

NEWARK DS & IT SCHOOL REGRESSION AND MORE

**ZARRAR AND JIUN
FEB 29, 2023**

Overview

- * Recap on data viz, cleaning, etc
- * Regression vs Classification
- * Recap on linear regression
- * More complex models
- * Cool applications

Overview

- * **Recap on data viz, cleaning, etc**
- * Regression vs Classification
- * Recap on linear regression
- * More complex models
- * Cool applications

RECAP ON DATA VIZ, CLEANING, ETC

<https://tinyurl.com/45e9xwfd>

Questions

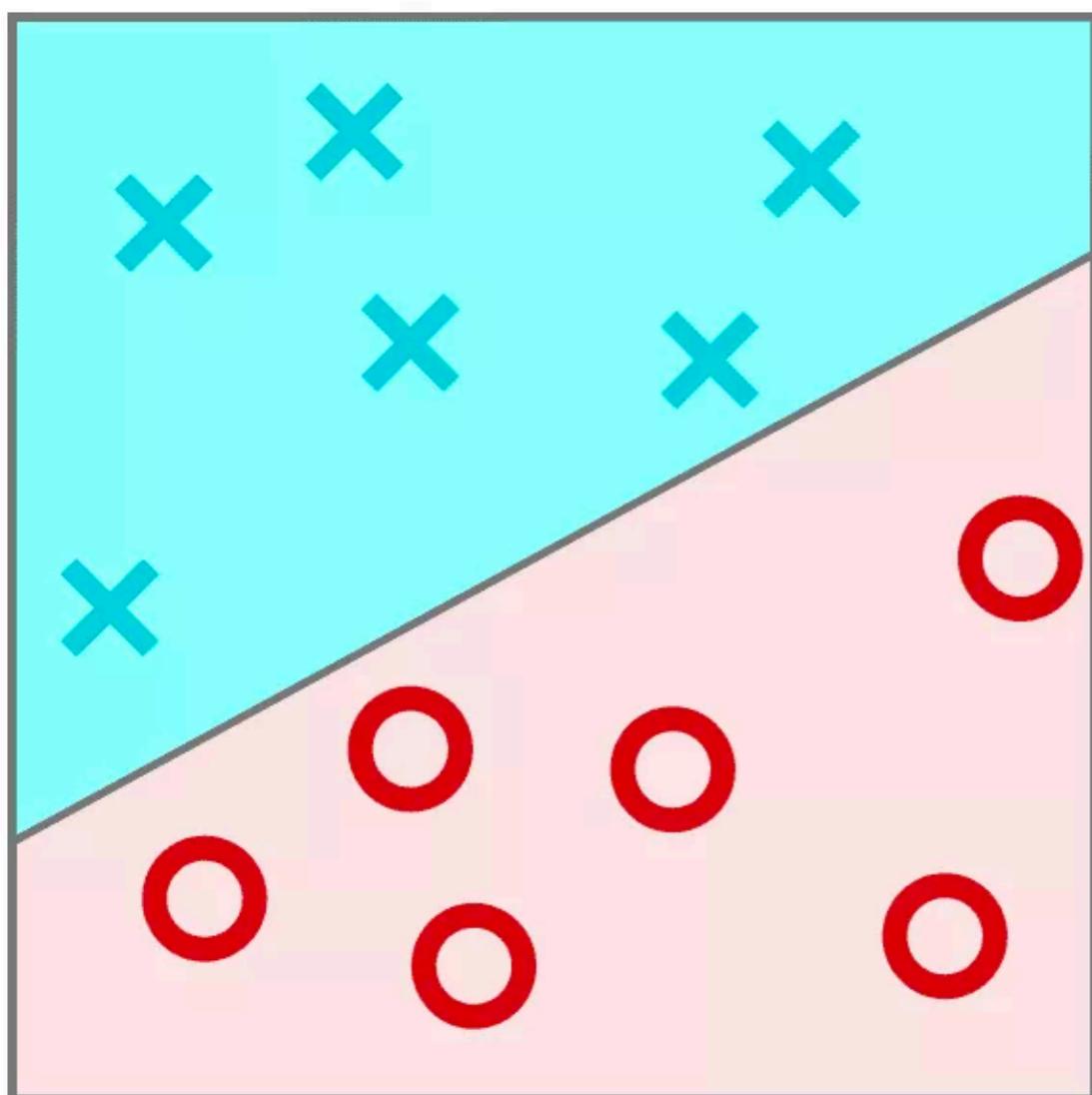
- * Were you surprised by how little code is actually for predicting the data?

Overview

- * Recap on data viz, cleaning, etc
- * **Regression vs Classification**
- * Recap on linear regression
- * More complex models
- * Cool applications

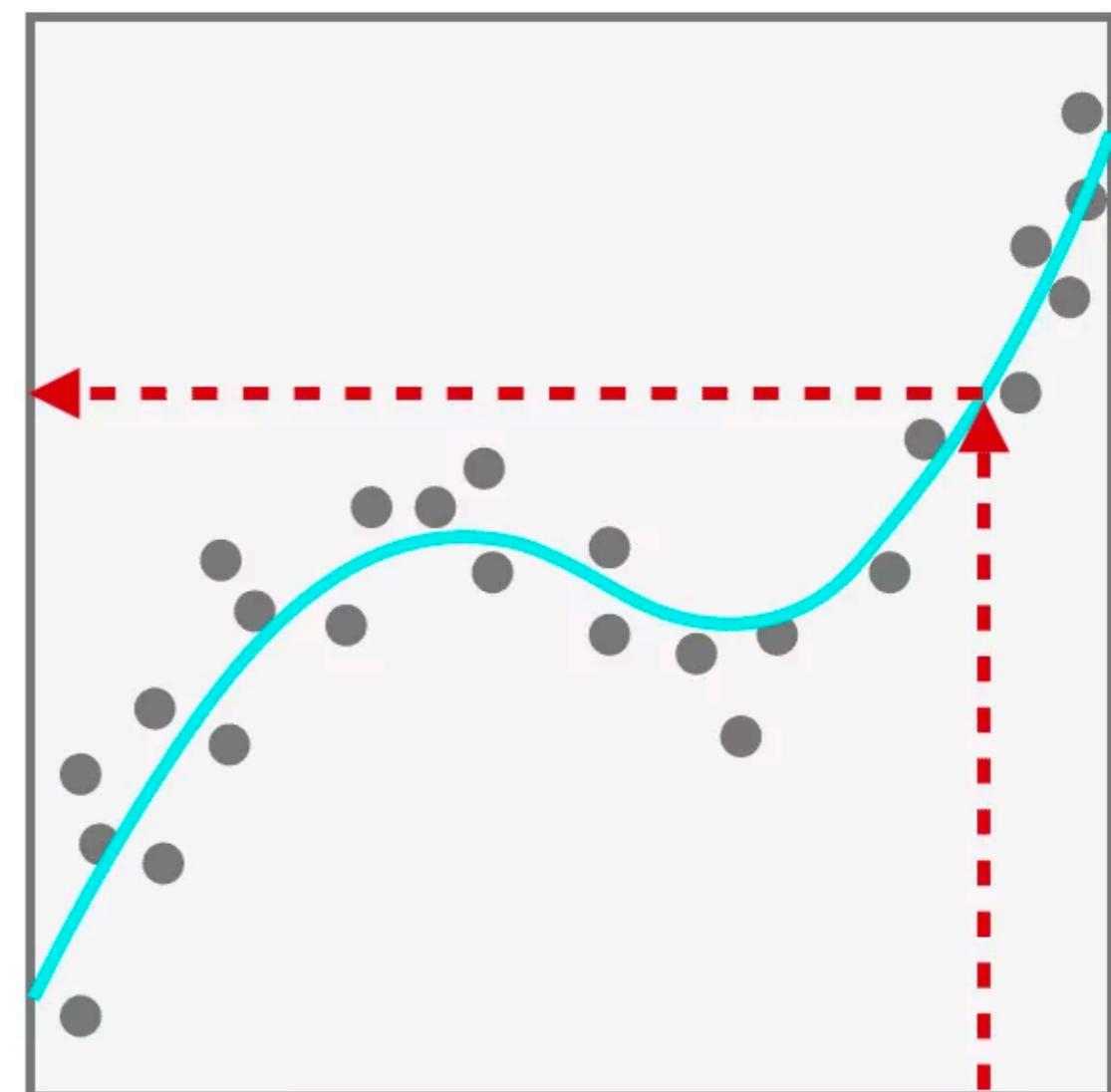
REGRESSION VS CLASSIFICATION?

Classification Groups observations into "classes"



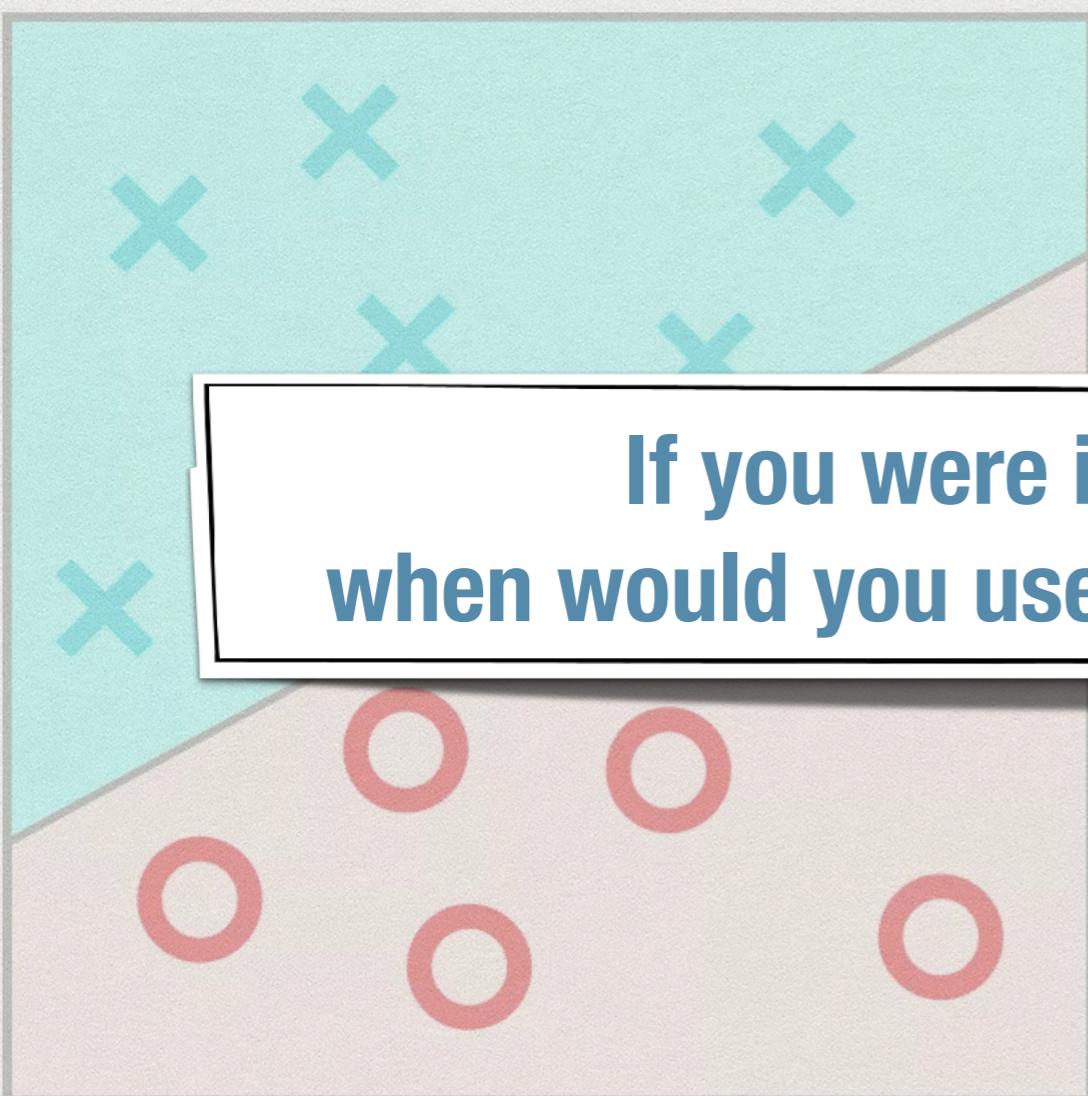
Here, the line classifies the observations into X's and O's

Regression predicts a numeric value



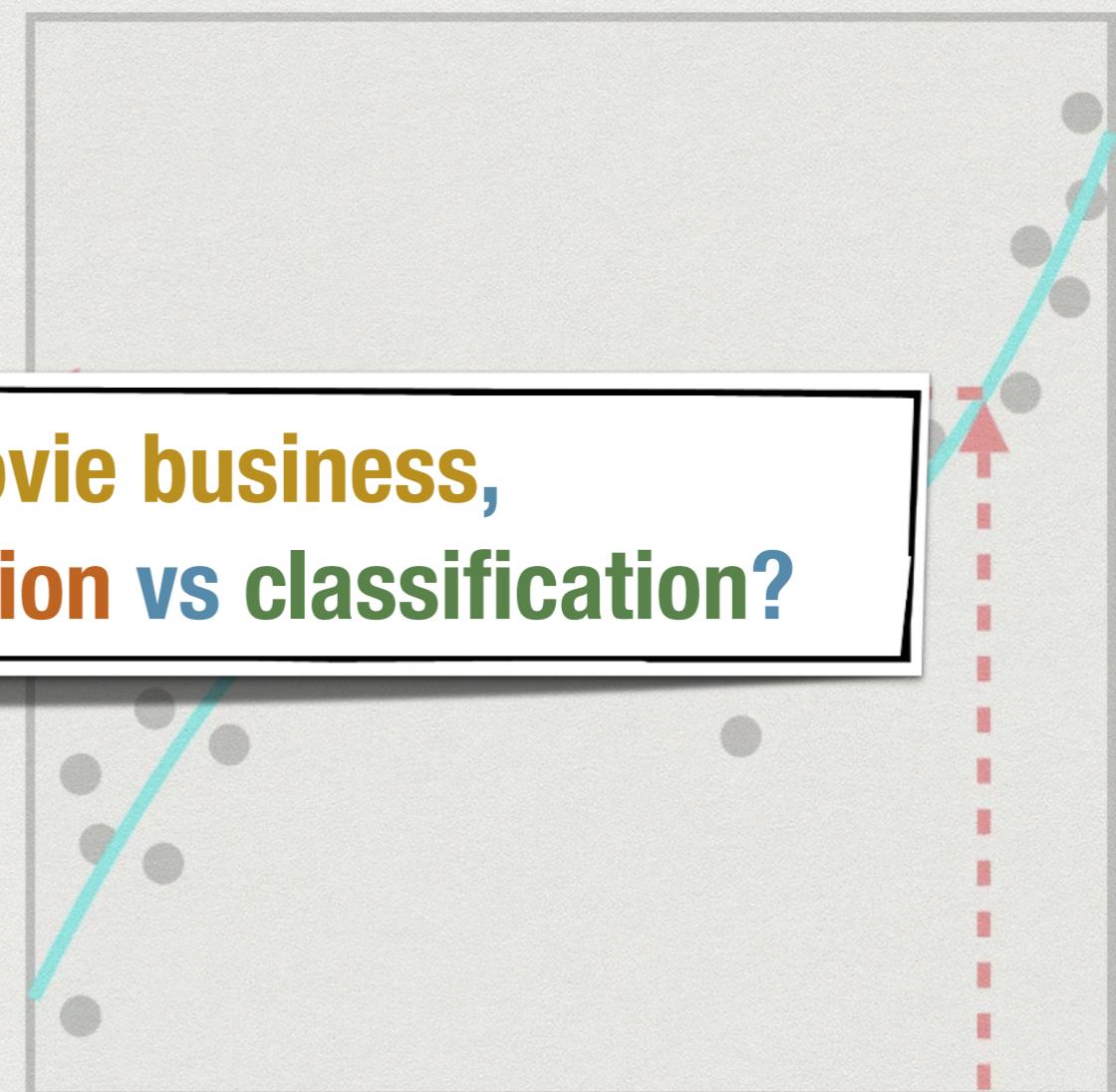
Here, the fitted line provides a predicted output, if we give it an input

Classification Groups observations into "classes"



Here, the line classifies the observations into X's and O's

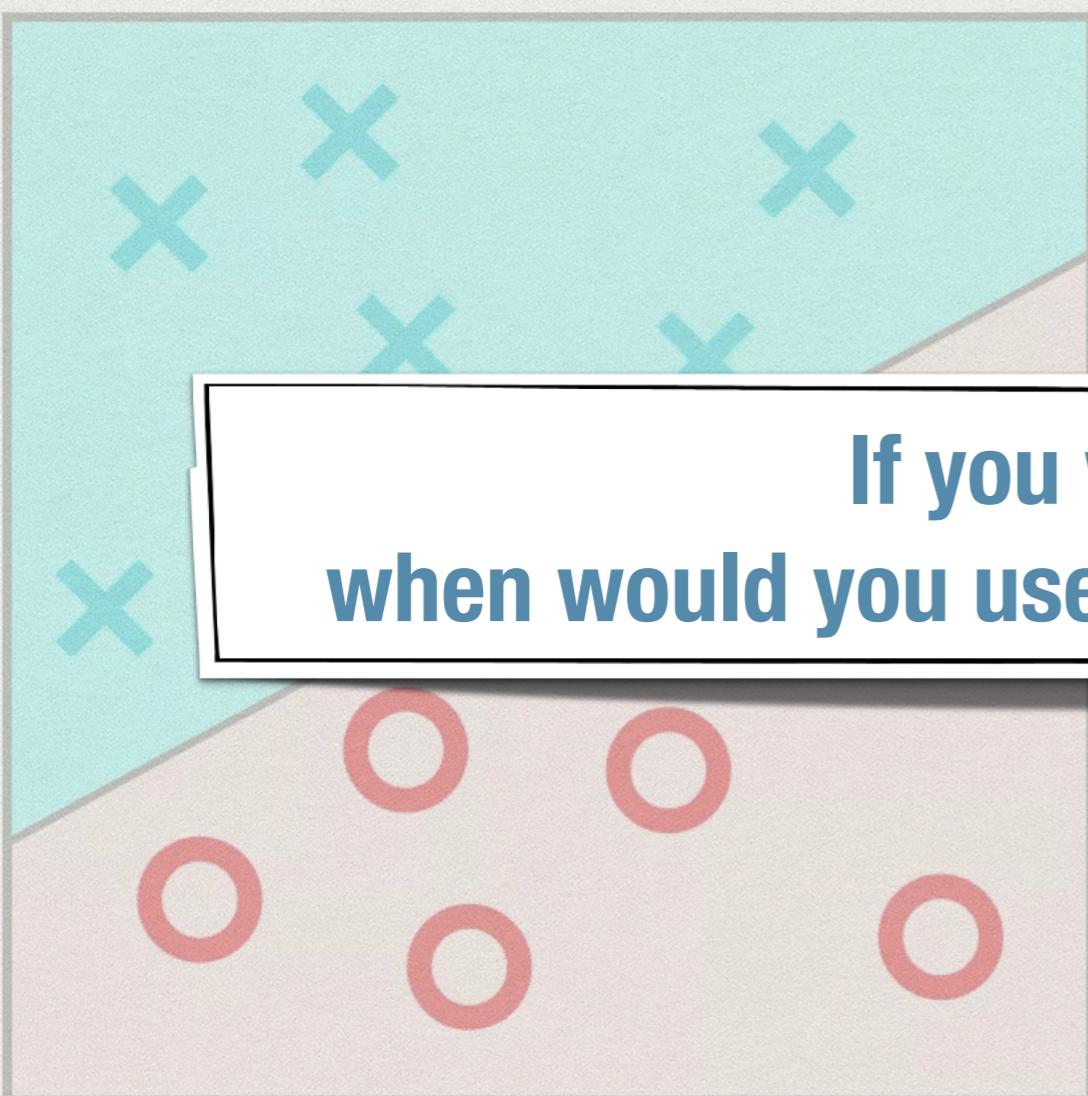
Regression predicts a numeric value



Here, the fitted line provides a predicted output, if we give it an input

If you were in the **movie business**,
when would you use **regression vs classification?**

Classification Groups observations into "classes"



Here, the line classifies the observations into X's and O's

Regression predicts a numeric value



Here, the fitted line provides a predicted output, if we give it an input

If you were at Audible,
when would you use **regression** vs **classification**?

Overview

- * Recap on data viz, cleaning, etc
- * Regression vs Classification
- * **Recap on linear regression**
- * More complex models
- * Cool applications

RECAP ON LINEAR REGRESSION

<https://audible-newark-ds-school.streamlit.app/>

Questions

- * Why is it important to predict a movie's revenue?
- * Why does the linear regression give a best fit line?
- * How might outliers effect the linear regression fit?
 - * How could you fix the problem of outliers?
- * Should you have the most complicated model?

Overview

- * Recap on data viz, cleaning, etc
- * Regression vs Classification
- * Recap on linear regression
- * **More complex models**
- * Cool applications

MORE COMPLEX MODELS

Too much data & features!!!

- * What's the problem to building complex models?

Too much data & features!!!

- * Imagine you are optimizing the search experience for **Audible** customers.
- * You might have 100s of features
Can you think of some examples?
- * You have millions of search results each month
Can you hold this data in memory?
- * *Hard to even visualize all this data*

Too much data & features!!!

- * We need to deal with highly correlated features
- * Data often so large that can't be held in memory

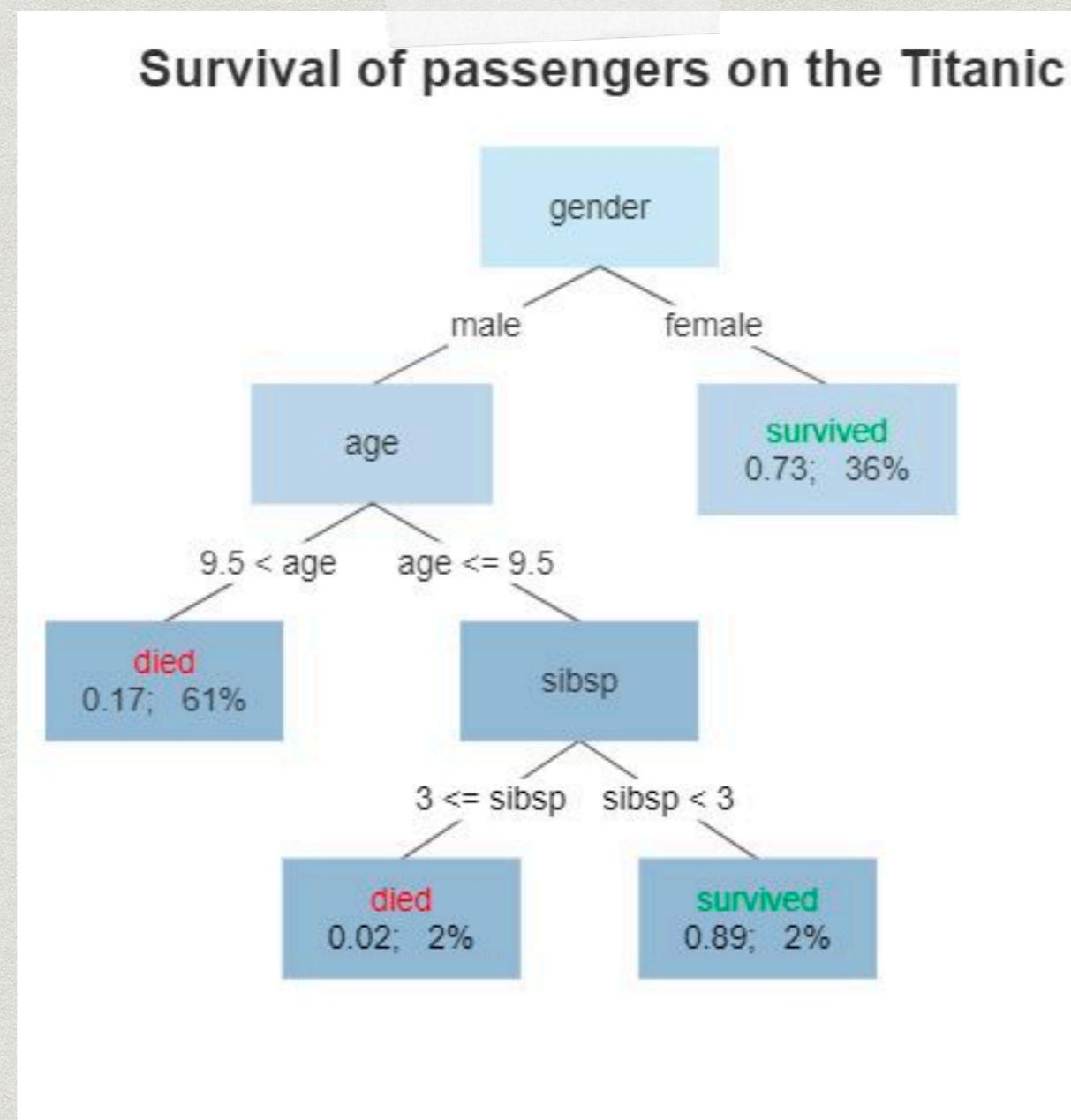
Models for a complex world

- * Last week discussed Ridge and Lasso
- * Decision Trees
- * Neural Networks

Decision Trees

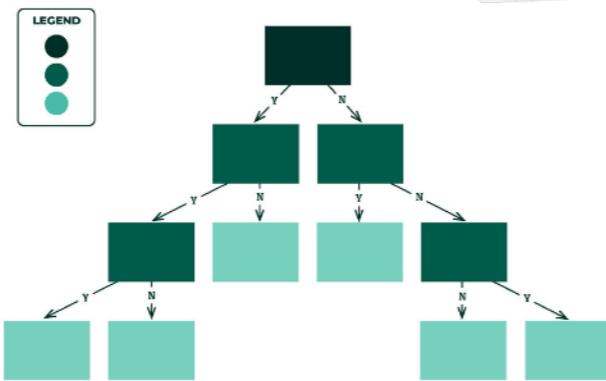
- * We use them at Audible to help rank the most relevant books to our customers!
- * Try a search on <https://www.audible.com>

Example Decision Tree

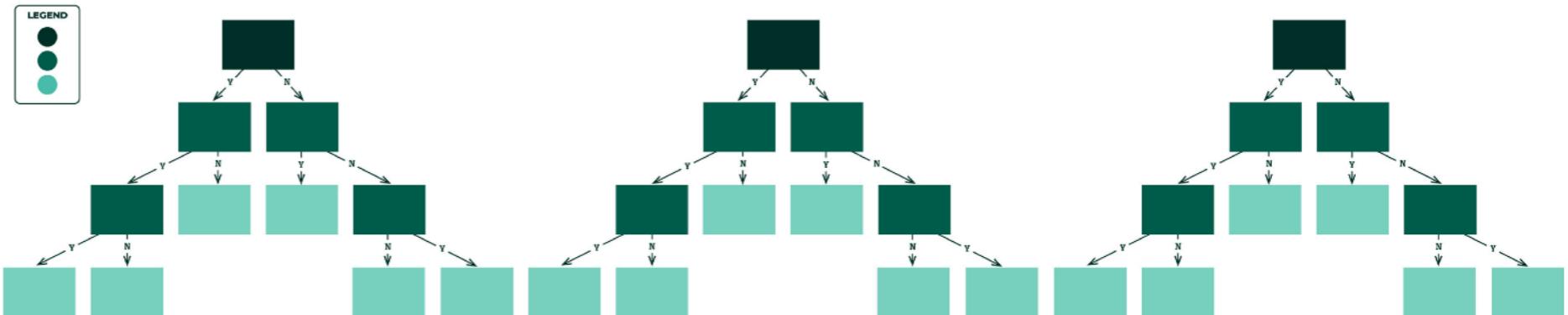


Scaling the Decision Tree

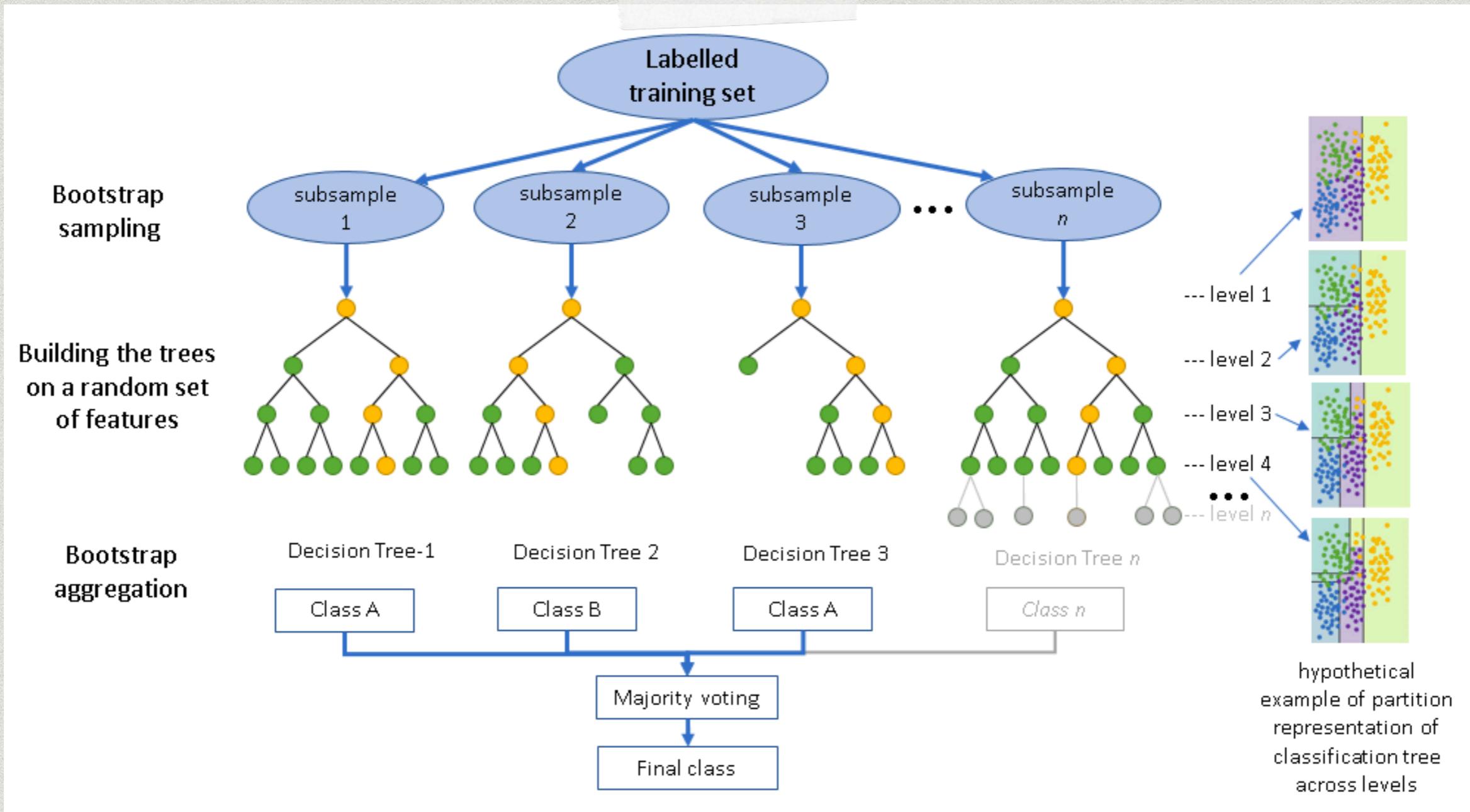
**DECISION
TREE**



**RANDOM
FOREST**



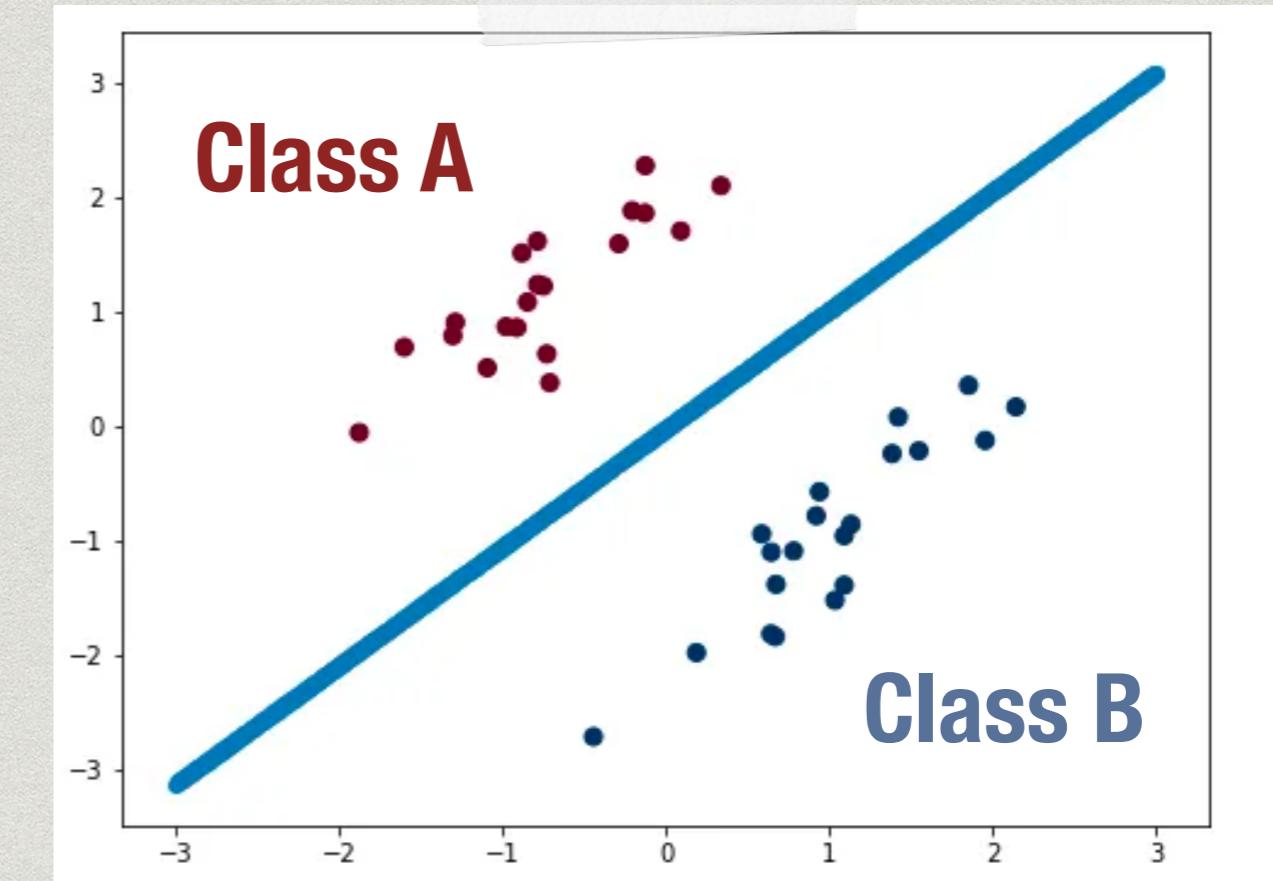
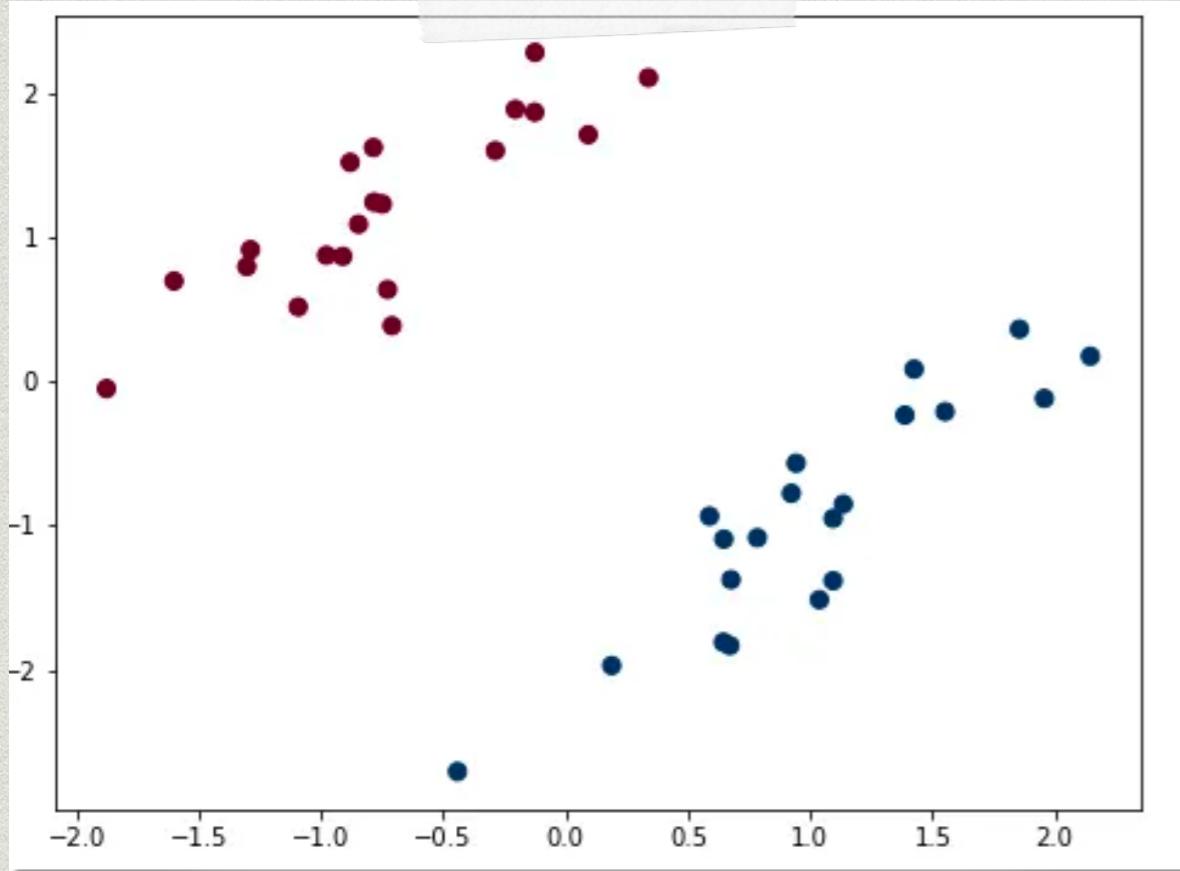
Scaling the Decision Tree



Neural Networks

- * We use these at **Audible** in many domains like recommendations or enhanced autocomplete
- * Other examples in face recognition, self-driving cars, conversational AI like ChatGPT.

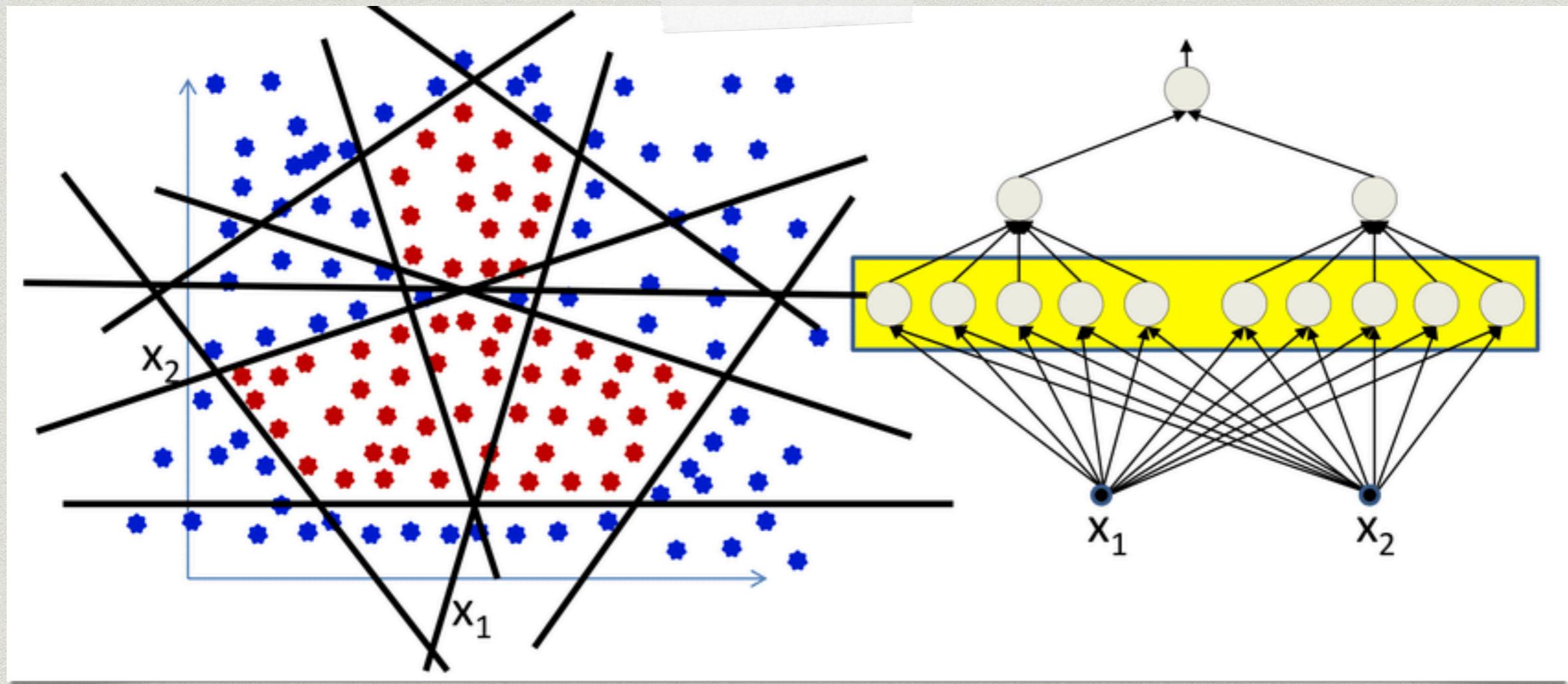
Intuition: Remember Regression



Here we are using the line to separate points into classes

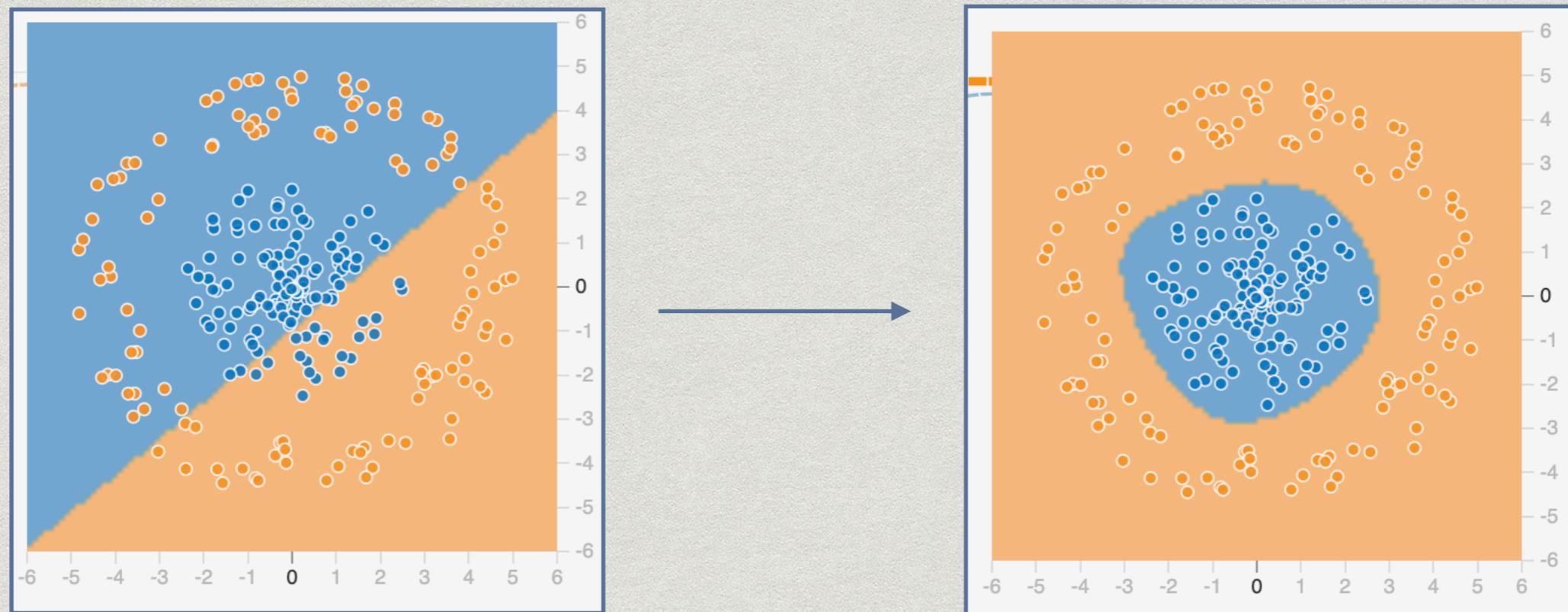
Think of how we do this categorization in our life?

Regression on Steroids



Example

- * <https://playground.tensorflow.org/>

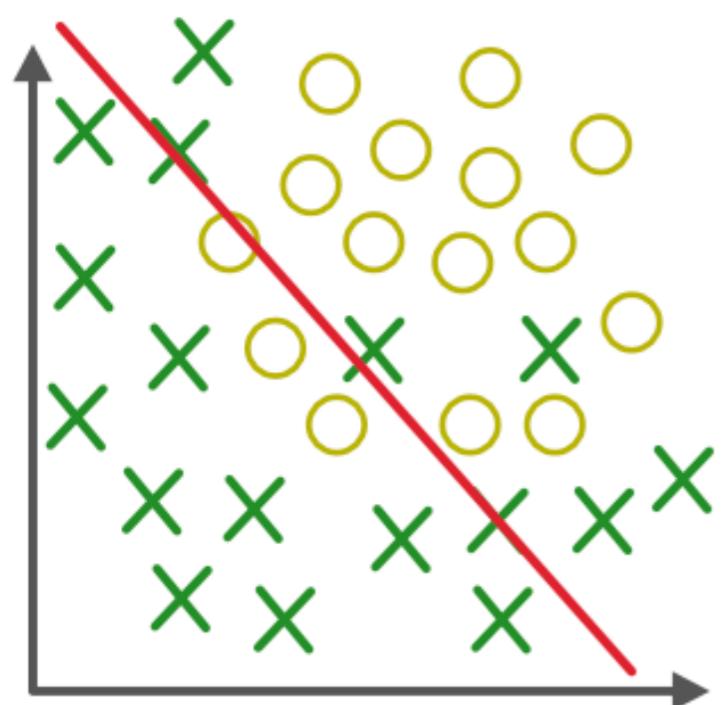


Why not build the most complex model?

Why not build the most complex model?

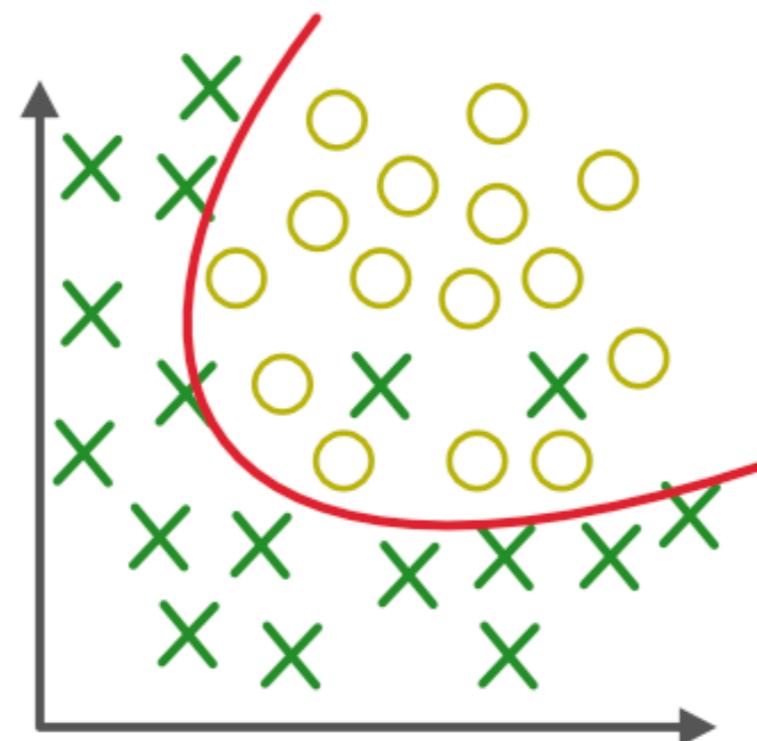
- * Complex models are more expensive \$\$\$
- * Complex models do not generalize ???
 - * This means they do well on the data they are trained on but not on new data.
 - * *Imagine you trained Audible data for only Christmas, would this generalize to the summer?*

Example

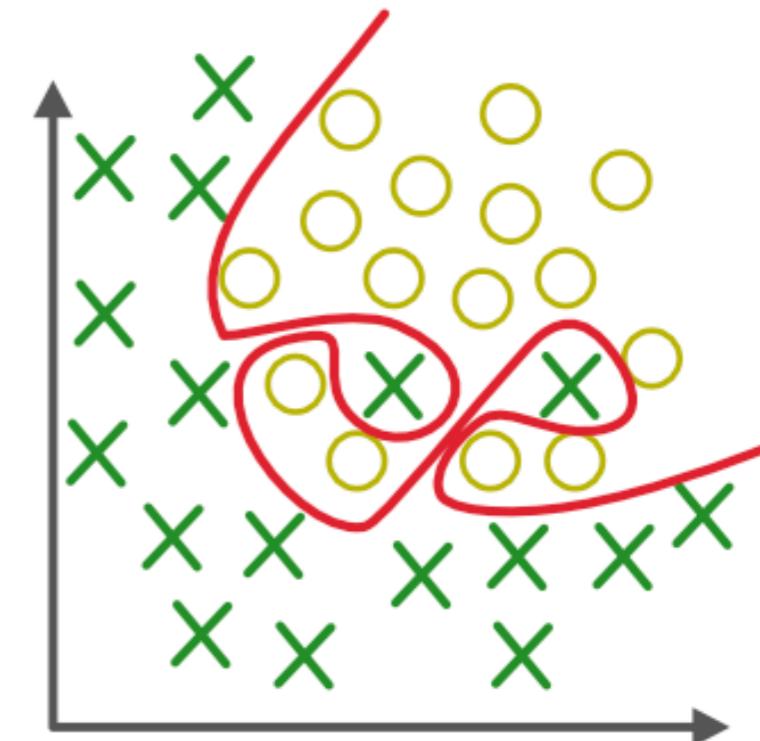


Under-fitting

(too simple to explain the variance)



Appropriate-fitting



Over-fitting

(forcefitting--too good to be true)

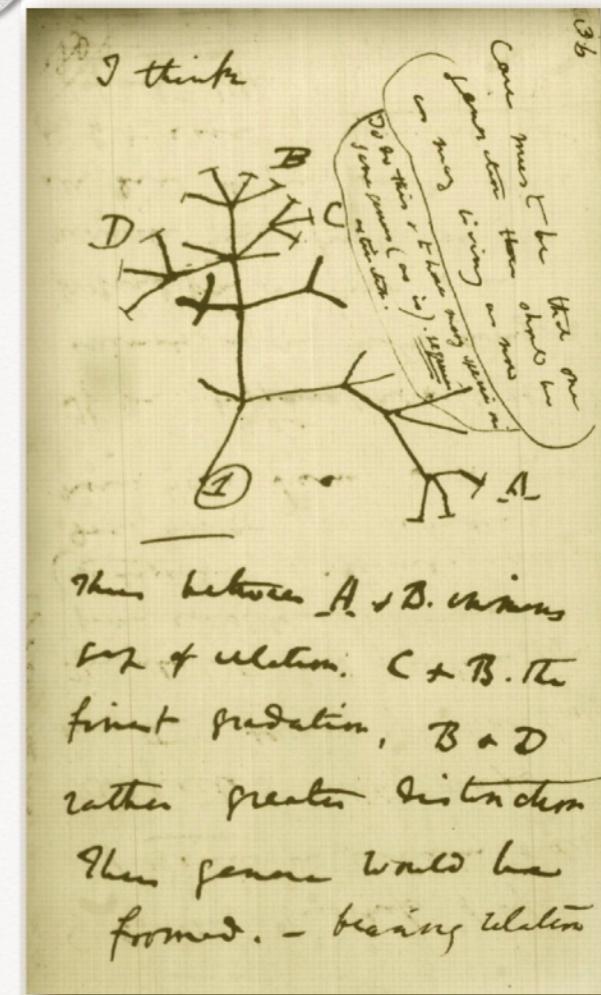
Example: Charles Darwin

Algorithms to Live By



What Computers
Can Teach Us About
Solving Human Problems

Brian Christian and Tom Griffiths



Had an 'early stopping' procedure to make decisions

Overview

- * Recap on data viz, cleaning, etc
- * Regression vs Classification
- * Recap on linear regression
- * More complex models
- * **Cool applications**

COOL APPLICATIONS

Cool Applications

- * Taking text and generating an image
(Stable Diffusion)
- * Conversational AI
(AlexaTM or ChatGPT)

To Learn More

- * Gentle Introduction to Machine Learning
https://www.youtube.com/watch?v=Gv9_4yMFhl
- * ChatGPT
<https://chat.openai.com/chat>
- * Generate image from text
<https://huggingface.co/spaces/dalle-mini/dalle-mini>

THANK YOU!

ANY QUESTIONS?