3.22.21

Presentation Day 2

Quick Remaining Agenda

3/29 - Project 2 Begins

20% HW Review, Presentation Discussion

30% What is Machine Learning, EDA for Machine Learning, Classification (Prediction)

50% Logistic Regression, Intro to other algorithms

4/5

20% Topics in Pandas for Sci-Kit Learn

60% Decision Trees

20% SVM's

Take Home EXAM 2

4/12

30% Review Models Thus Far

50% How good are models? How do we tell? What do we tweak? Feature Scaling, Feature Selection

20% What else is out there??

4/19

60% Deep Learning (Neural Networks)

20% Advanced Applications

20% Review, Feature Generation, Steve's Presentation

4/26 Project 2 - Also when are finals?

Speaking Tips

Confidence

Slow down

We have never seen your data before, heard any of those acronyms and possibly never heard of your topic

Be yourself, don’t change for others...even though you are seeing them perform

Audience Tips

Take Notes? Possibly...

Active Listening

You may be very interested...express this and ask questions

Your data or interests may be closely related!

RESPECT

Jupyter Notebook Essentials

What we want by next week

- Understanding of our main function calls
  - Groupby, Value Counts, Cutting and Filtering
  - Intuition of why rows, columns robustness is important
  - Visualization examples

Try and put in the 3-4 hours…

- This is an ART as much as a SCIENCE
  - WHY did you do what you did (or someone else did)

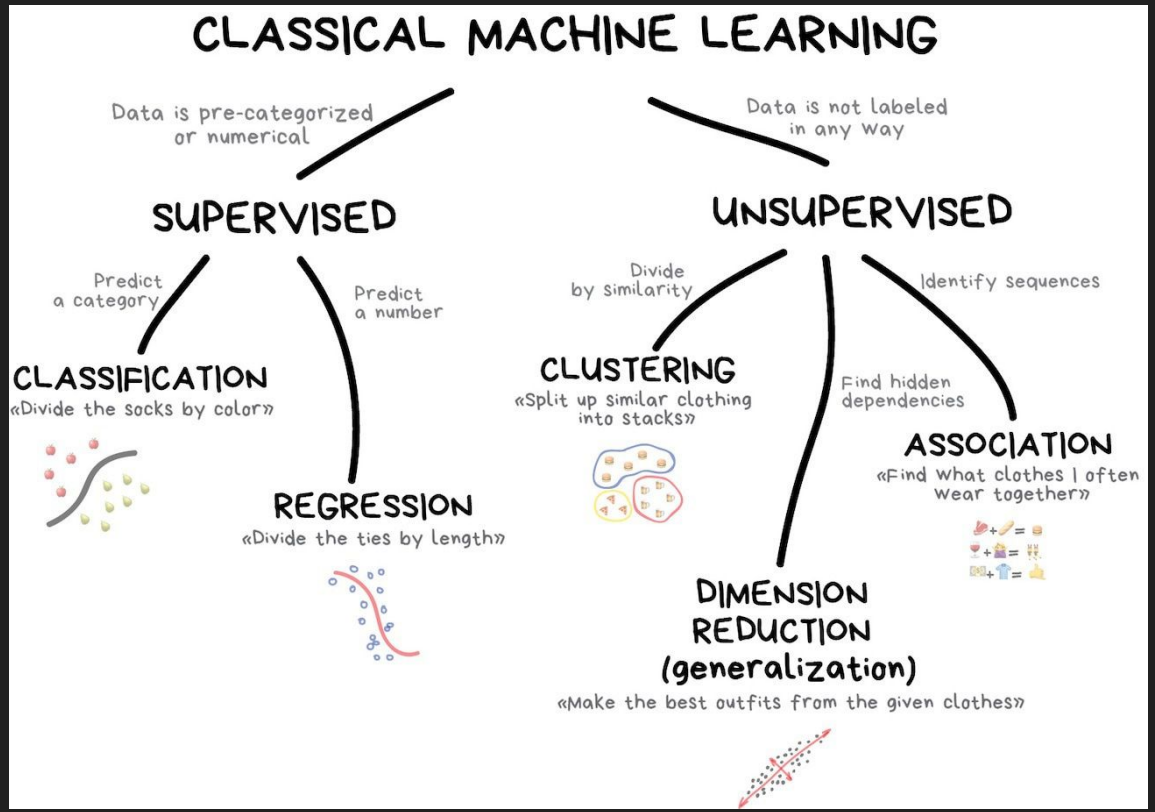
Thinking about the final project...

8 Minutes

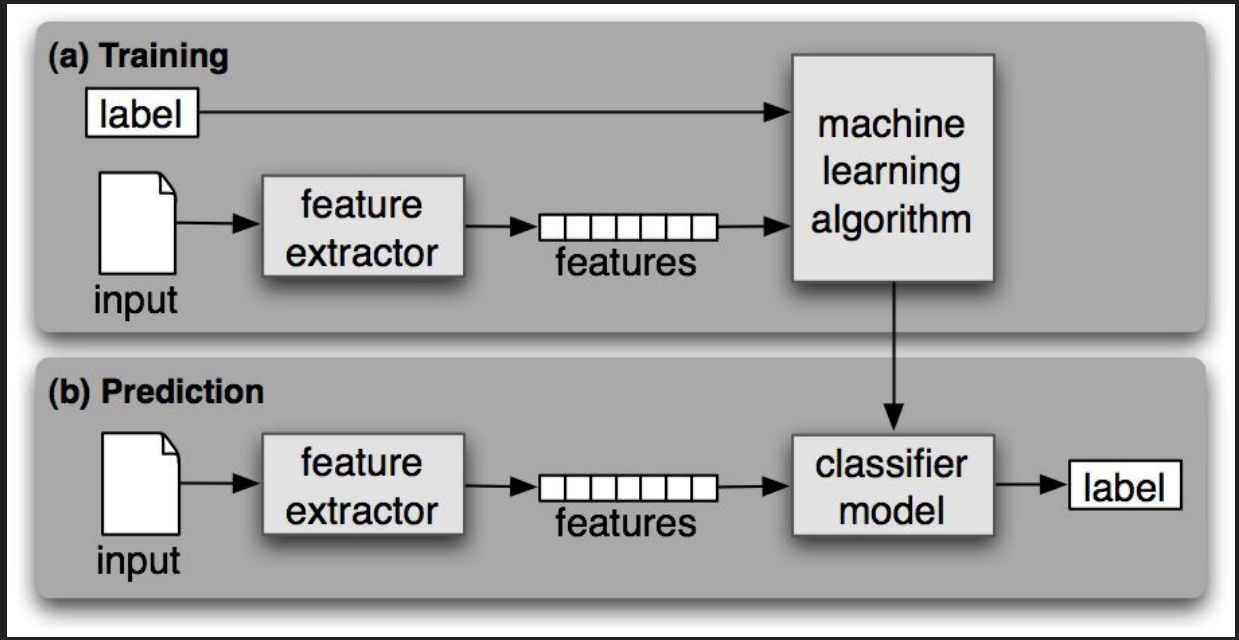
2-3 Intro to data & EDA

3-4 ML & Analysis

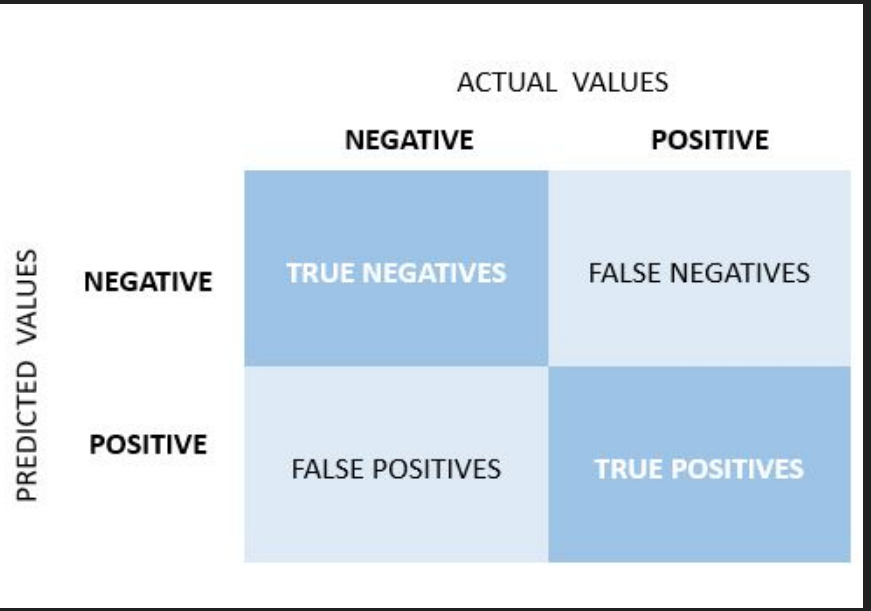
Classification vs. Clustering



Recall our algorithm for algorithms



How Good Is A Model



Metric Name	Formula from Confusion Matrix
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP+FP}$
Recall, Sensitivity, TPR	$\frac{TP}{TP+FN}$
Specificity, 1-FPR	$\frac{TN}{TN+FP}$
F1	$\frac{2 * precision * recall}{precision + recall}$