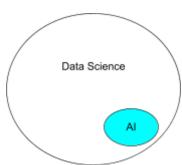**1/10/24**

Intro to AI and ML

- Data Science - the study of data to extract meaningful insights, especially for business
  - Artificial intelligence is a small part of datascience as a whole, it is a tool that helps make datascience easier
- Artificial intelligence - AI is the simulation of human intelligence through technology
- Machine learning - an extension of AI that speacializes in computers taking in and learning from data to provide a more desirable output after many generations of trials

1/31/24

Regression Models
- **Linear Regression**
  - best for a distribution that resembles a strait line
  - Most simple form of regression
  - Linear regression works best with data that only has 2 variables, and x and a y, where the x directly influences y
- You can randomly generate values for m and b within a set range that meakes sense for the data and then check how well the line fits the data by calculating the rmse. Do this a couple 100-1000 times and find the best equation fit
- Residuals are the distance between the line and the actual datapoint, the rmse (root mean squared error) is the square root of the sum of all the residuals squared. This tells you how far off on average your line is from the actual data

- **Decision Tree Regression**
  - A flow chart that splits the data into categories that are too different from each other to allow a single equation to explain it
  - The best data for this regression model is one with multiple variables or one that has datapoint that dont seem to follow a linear or even polynomial pattern
- First find the threshold between two data points that results in the lowest sum of squared residuals and that is the root of the tree, then continue doing that from the right side of the graph until you find all of your branches. The leaves of each branch will be the mean value within that branch
- The rmse is used to find how good a branch or root node is fitted to the data by averaging the distance between the mean of the branch and the actual data