Introduction to Machine Learning

IST 402

Calvary Gollwitzer, Tevin Julien, Kevin Walters

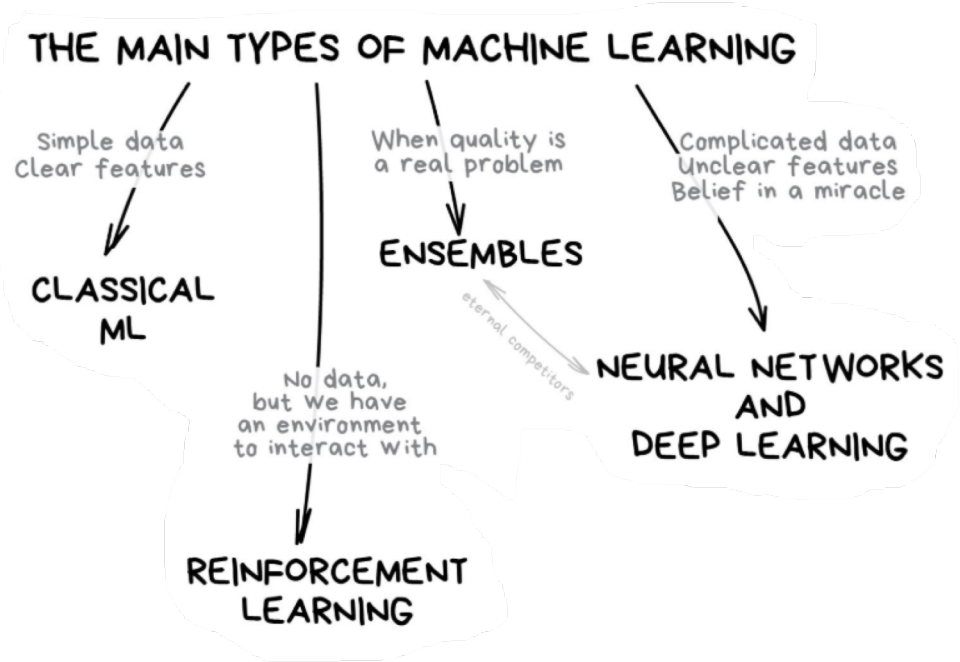
Needs for Machine Learning

- Consists of understanding that humans can use a computer to recognize algorithms or patterns.
- Computers are far more capable than humans to recognize, analyze and comprehend data and the hidden patterns that lie within them.
- This realization birthed the creation of machine learning.
- The single true goal of machine learning is to predict results based on incoming data. THAT IS IT.

Different types of Machine learning

The idea of machine learning can be broken up into 4 main categories.

- Classical Learning
- Reinforcement Learning
- Neural Nets And Deep Learning
- Ensemble Methods

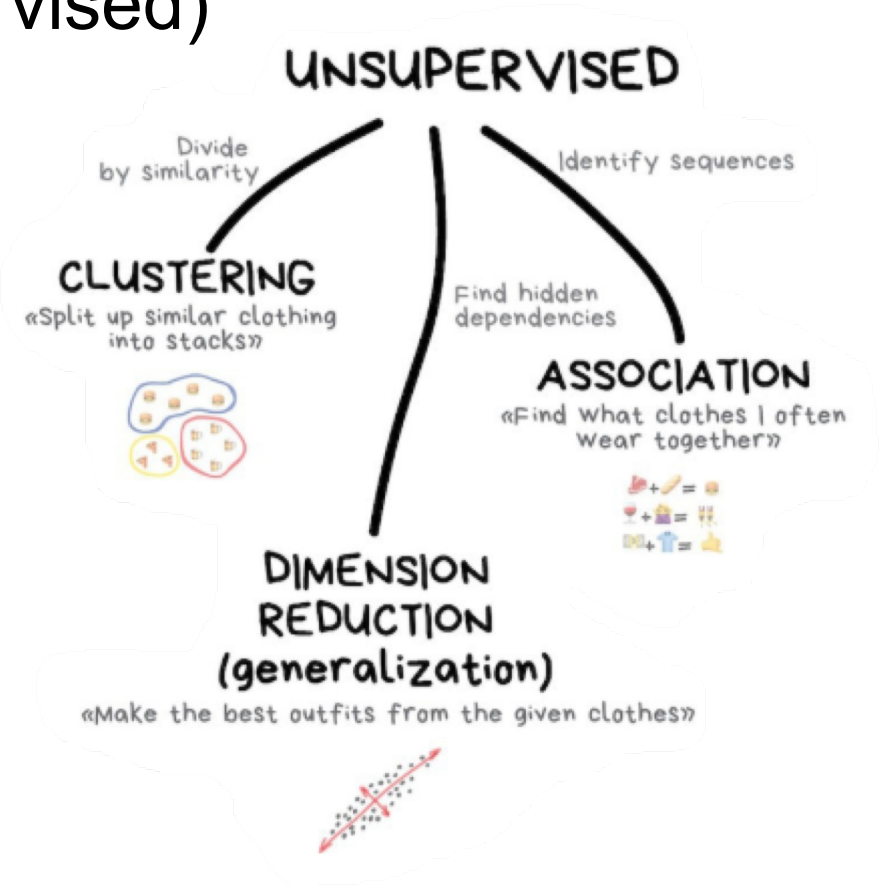


Classical Machine Learning

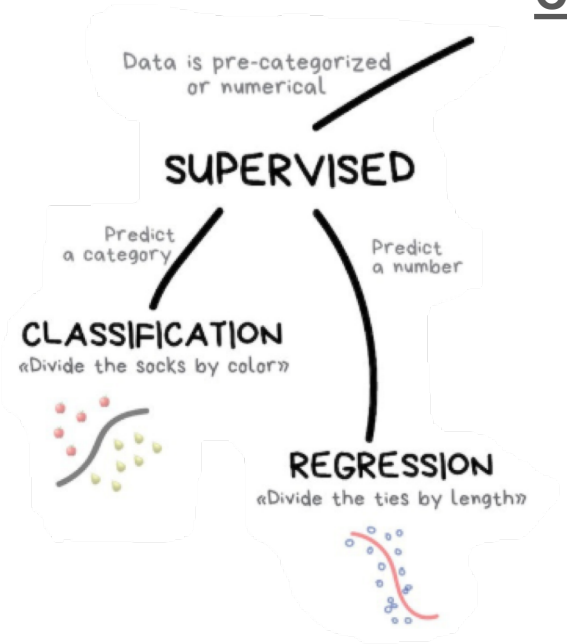
- Encompasses Supervised and Unsupervised learning
  - As well as all the subcategories that fall within those
- Also includes Ensemble Learning Methods
  - Several classical ML models, aggregate their predictions to get final prediction
- Is one of the earliest forms of general machine learning because it is a basic concept.
  - Has been around since the 50’s
  - Came from solving pure math tasks like pattern in numbers and proximity of data points

Classical Learning (Unsupervised)

- Unsupervised learning occurs when a machine learning model is given data without labels.
- The model simply looks for patterns within the features of the data.
- Ex.
    - Machine is given a dataset of people who do and don't have cancer. The machine only has the different features of each person to recognize patterns between them that may associate to cancer.

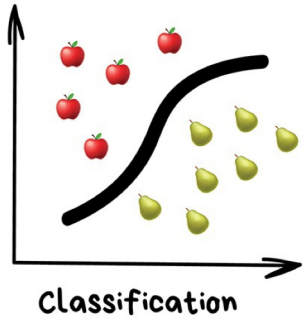


Classical Learning (Supervised)



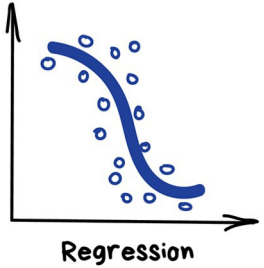
Classification

- **Binary** - two outcomes (yes/no)
  - Ex. Does patient have cancer?
- **Multi-class** - more than two possibilities for the variable being predicted
  - Ex. outcome of an election



Regression

- **Linear Regression Models**
  - Used when the variable you are looking to predict is continuous and can be plotted on a line
  - Ex. crop yields
- **Polynomial Regression Models**
  - Used when the variable can be predicted on a curved line



Binary Classification

Accuracy

Out of all ground truth labels, how many were predicted right?

$$\frac{TP+TN}{TP+FN+FP+TN}.$$

*Dumb classifiers that have good accuracies = data imbalance or data skew

Precision

Out of all positive predictions , how many actually are positive ?

$$\frac{TP}{TN+FN}.$$

Recall

Out of positive values in the data set, how many are predicted positive?

$$\frac{TP}{TP+FN}.$$

*Reverse functions to focus on negatives (different classifier)

Confusion Matrix- How Accurate is Your Model?

True Positive TP	False Negative FN
False Positive FP	True Negative TN

TP - correctly classified as 1
TN - correctly classified as 0
FP - falsely classified as 1 (correct is 0)
FN - falsely classified as 0 (correct is 1)

Ground Truth Label	Output of Hypothesis function h(x)~y	
1	0	FN
0	0	TN
0	1	FP
1	1	TP

Multi-Class Classification

- Predicts which individual will win an election (multi-class problem)
  - Data put into the model would be features about prior elections and other factors, and the variable would be who won the election.

- Binary Class: F: for the election (one k-dataset)
  - How many candidates stood
  - Year of election
  - How many from the democrats party
  - How many from republicain
  - Label: won

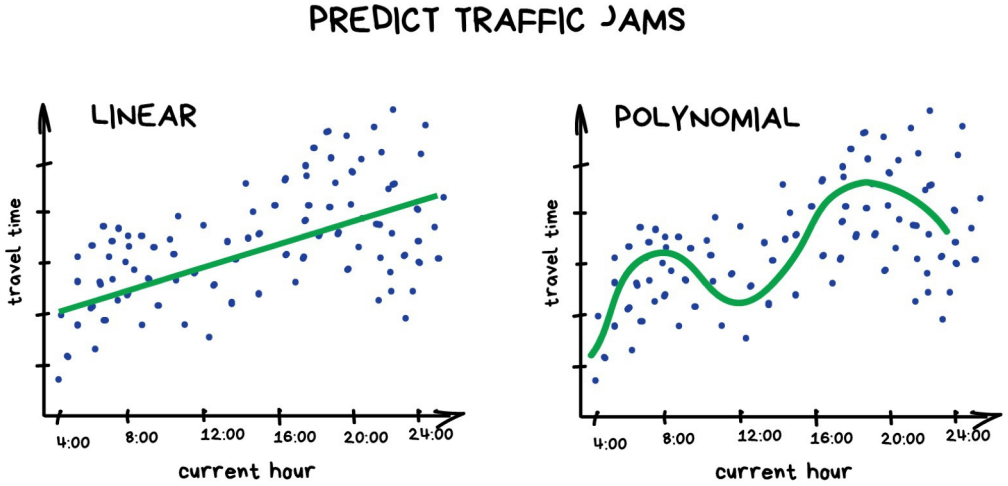
- **The easiest way to create a multi-class classification model is to break it down into multiple binary classification models in a one vs. all classification**

- **Take the predictions from the majority of the binary classifiers to determine overall decision**

Regression Models

- Regression models are used to predict variables that are continuous and can be put on a number line somewhere.
    - Ex. Price of a stock, Crop yields, Real estate prices
- Simple way to think of it is to predict a variable based on a condition, like traffic based on time of day.

- Computer looks for correlation between features but can take in many more variables then a human and is much more precise in its prediction.



REGRESSION

Reinforcement Learning (A brief overview)

- Reinforcement learning is a category of machine learning separate from classical learning.
- In RL, the model is not given any data to work with…
- Instead, the model has an environment that it can interact within where it is given a reward system for performing well towards its goal.
    - Ex. Model is tasked with trying to balance a bat on a point.
    - Simulation is run over and over again and the model slowly learns over time what actions prove most fruitful for it (how long the bat stays balanced)
    - Eventually the model becomes proficient at balancing a bat because it has learned what actions and reactions work best for it.