Intro to Machine Learning

What is it?

* Nvidia: Practice of using algorithms to parse data, learn from it, then make determination or prediction
* Stanford: Science of getting computers to act without being programmed
* UW: Algorithms figure out how to perform tasks by generalizing from examples
* **Machine learning is a subset of Artificial Intelligence whose algorithms find patterns in data**
* Supervised learning: Learning a pattern from labeled training data that allows us to make predictions about unseen or future data
  + Regression
  + Classification
* Unsupervised learning: Learning to recognize patterns in unlabeled data by disconvering an underlying structure in the data
  + Clustering
  + Data Transfoormation
* Reinforcement Learning: Learning by trial and error to reach a particular objective
* Deep Learning: Subfield of machine learning concerned with algorithms, known as artificial neural networks, inspired by systems and structures of the brain

## Regression

* When output variable is a real value, such as dollars or weight
* Problem: What should be value of each weight?
* Regression algorithm will learn what weights should be based on a training set of data that contains feature information and selling price of houses in that area

Classification

* When output is a category – like red or blue, disease or not
* Example: You just received an email saying “Congratulations! You won”
* Problem: Need to estimate probability that words found in email and words missing indicate it could be a spam email. (e.g cheap!, meeting)
* A classifier learns these probabilities from a training set containing spam, or not spam emails.

Clustering:

* Partitioning data into distinct groups with similar features or characteristics
* Example: Customer segemnetation – identify customers who may be potential users of product or service
* Problem: How to define “similar”

Data Transformation:

* Create a new representation of the data that is easier for huyman or other machine learning algorhuths to understand
* Example: Dimensionality Reduction – reducing a higher dimension representation of data, consisting of many features, down to two dimensions for visualization
* Problem: Which features are important for representation?

Reinforcement learning: Learning by trial and error to reach a particular objective

* Example: Teaching 4 legged robot to walk
* Problem: Initially have no idea what correct instructions are
* Reward function might give robot positive rewards for moving forwards, and negative for failing

Deep learning:

* Subfield of machine learning concerned with algorithms, known as artificial neural networks, inspired by function and structures of the brain

Role of data

* ML algorithm learns from data you provide it
  + Learns and maintains biases!
  + If your previous best employees had white names, black names in resumes will be biased against
* Maybe no ground truth at all