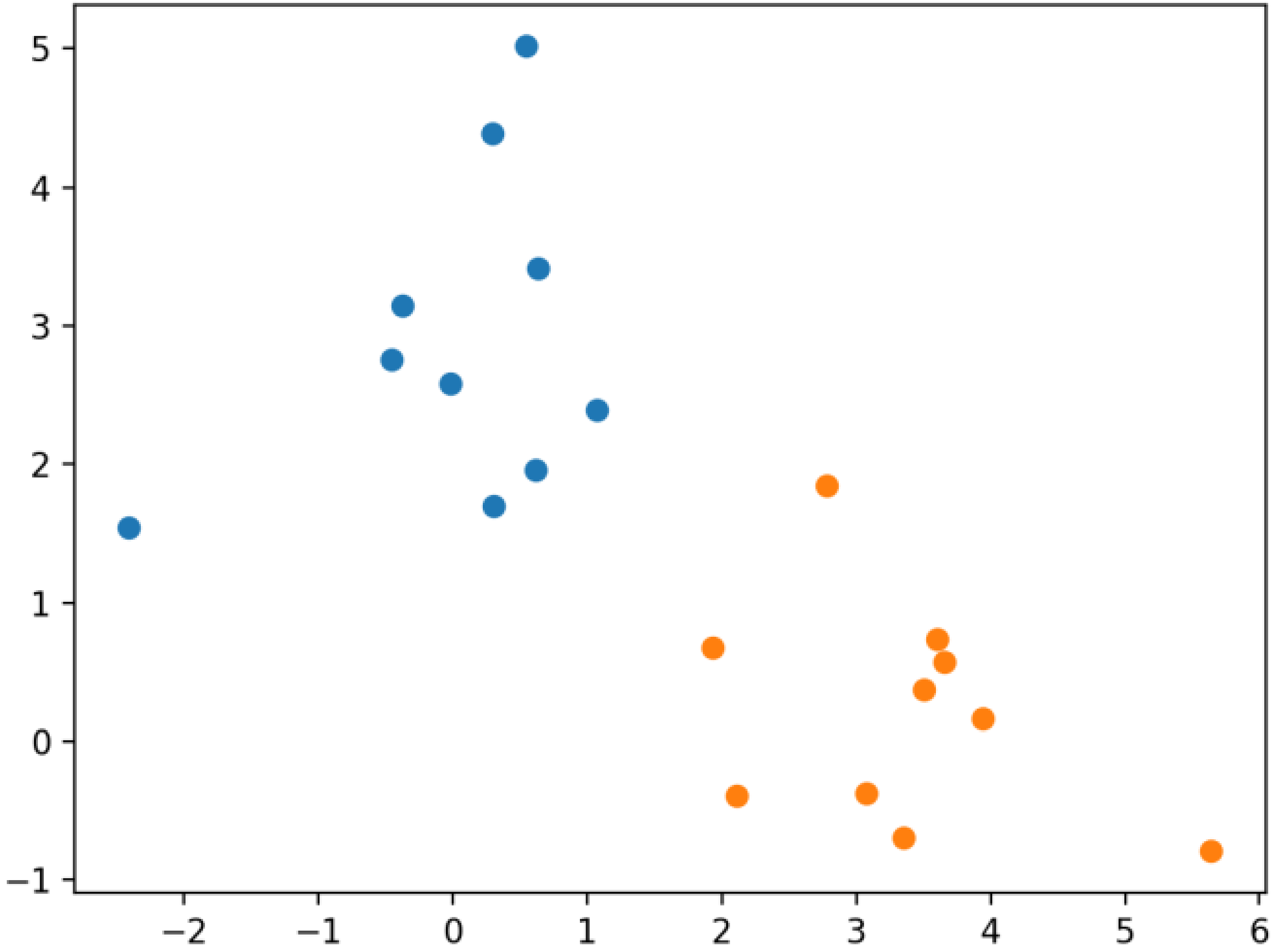


# Online Learning

- 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

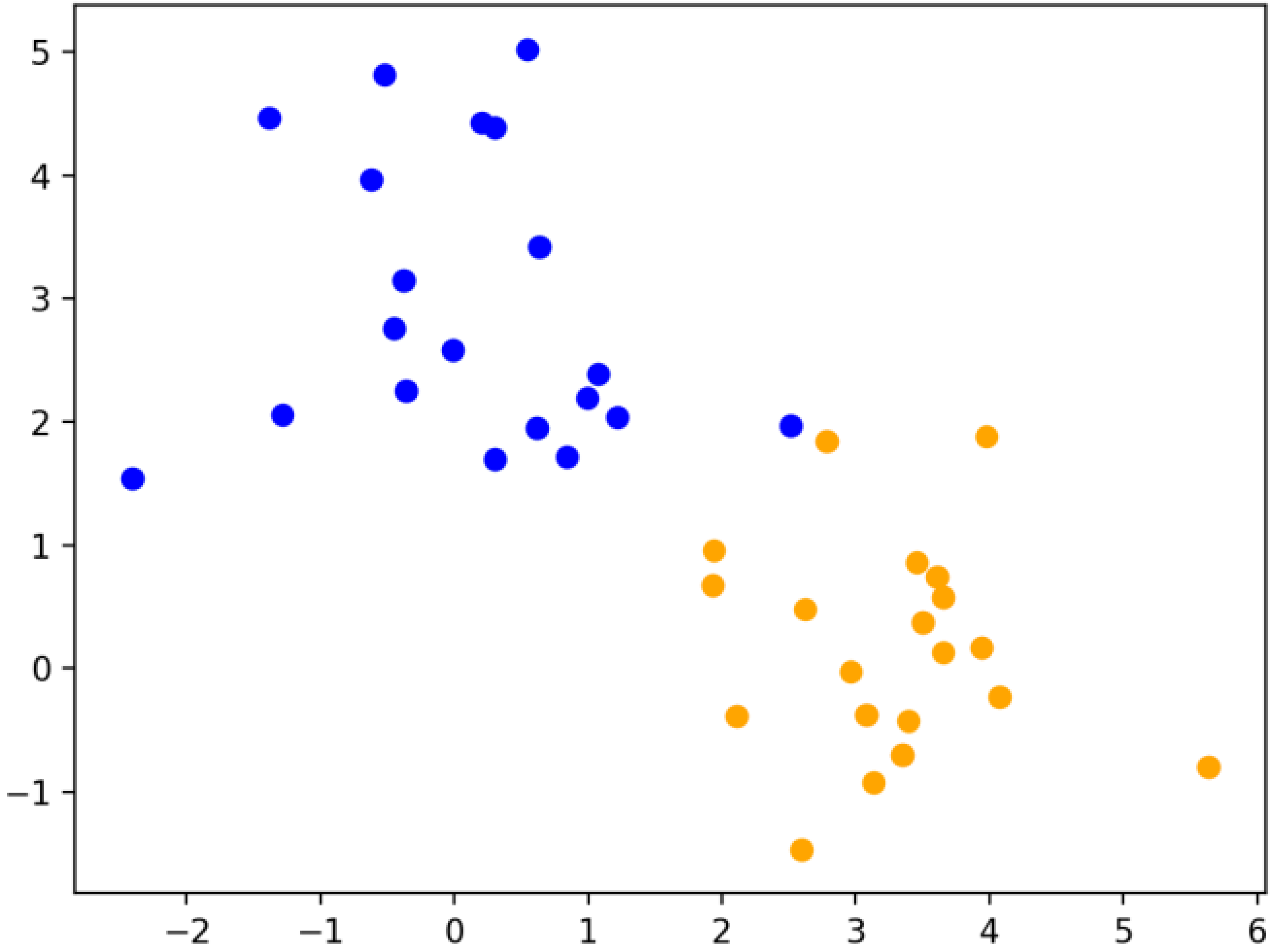


# Online Learning

- 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

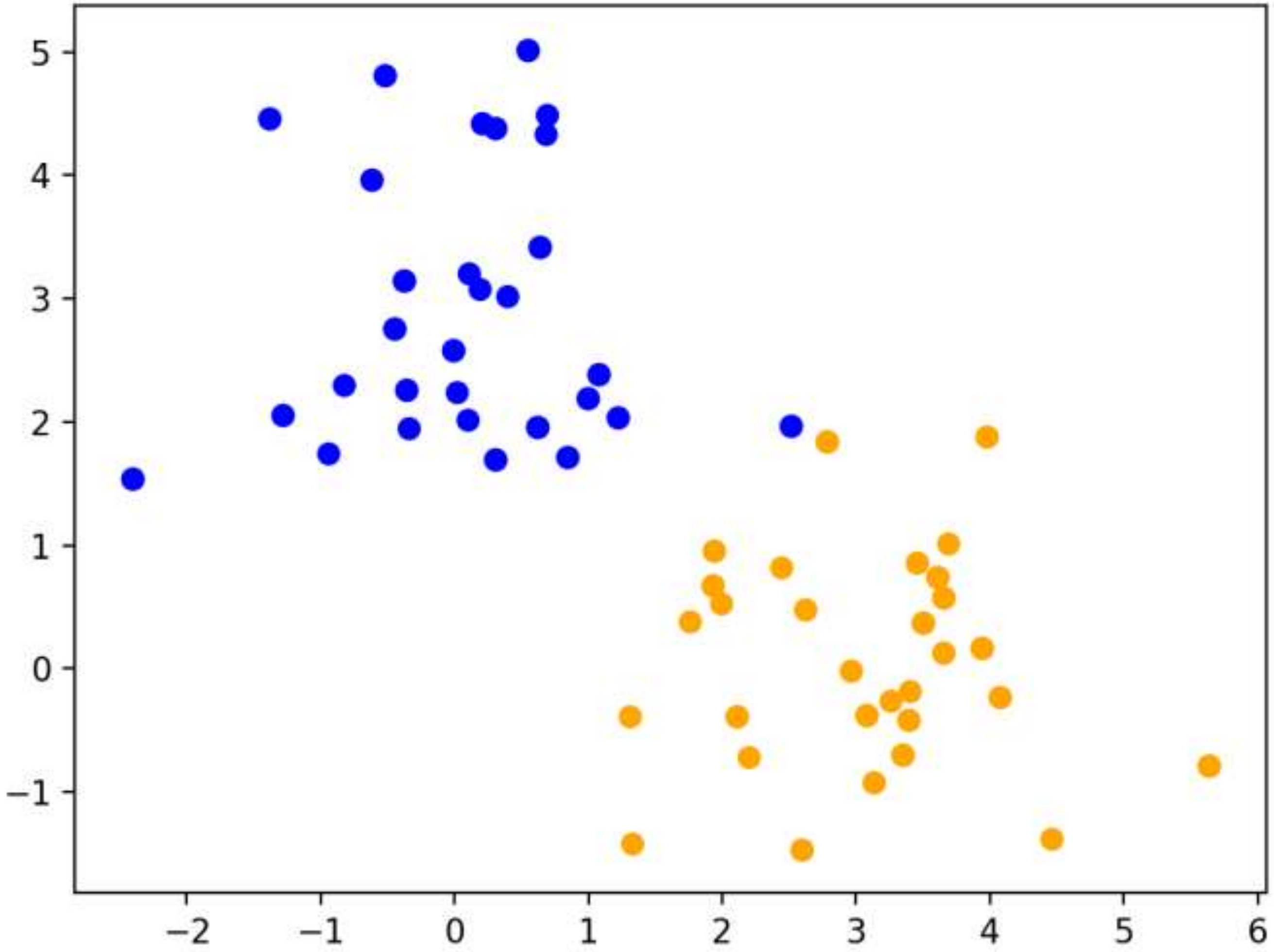


# Online Learning

- 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

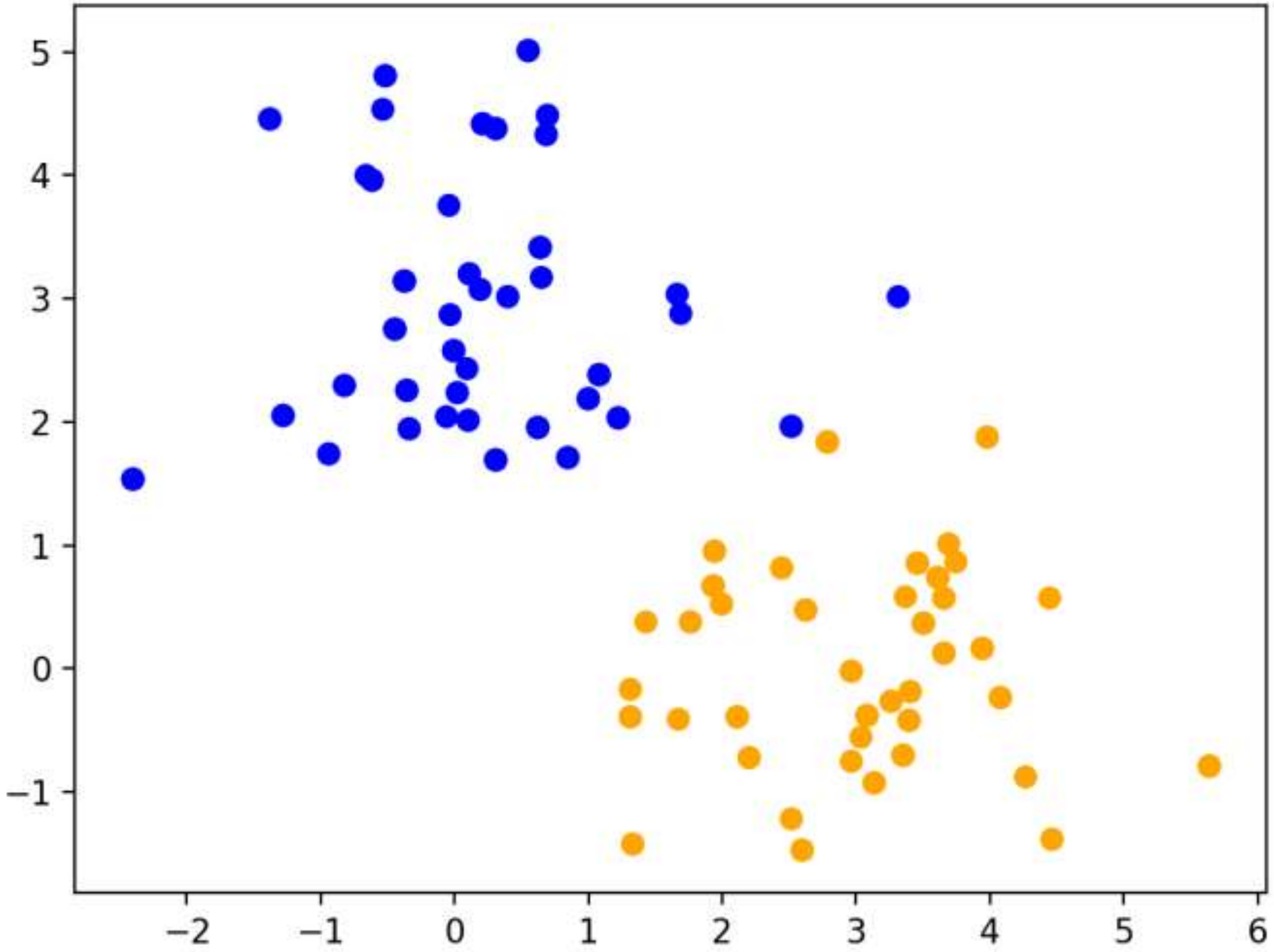


# Online Learning

- 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

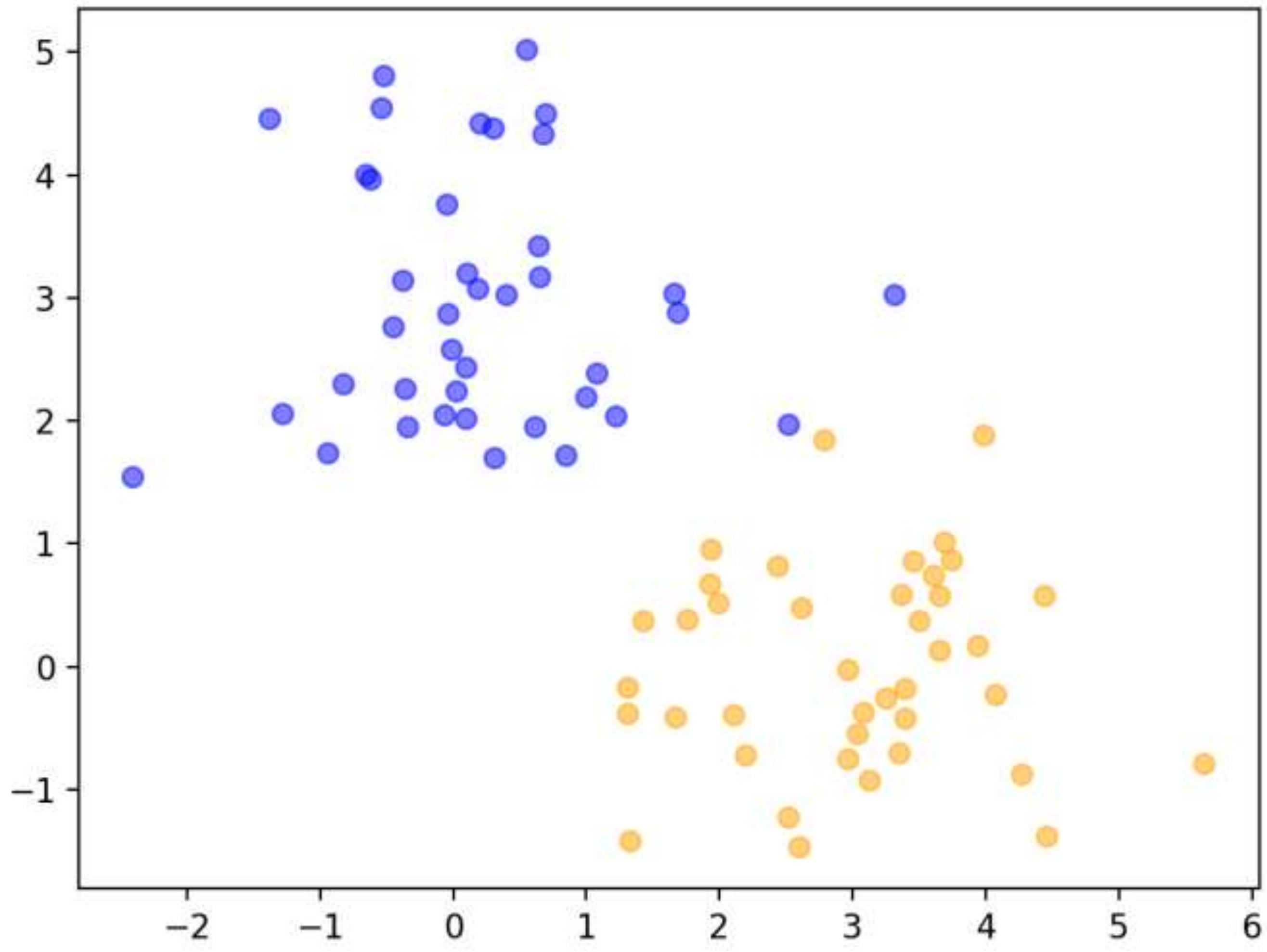


# Lifelong Learning

- Similar to lifelong learning; also uses a learning rate

- Data characteristics change over time

- Continually refine our classifier to adjust to these changing characteristics

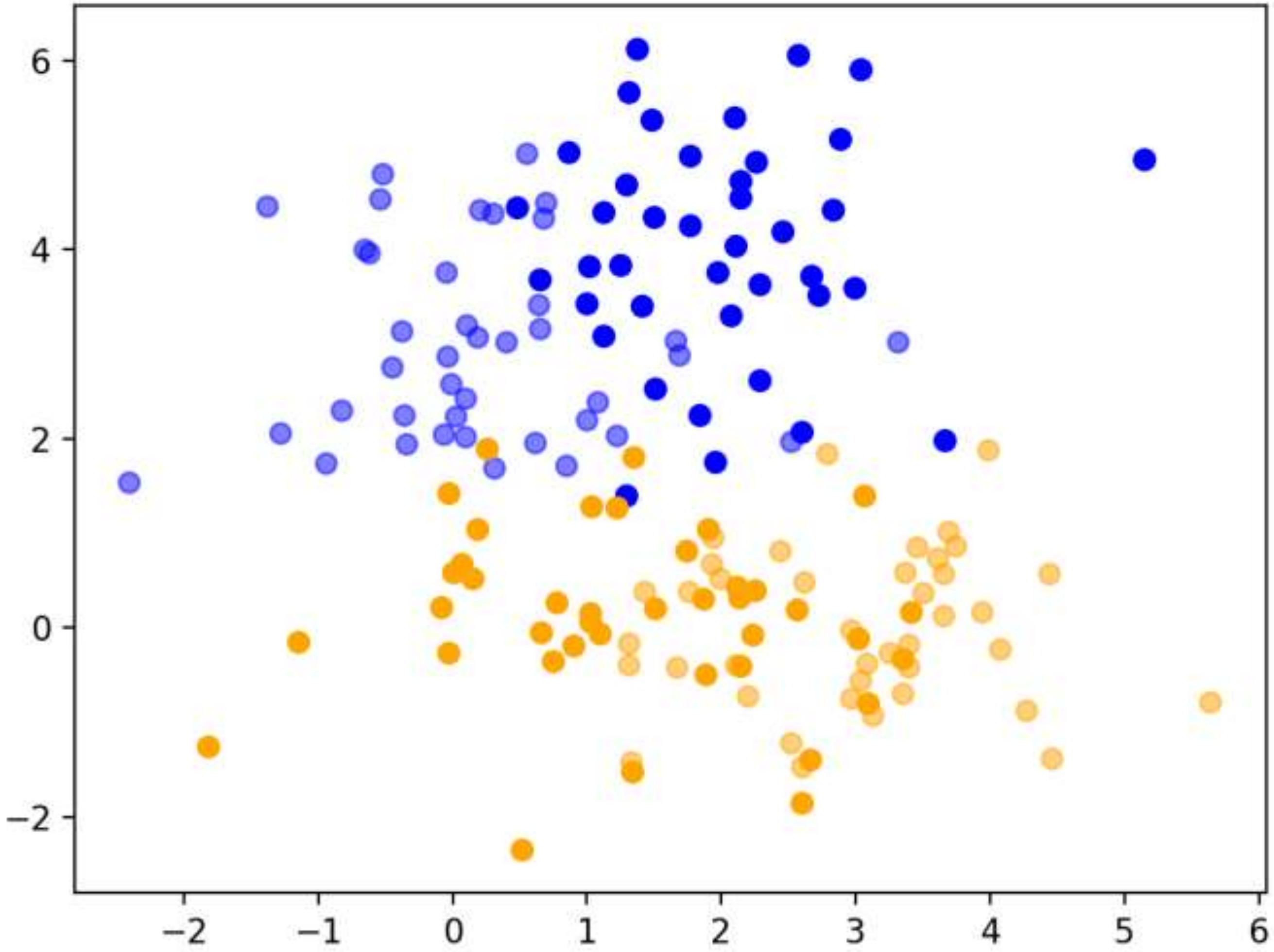


# Lifelong Learning

- Similar to lifelong learning; also uses a learning rate

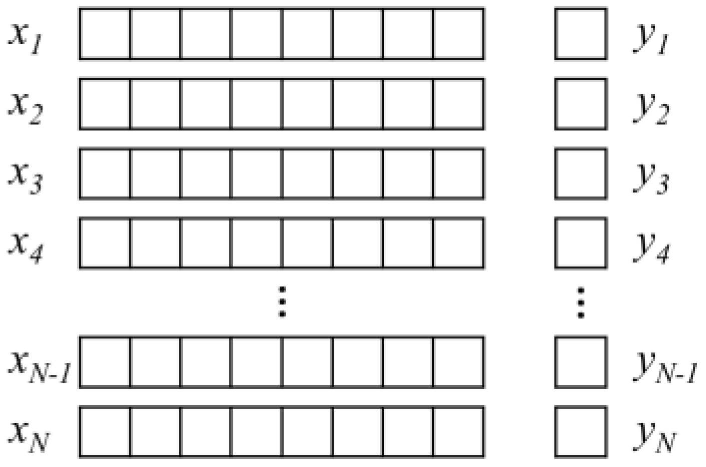
- Data characteristics change over time

- Continually refine our classifier to adjust to these changing characteristics

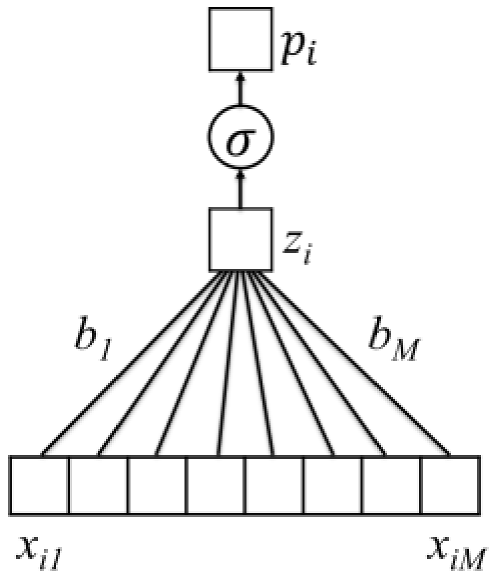


# WE HAVE COVERED A LOT!

# Learning Model Parameters

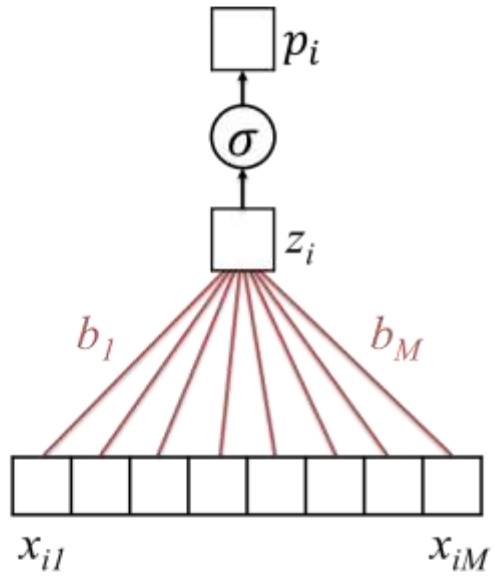



Training Set




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

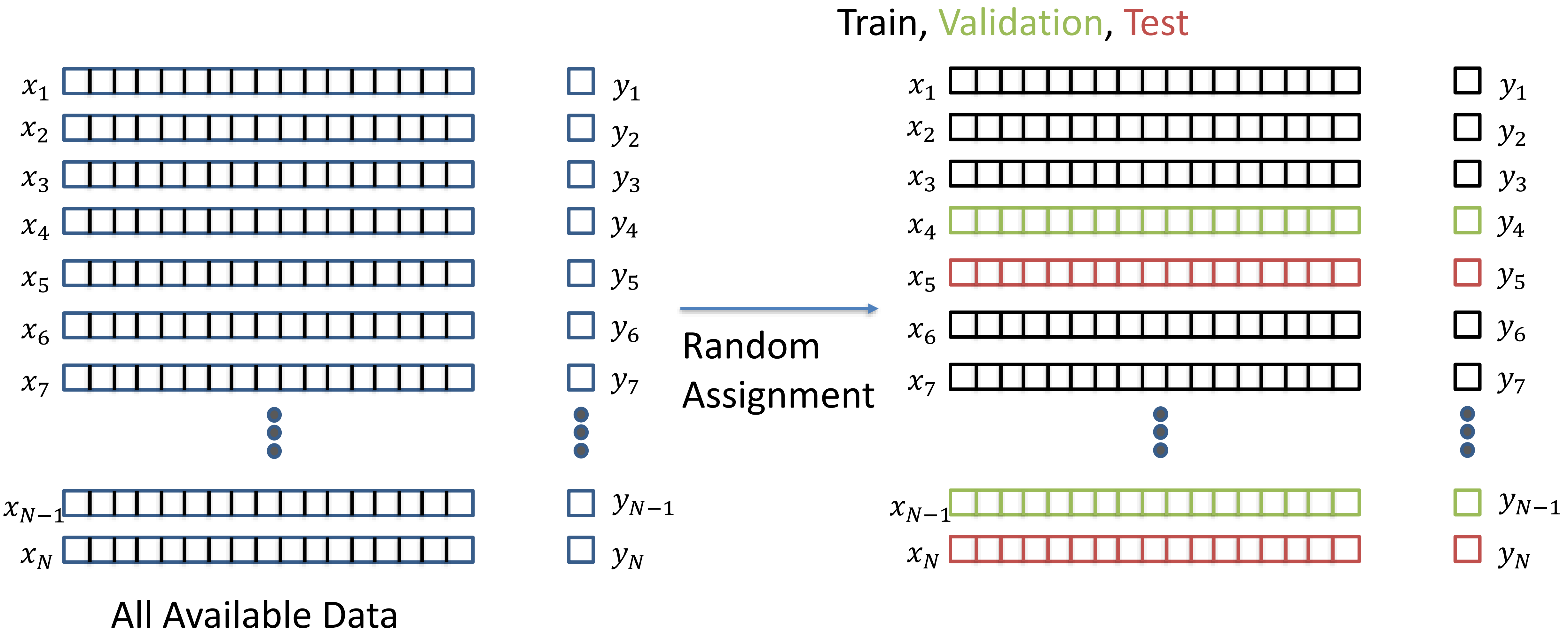
Untrained Logistic Regression Model (or “Network”)



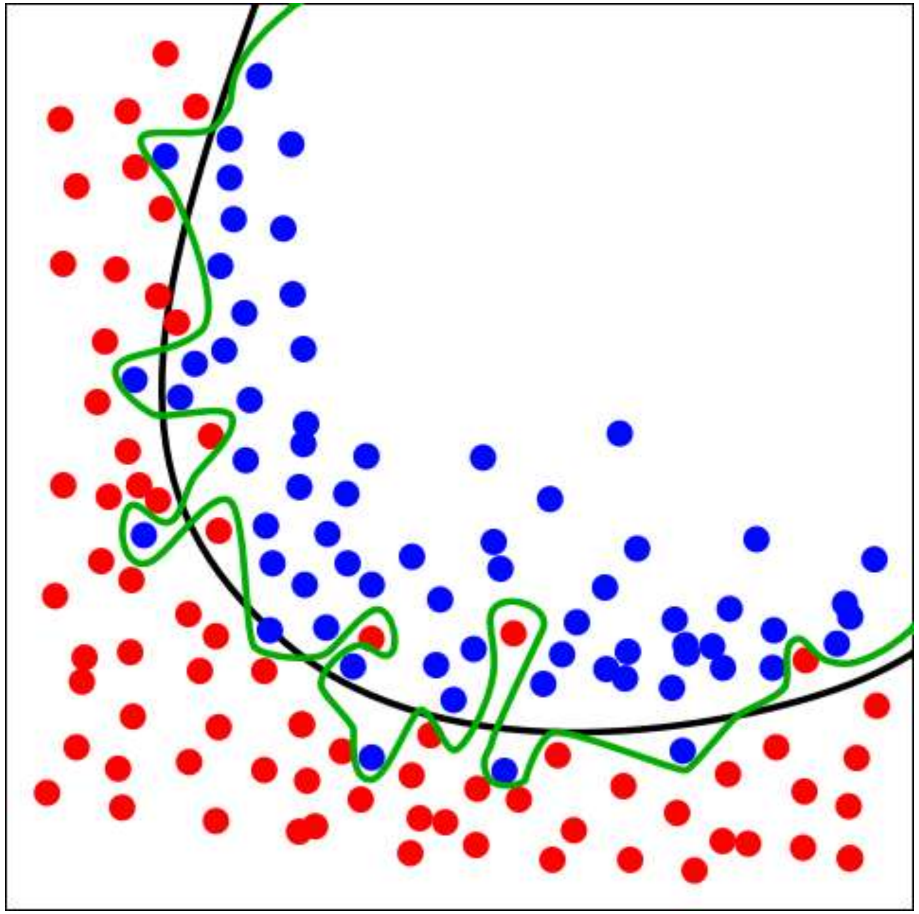

$$b = (b_0, \ldots b_M)$$

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:

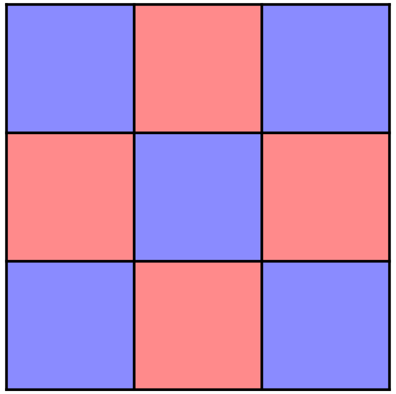
- This is overfitting

Black boundary:

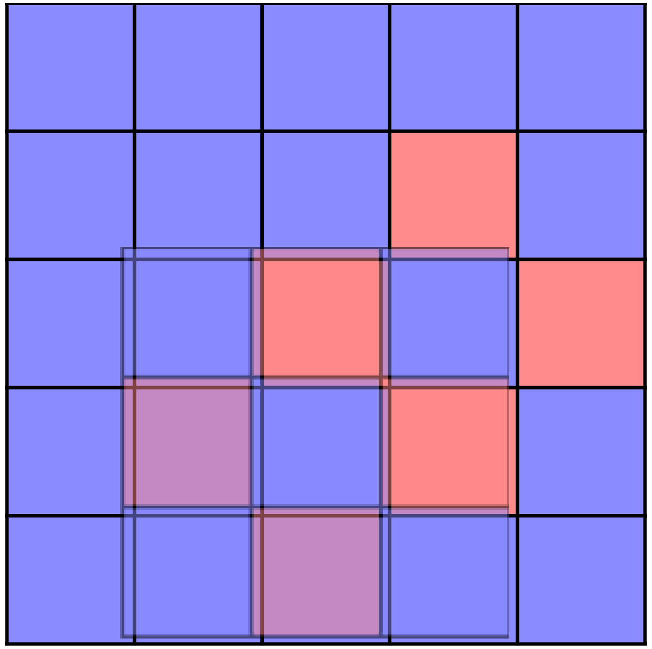
- Balance between fit and model complexity

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

# An Example…



filter

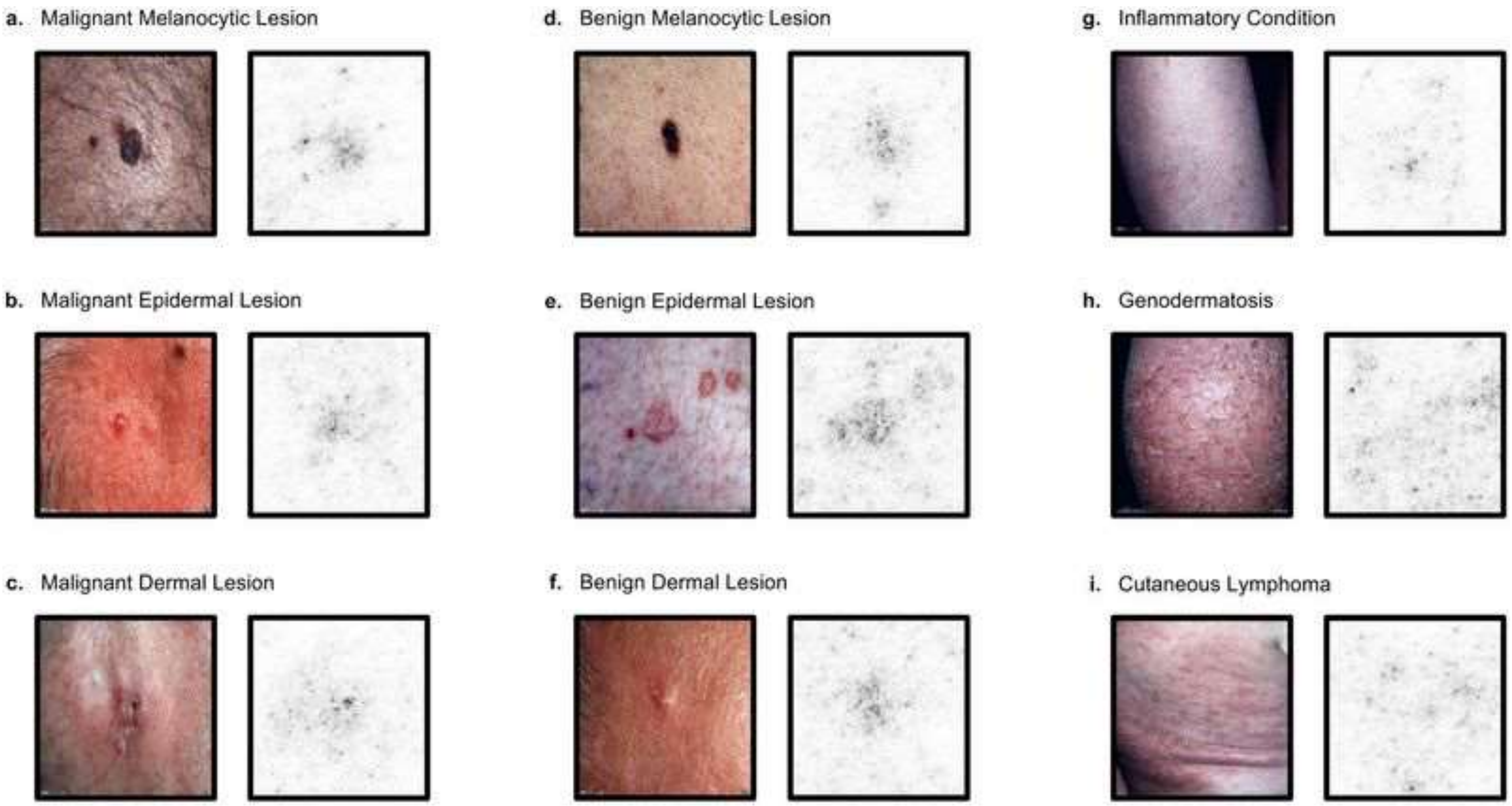


image

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

$x_i^R \odot b$

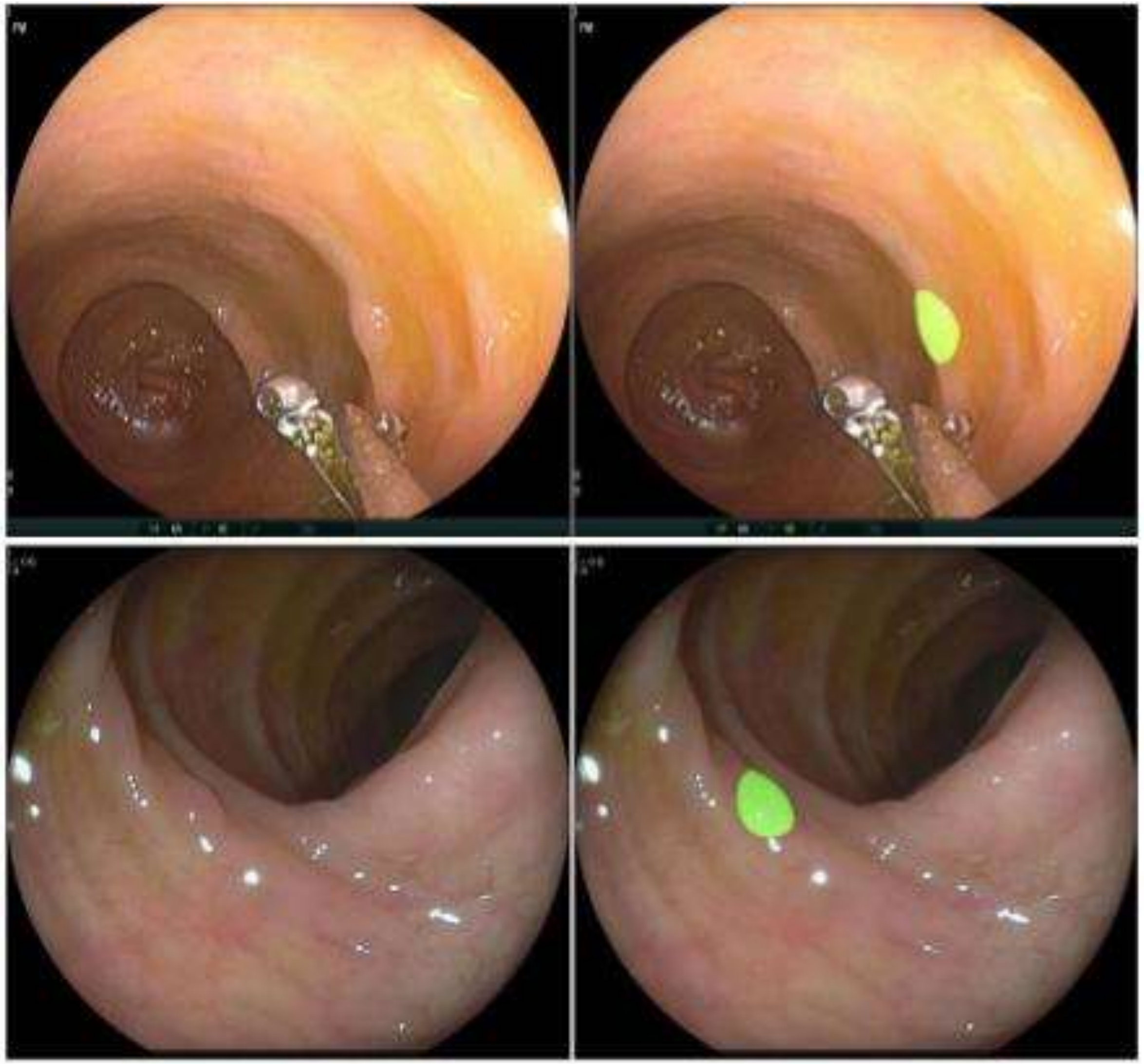
# Saliency maps for example images



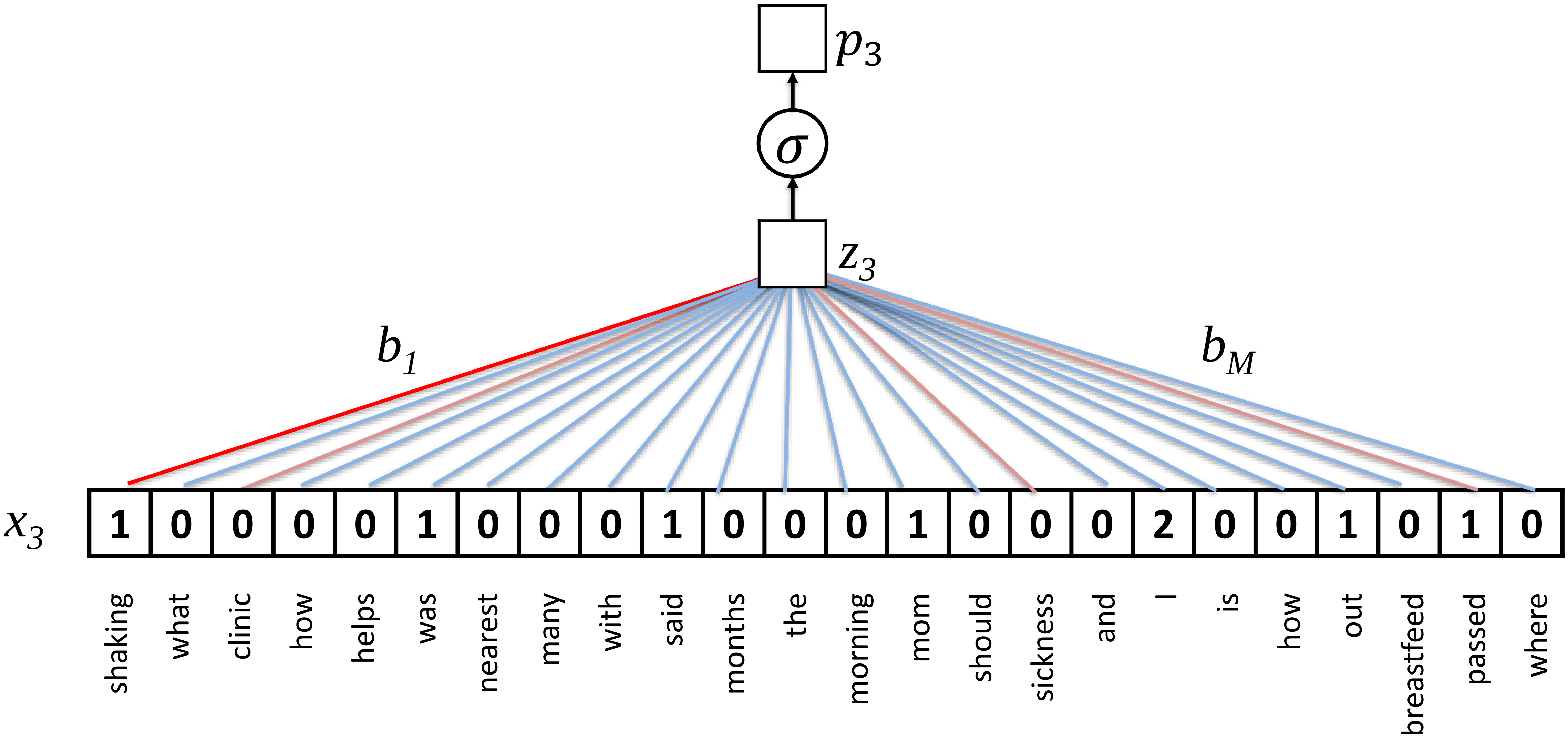

Saliency maps show gradients for each pixel with respect to the CNN's loss function. Darker pixels represent those with more influence.

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

# Precisely Identify Boundaries

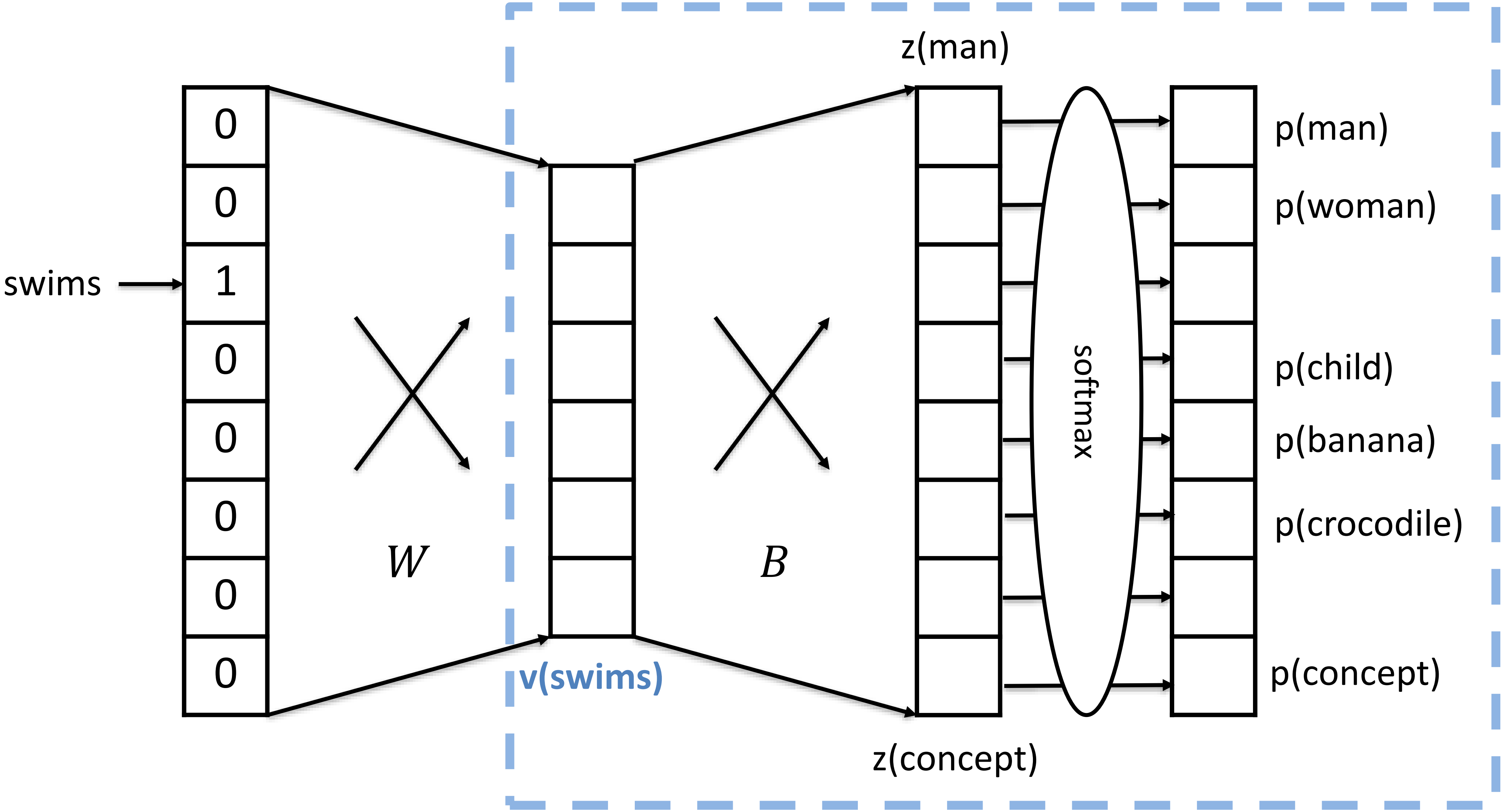


# Logistic Regression for Text Classification

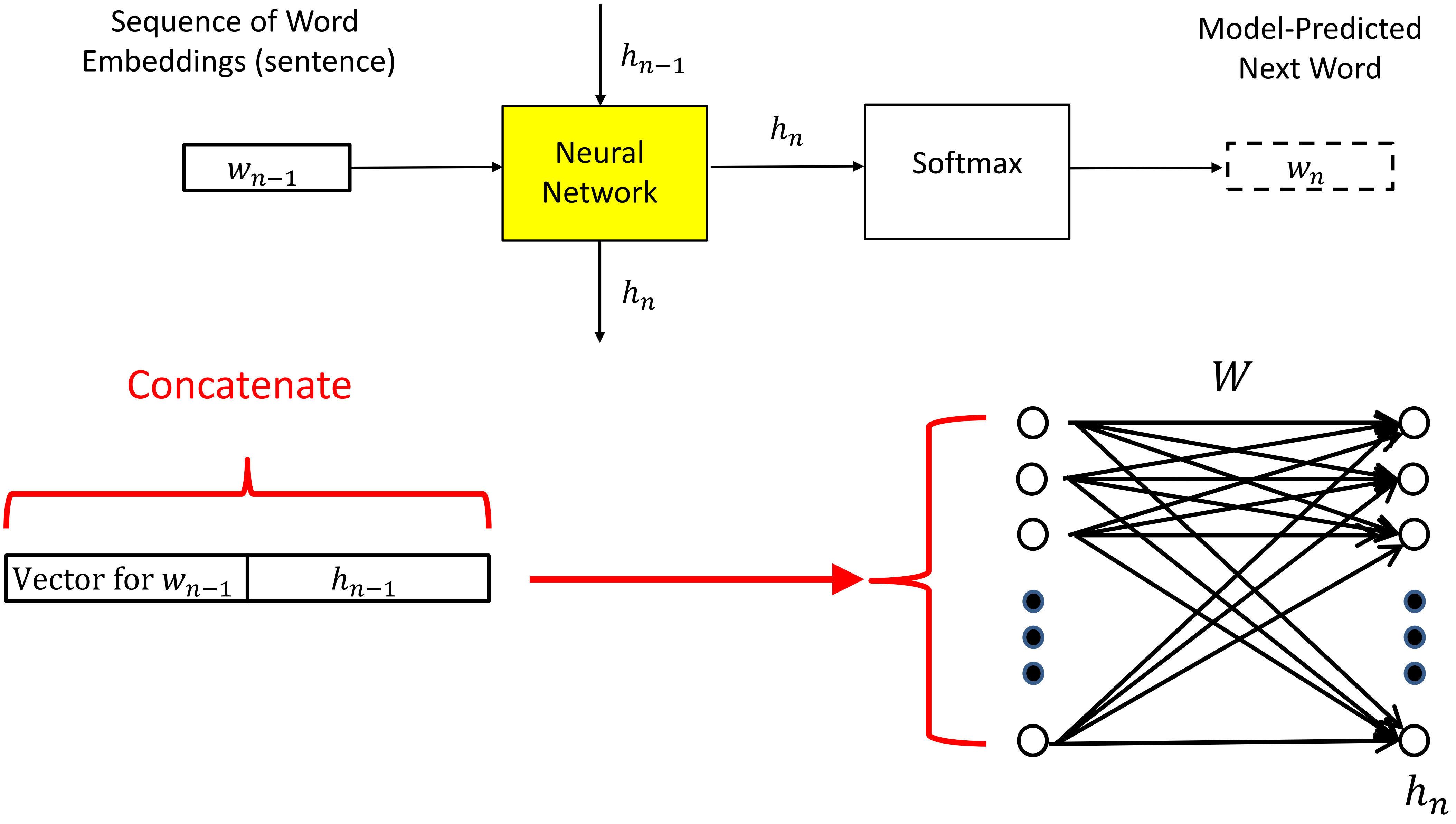



I passed out and Mom said I was shaking

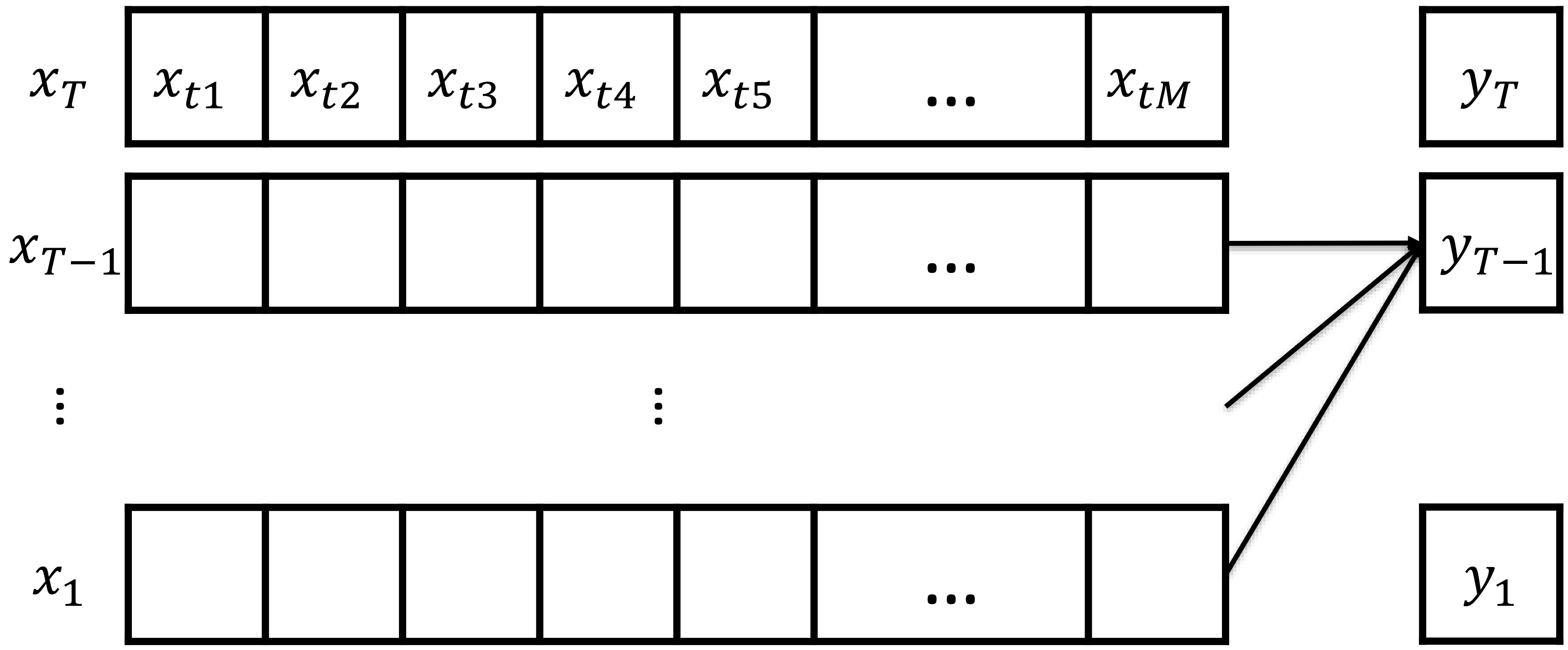
# We now have a distributed representation of word ***meaning*** based on ***context***



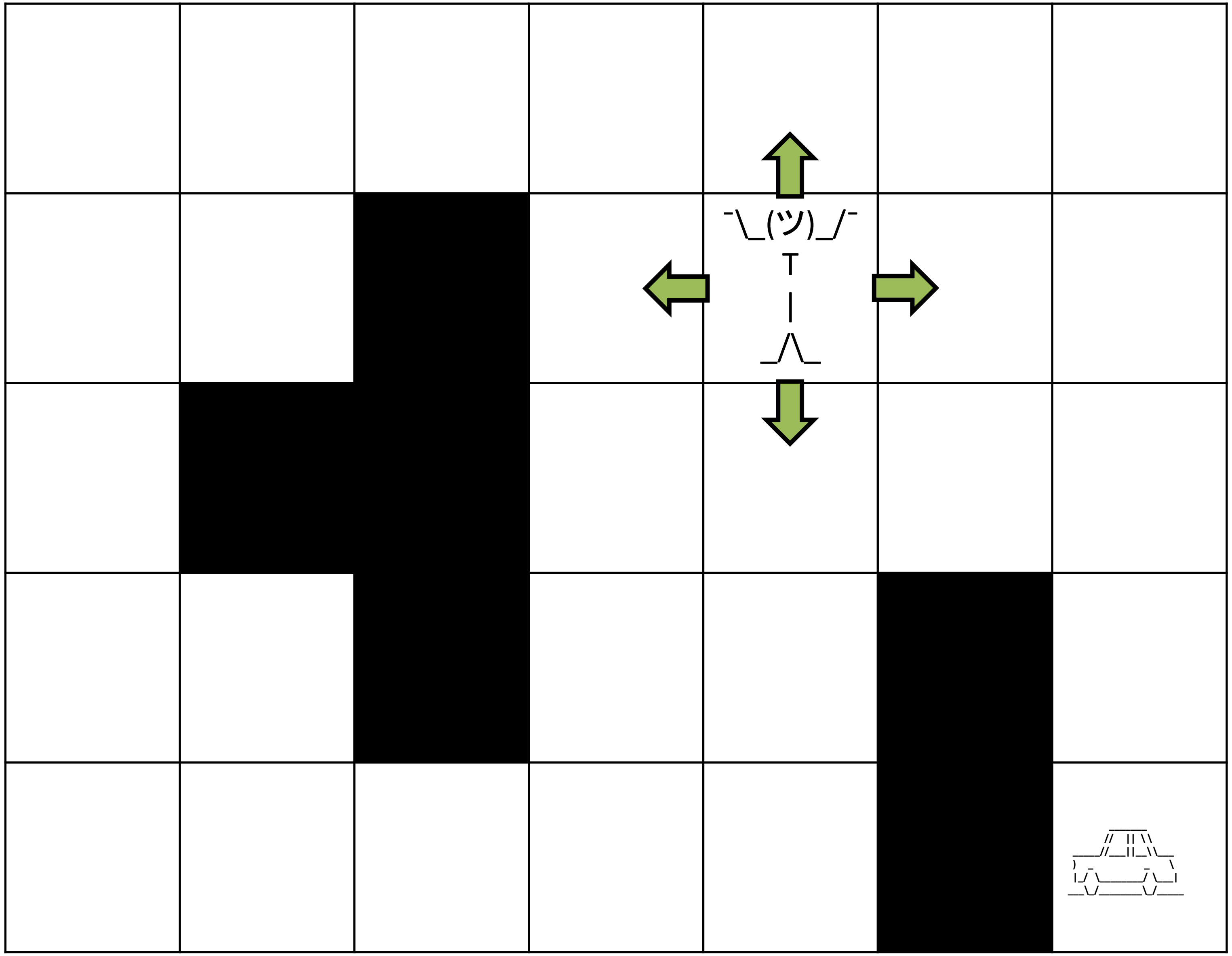
# Recurrent Neural Network



# Prediction task B: one label per time step




Goal: predict whether pt will become hypoxemic <u>during the next 5 minutes</u>




**GOAL:**

Learn a <u>policy</u> $\pi: S \rightarrow A$

that **maximizes expected reward over time**

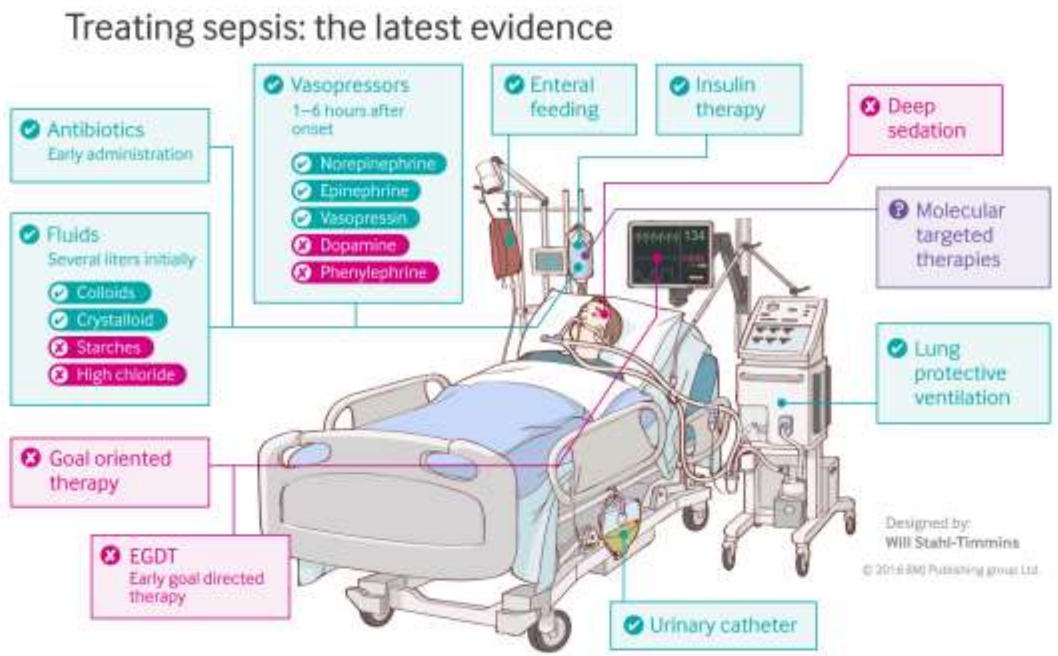
**HOW?**

Learn the value $Q(s, a)$ of action *a* when in state *s*

$$Q(\text{current square}, down)$$

# Sequential Medical Decision-Making:

# Sepsis Management




An agent

takes actions

based on the state of a system

to maximize reward

A clinician

gives fluid and/or vasopressor

based on the patient's physiologic status

to maximize chance of survival

Be in touch: m.engelhard@duke.edu

# THANK YOU!