20180327

Machine learning = about predicting new data with a model made by old data

statistics = analysis of data that is already there, make statements about data

Machine learning steps:

1) Define question/problem

2) Aquire training and test data

3) Wrangle, prepare, cleanse data

4) Analyze, I dentify patterns, explore data

5) model, predict, solve

**Models:**

**logistical regression:** use when outcome is binary (ie survived or not survived), data is quantitative, no correlated features, no outliers.

Made taking the natural log of the odds ratio. Below and above the line classifies the sample.

**Linear Discriminant analysis:** simlar to PCA, but tries to maximize seperation between classes, while PCA tries to maximize variance. Use when outcome is binary, input is quantitative. Auumes classes are normally distributed, variance, between outcomes is equal.

**KNN Nearest Niehgbors:** Allows multiple outcomes(ie, Iris species), no assumtptiosn about distribution, outcome classes already need to be distignuishable in problem space. Works by finding nearest neighbors to sample, and counting what class they are in, then assign the most represented class to sample.

**Classification and Regrassion Trees:** Can use numerical and categorical input, no assumptions about distribution, multiple outcomes. The approach is that it splits source data based on attribute value and makes a set of logical if then conditions. Its basically a flow chart! Classfication Tree is for categorical data, Regression trees are for continous outcome.

**Gaussian Naive Bayes:** Assumes high independence between input data (no colinearity), multiple outcome classes. It works by reling on Bayes thereom – look at probability or two outcomes. Different algorythyms based on outcome distribution.

**Support Vector Machines:** Multiple outcome classes, distribution of outcome is unknown. The approach is to fit ‘planes’ to data that give maximal ditance from classes, and seperate classes.

**SKLEARN HAS LOTS OF MODELS AND TUTORIALS FOR EACH- CAN LOOK INTO FO R MODEL!**