Team 8: Tanmay Patel, Alexandra Case, Aidan Lane, Bryan Jurado

What is the setting of your project? is it supervised, unsupervised, or semi-supervised? (I think everybody got supervision for this class though).

- Supervised! Supervised learning is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. The outcome of this project is to classify the quality and type of wine (red or white). This will be supervised learning and a classification problem where the model will classify the wine from 0-10 based on the learning model.

How you represent X (Features/Representation and their explanation, justification)

- There are 11 inputs in the input space, they are our features/representation!
- These inputs include: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, and alcohol
- X will be all of the characteristics that make up the wine.

How do you represent Y (binary? multi-class? sequences? trees? graphs?)

- The output is the quality grade of the wine (scoring between 0-10) . We represent quality as multi class because it quantifies the models ability to distinguish between each class.
- The color of the wine (white , red). Wine has a binary representation.

Which libraries will you use?

- We are using Scikit-learn because it is the most useful and robust library for machine learning in python. It provides a selection of efficient tools for machine learning and statistical modeling including classifications (which is what we use). For training, testing and splitting scikit-learn is beneficial.
- Pandas to read in the csv files. The Pandas Library is very good for data representation which is why we chose it.
- Matplotlib is the library we use to create a static and animated interactive visualization in python. Did you know that matplotlib makes things easy, and hard things possible! So how could we not use this feller.
- We are using NumPy because the NumPy arrays are faster and more compact than python lists. Eventually we are going to have to use arrays because we chose to do classification. Arrays consume less memory and are convenient to use. Numpy also uses less memory to store data and it provides a mechanism of specifying data types.