Wine Classification with a Neural Network

Group 6: Vikram P.T., Suyang Wang, Rui Li, Tyler Youngberg

Introduction

Data Set

Wine Quality: http://archive.ics.uci.edu/ml/datasets/Wine+Quality

Research Goal

Classify Two Types of Wine by Analyzing physicochemical variables

Data Cleaning

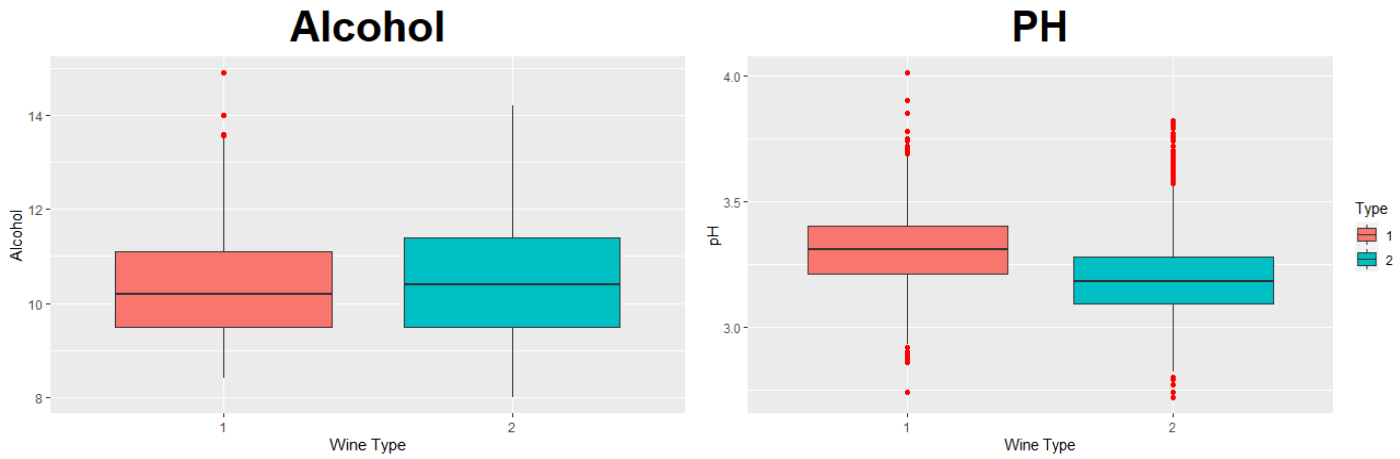
Combing Two data sets, Red Wine and White Wine, with the `rbind` function in R, and delete the non-using columns.

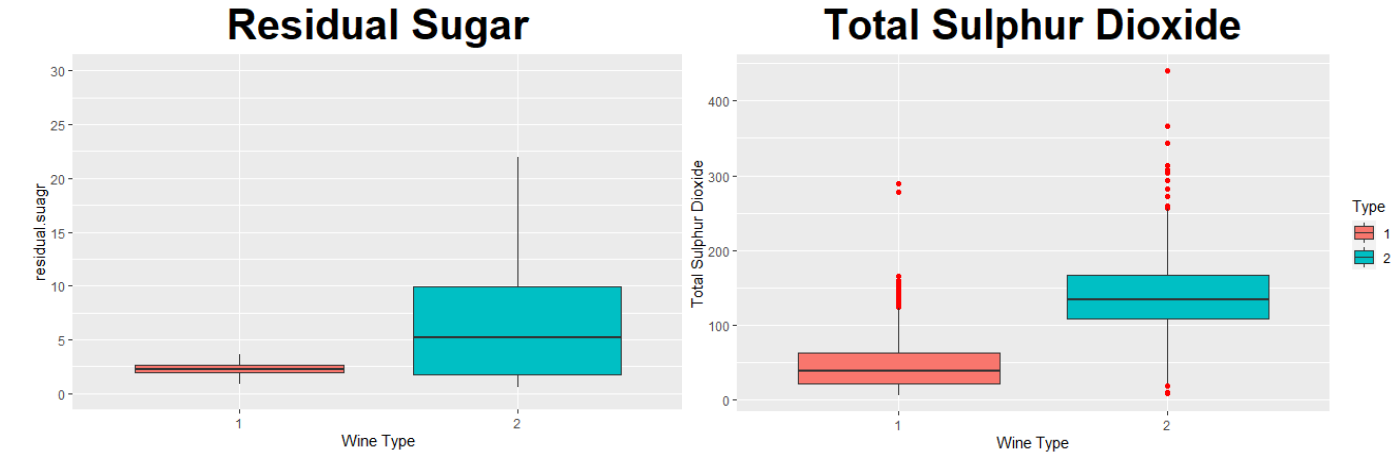
Type	Number	Relabel
Red	1599	1
White	4898	2

Data Description

Predictors							
	Min.	1st Qu	Median	Mean	3rd Qu	Max.	Units
fixed acidity	3.8	6.4	7	7.215307	7.7	15.9	g(tartaric acid)/L
volatile acidity	0.08	0.23	0.29	0.339666	0.4	1.58	g(acetic acid)/L
citric acid	0	0.25	0.31	0.318633	0.39	1.66	g/L
residual sugar	0.6	1.8	3	5.443235	8.1	65.8	g/L
chlorides	0.009	0.038	0.047	0.056034	0.065	0.611	g/L
free sulfur dioxide	1	17	29	30.52532	41	289	mg/L
total sulfur dioxide	6	77	118	115.7446	156	440	mg/L
density	0.98711	0.99234	0.99489	0.994697	0.99699	1.03898	g/mL
pH	2.72	3.11	3.21	3.218501	3.32	4.01	none
sulphates	0.22	0.43	0.51	0.531268	0.6	2	g/L
alcohol	8	9.5	10.3	10.4918	11.3	14.9	percent

- **Fixed Acidity:** Fixed or nonvolatile acid of wine (do not evaporate readily).
- **Volatile Acidity:** The amount of acetic acid in wine.
- **Citric Acid:** Found in small quantities, citric acid can add 'freshness' and flavor to wines.
- **Residual Sugar:** The amount of sugar remaining after fermentation stops.
- **Chlorides:** The amount of salt in the wine.
- **Free Sulfur Dioxide:** The free form of SO2 exists in equilibrium between molecular SO2 and bisulfite ion.
- **Total Sulfur Dioxide:** The amount of free and bound forms of S02.
- **Density:** The density of wine.
- **PH:** How acidic or basic a wine is.
- **Sulphates:** A wine additive which can contribute to sulfur dioxide gas (S02) levels.
- **Alcohol:** The percent alcohol content of the wine.





Overview of Analysis to Follow

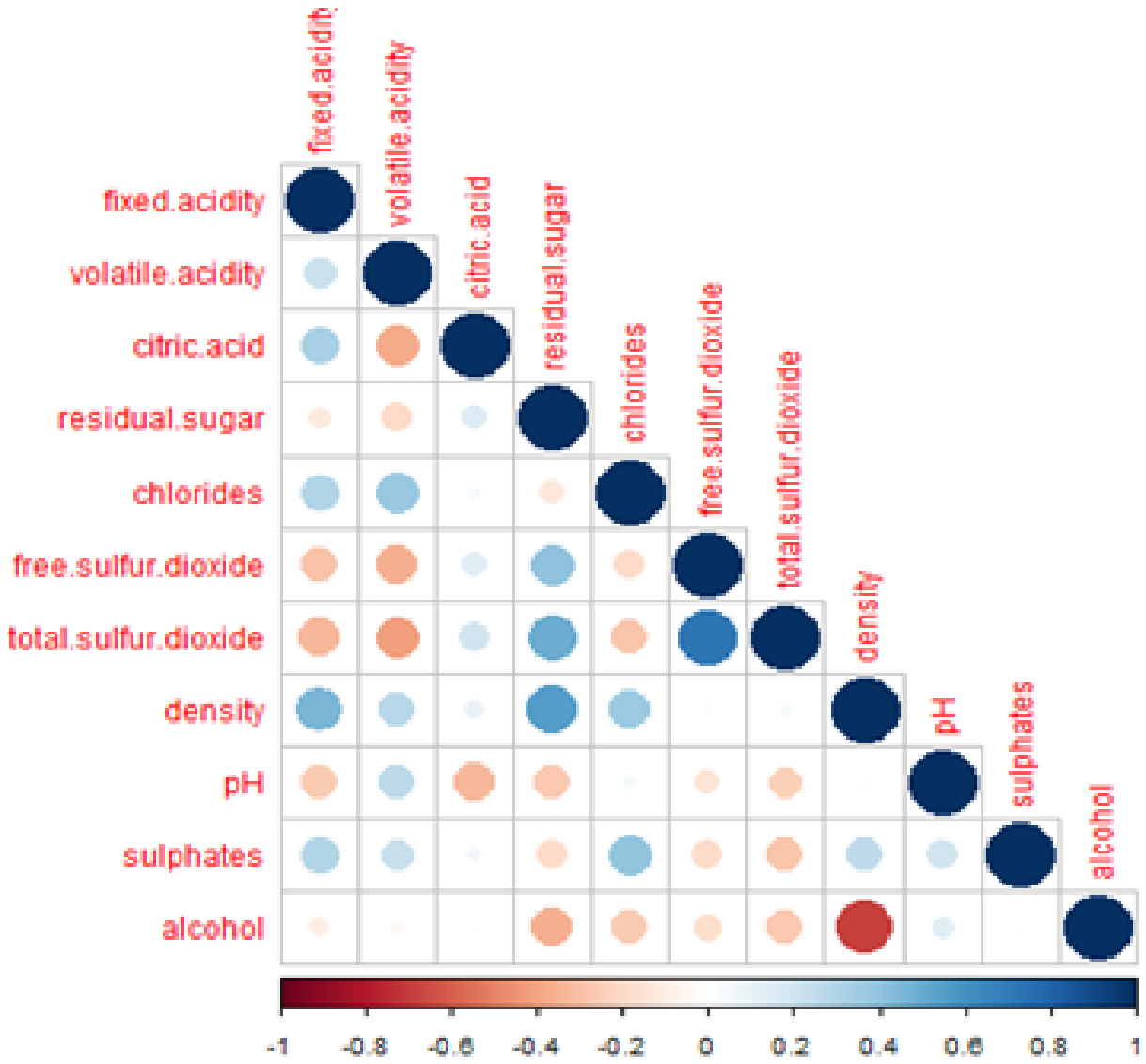
- Box Plots: We first create box plots for each predictor and focus on the boxplots spreading significantly different for Red and White Wins.

Overview of Analysis to Follow

- Correlation: We deeper investigate to the correlation between the predictors.

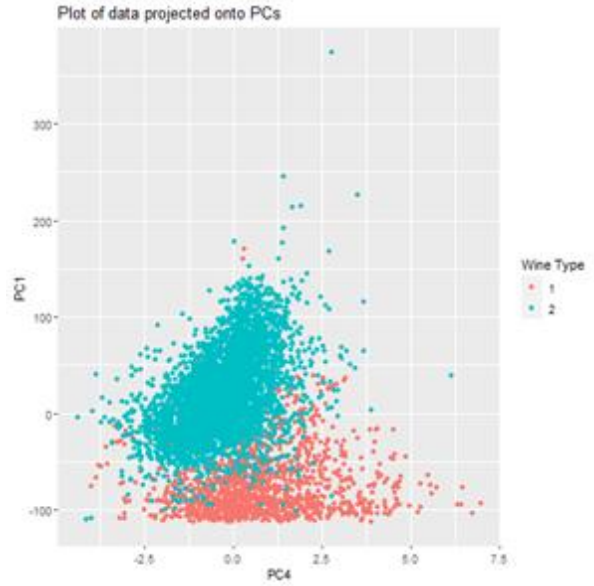
- **Density = concentrations of (Water + Sugar + Alcohol).**

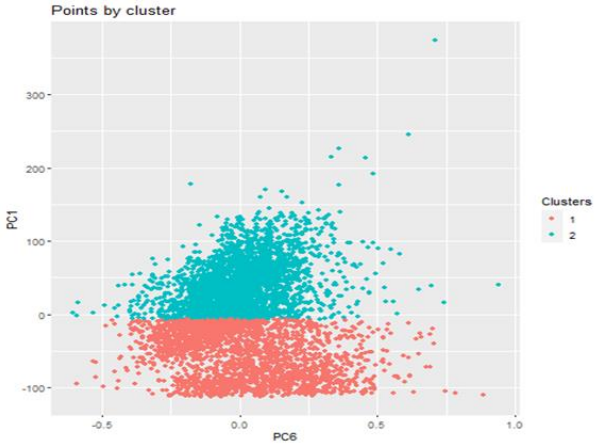
- **Water concentration = 100%(total concentration) – alcohol (concentration) - (sugar concentration)**

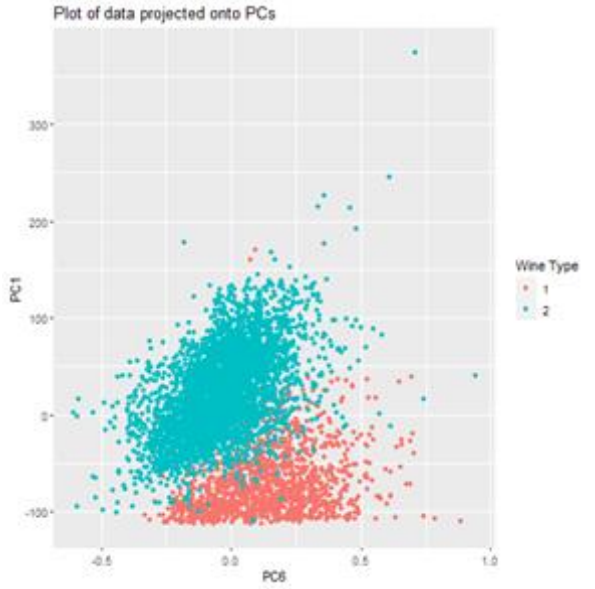


Overview of Analysis to Follow

- PCA: We use PCA to visualize whether there is separation of the 2 classes.
- Clustering: The K-means of the clustering has a very high False-Positive Rate.
- Neural Network: We use Neural Network to improve performance.

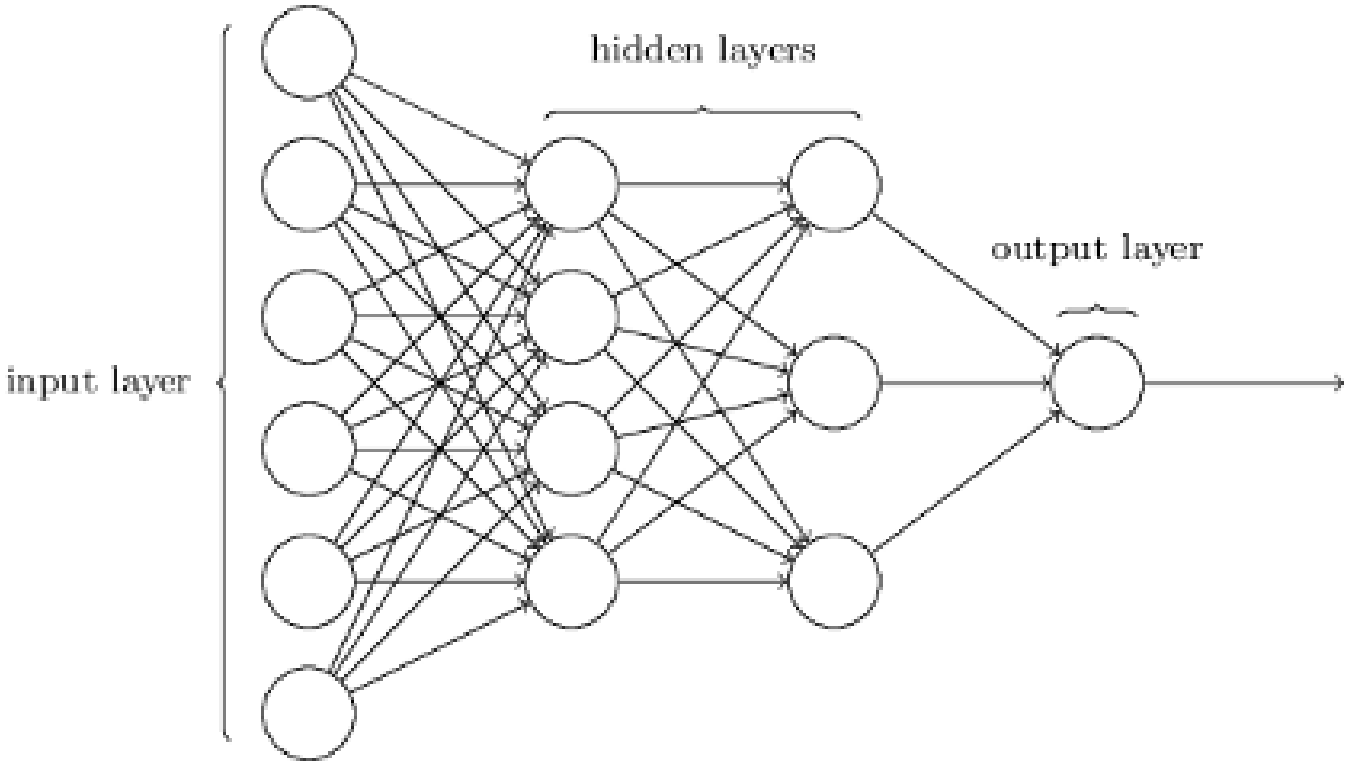


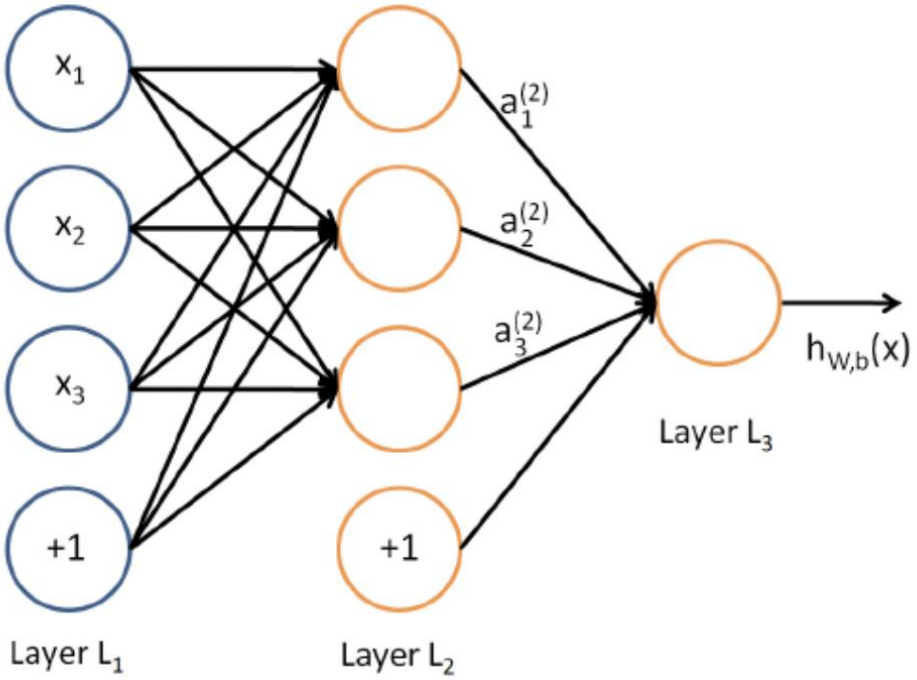




Neural Network

- "A neural network is put together by hooking together many of our simple 'neurons', so that the output of a neuron can be the input of another." (Andrew, 2011, p. 3)





Iterations of:

1. Forward Propagation
2. Backpropagation
3. Gradient Descent

To update parameters (W, b) minimizing the overall cost function.

Computation of activations

$$a_i^{l+1} = f(W_{i1}^{(l)} a_1^l + W_{i2}^{(l)} a_2^l + \cdots + b_i^{(l)})$$

Overall Cost Function

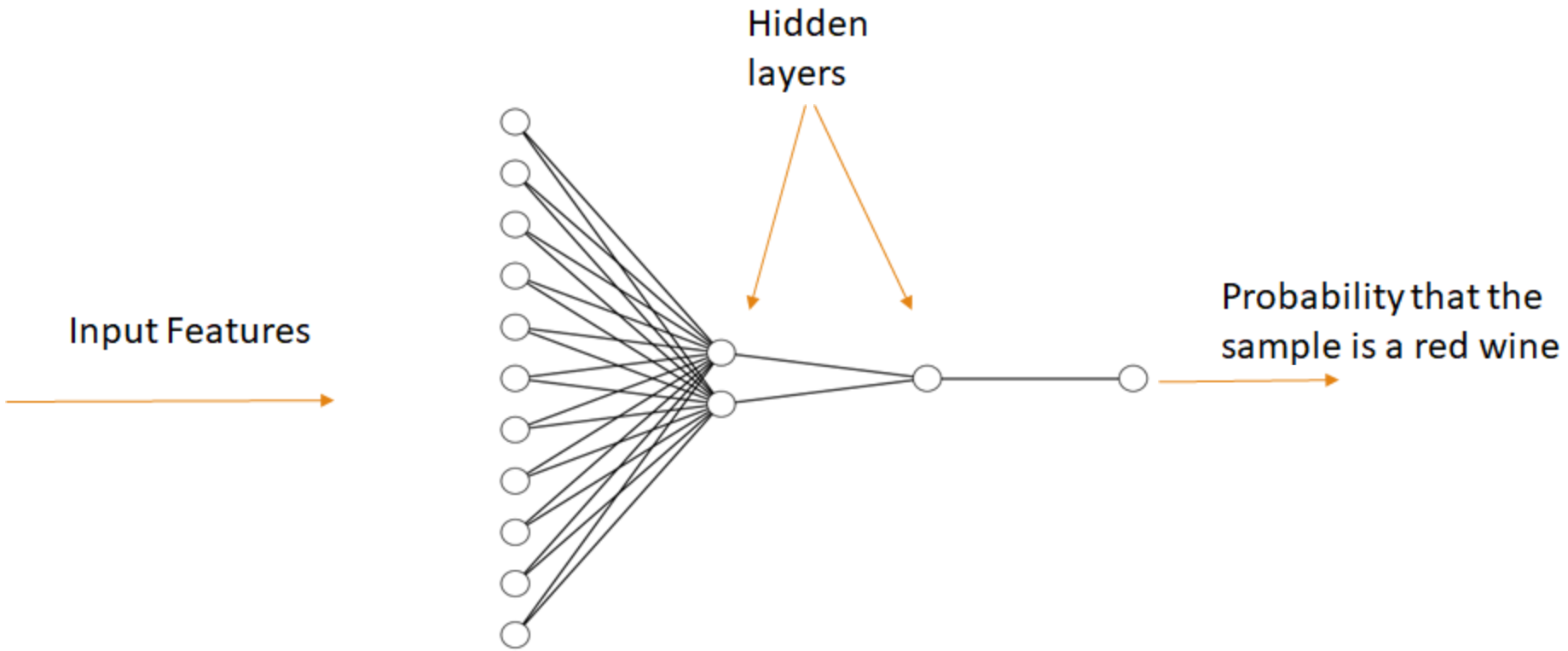
$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m} J(W,b;\, x^{(i)},\, y^{(i)})\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^{(l)})^2$$

Gradient Descent

$$W_{ij}^{(l)} := W_{ij}^{(l)} - \alpha\frac{\partial}{\partial W_{ij}^{(l)}} J(W,b)$$

$$b_i^{(l)} := b_i^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}} J(W,b)$$

Our Model

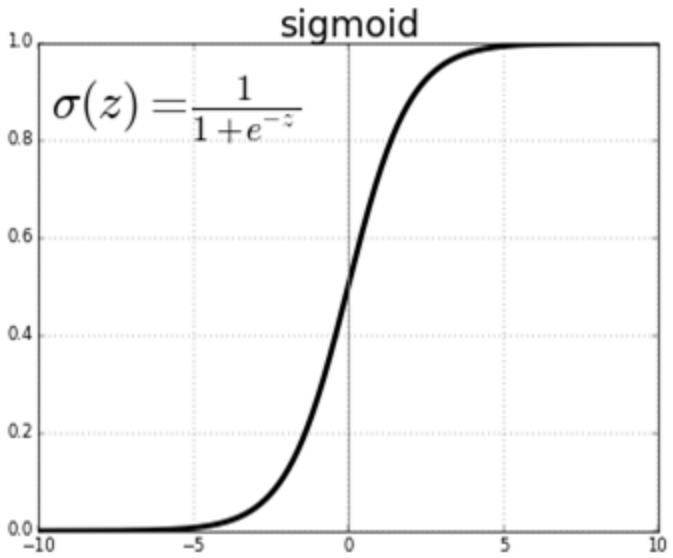

Hidden
layers
Input Features
Probability that the
sample is a red wine

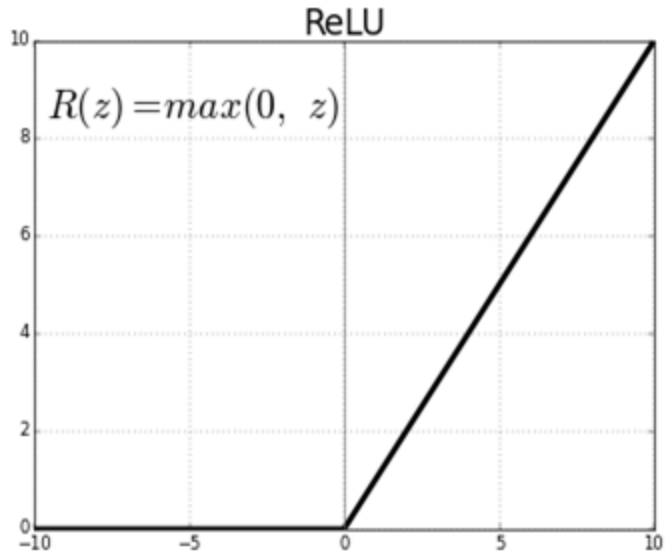
Our Model

- 11 input nodes correspond to the features of our data.
- 2 hidden layers and 1 output layer
- The activation function used for the hidden layers is RELU and for the output is Sigmoid.
- The number of hidden layers and units in each layers a hyper-parameters and can be tuned using cross-validation or using a test set.
- Used the binary cross entropy as loss/cost function.

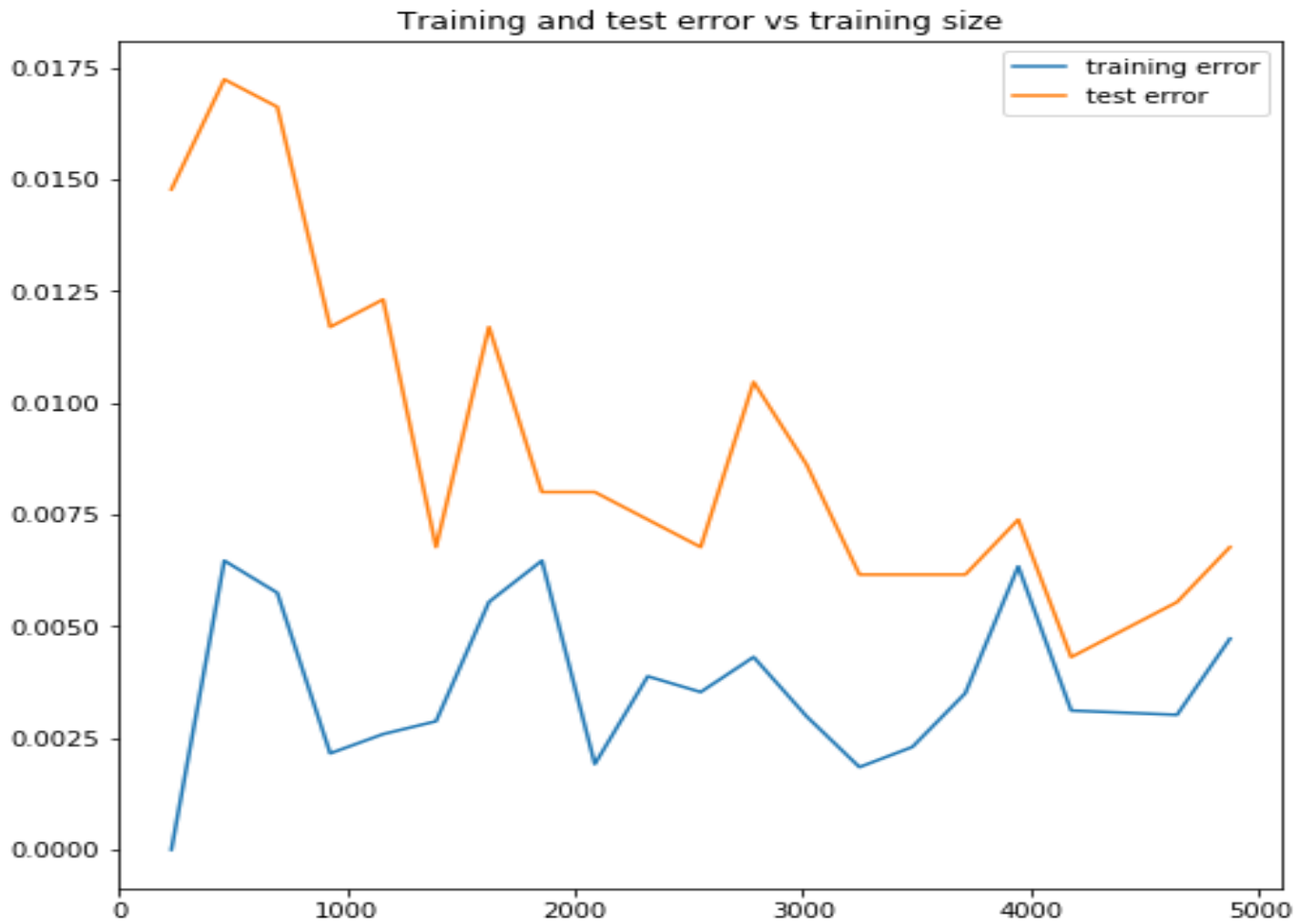
$$Error = \sum_{i=1}^{n} -(p_i \log q_i + (1 - p_i) \log(1 - q_i))$$

RELU and Sigmoid



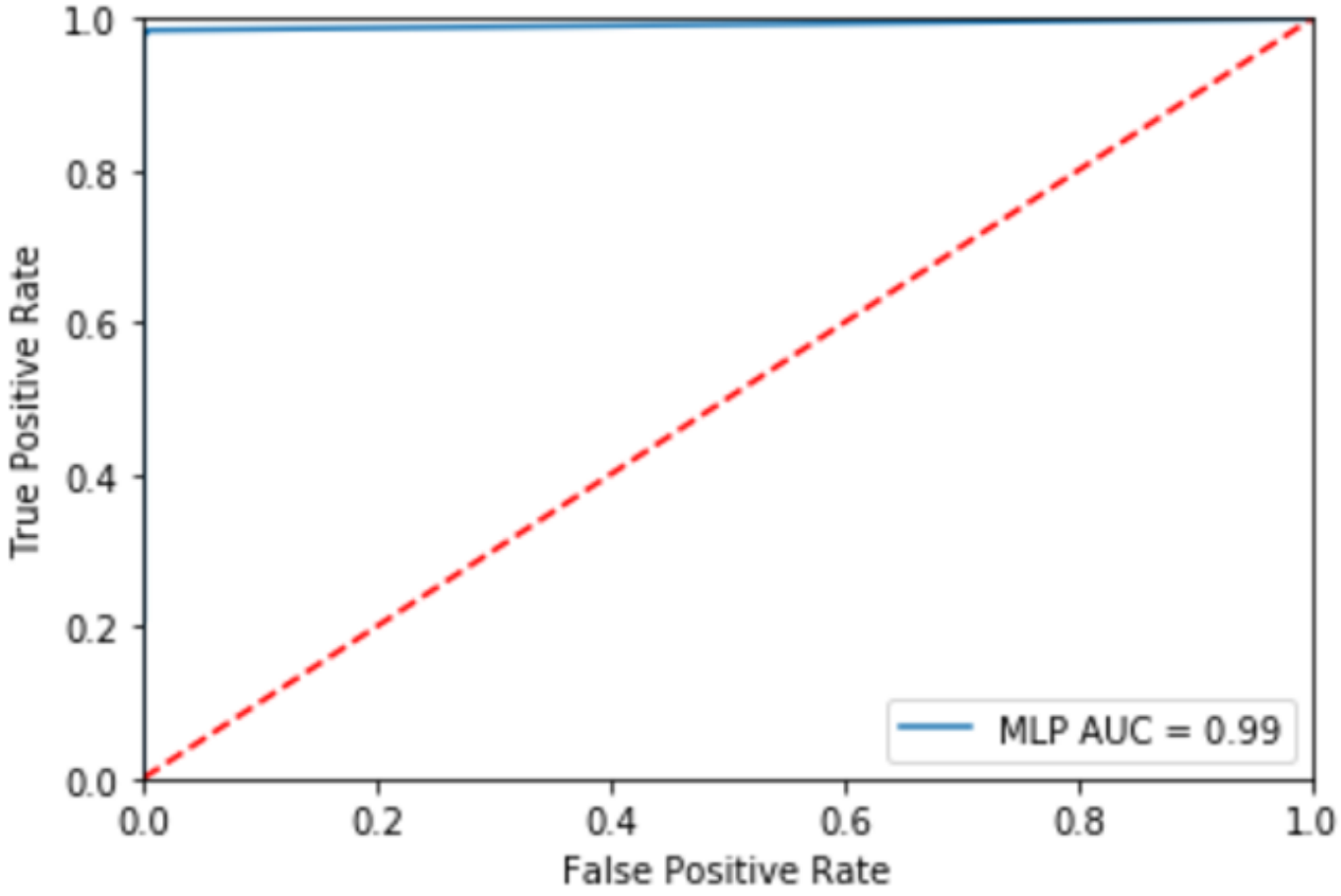


Training and Test error vs Training set size



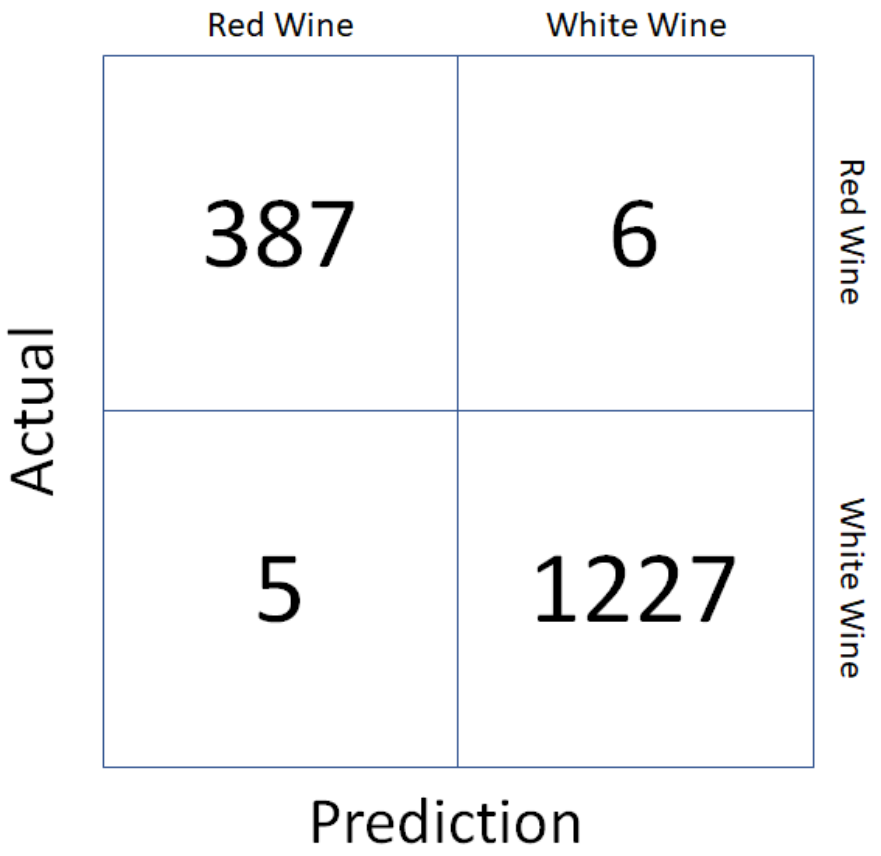
Receiver Operating Characteristic (ROC) Curve

- Shows the performance of the model against all possible thresholds
- Area Under the Curve (AUC) is 0.99
  - 0.5 = random classification
  - 1.0 = perfect classification



Confusion Matrix

- 0.0066 overall error rate
- 0.0153 error rate for red wine
- 0.0041 error rate for white wine



Conclusion

- Neural network demonstrates that it is possible to discern between red and white wine with high accuracy based on these predictors
- Does not reveal *how*
- A simpler method, such as logistic regression, could be used to give an idea of how the predictors differ between the two

THANK YOU !