**AI and Machine Learning Week 1**

* ML is mainstream now
* Uses data to predict trends and get insight
* Examples: recommendations engines on websites, crop yields
* Libraries that do heavy lifting
* What is machine learning?
  + We want to write a program or a “machine” to create its own equation based on the dataset. The program learns how to calculate future data based on previous data
  + Algorithms
    - An algorithm can be thought as a general equation that may fit the task at hand
    - You do not come up with new algorithms
    - You pick the algorithm
  + Models
    - Model is a specific version of an algorithm that fits the data
    - Specific version is determined using past data, called training
  + Constraints and Hyperparameters
    - We may constraint the algorithm based on the problem
    - Example: salary can't be negative
    - The different constraints can be set through hyperparameters
  + Types of machine learning algorithms
    - Supervised
      * Regression: output is continuous
      * Classification: categorize data into given categories
    - Unsupervised
      * Clustering: Categorizing data into categories not given by the user
* Machine Learning Blueprint
  + 1. Data Collection
  + 2. Feature Engineering
    - Cleans the data
  + 3. Clean data is broken down: model based on the Training set, test set with data. Validation set is data that is retrieved separately from the training set data
  + 4. Algorithm -> Model -> Predictions
  + 5 live analysis
* Data engineer vs Data Scientist
  + Data engineer works on getting the data and putting it a form suitable for the ML
  + Data Scientists choose the algorithms and come up with models
* Feature Engineering
  + Features vs labels
  + Any supervised learning problem will have 2 things
    - Input called features
    - Output called labels
  + Bad data in, bad data out
  + Data must be clean, relevan and meaningful
  + Step 1. Data Cleaning
    - Identify bad records by checking validity
    - Fix or remove them
    - Validity checks
      * Out of range values
      * Impossible combinations = 20 years of experience, 19 years old
    - Step 2: Analyze Features
      * Make sense of data
      * Visualize data
      * Remove irrelevant features
      * Create new features
        + Binning -> combine several features into one for clarity
    - Step 3. Prepare the features for the algorithm
      * Features data categorized into 2 areas
        + Continuous: infinite options
        + Categorical: finite number of options
      * Non-numerical data cannot be fed to a ML algorithm
    - Step 4. Feature scaling
      * Normalization: Min Max Scaling
      * Scaling feature data down
* Linear Regression Algorithm
  + Regression model helps us predict a continuous value
  + Ordinary least Square method
* Draw line, sum of errors
  + Line of best fit is minimum errors
  + Hyperparamaters for Linear regression
    - Could be the intercept
    - Learning step
    - Stopping criteria or tolerance
  + Accuracy of Linear regression
    - No ML model will be 100%
    - We need to have a measure of accuracy
    - For regression problems, we use coefficient of determination