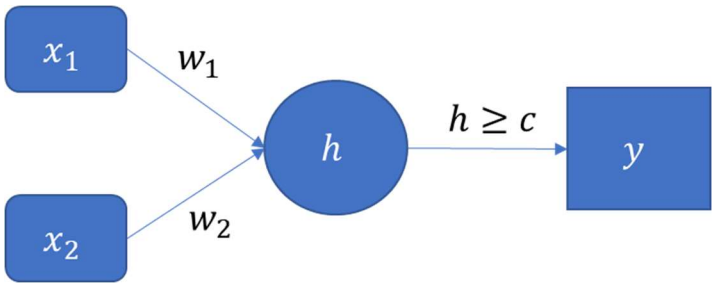
CS 584-04: Machine Learning

Fall 2018 Assignment 5

Question 1 (40 points)



a) (10 points). If we restrict the values of the parameters w_1, w_2, and c to positive integers, then specify the lowest possible values for these parameters such that the perceptron can implement the logical AND function.

W1 = 1

W2 = 1

C = 2

b) (10 points). If we restrict the values of the parameters w_1, w_2, and c to positive integers, then specify the lowest possible values for these parameters such that the perceptron can implement the logical OR function which can be represented by the following table:

x_1	0	0	1	1
x_2	0	1	0	1
x_1 OR x_2	0	1	1	1

W1 = 1

W2 = 1

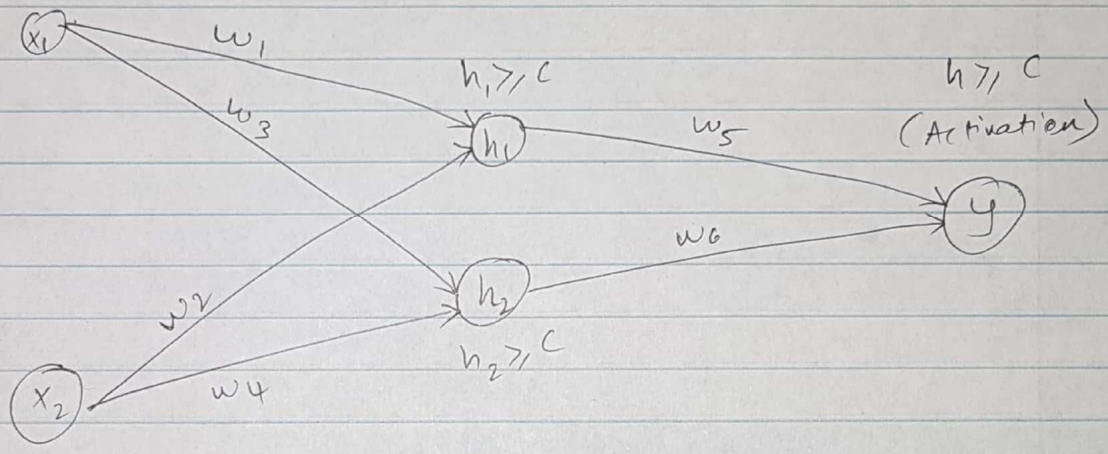
C = 1

c) (20 points). The logical XOR function (i.e., the Exclusive OR) returns TRUE only when one argument is TRUE and another is FALSE. Otherwise, it returns FALSE. This can be represented by the following table:

x_1	0	0	1	1
x_2	0	1	0	1
x_1 XOR x_2	0	1	1	0

Consider a neural network which has two neurons in a single hidden layer. Specify the four synaptic weights and a threshold value such that the neural network can implement the XOR function. The parameters are still integers, but we allow negative integers.

A single layer neural network cannot implement XOR gate because the classes in XOR aren't linearly seperable. We cannot draw a line to seperate inputs (0,0), (1,1) and inputs (0,1) (1,0) Thus we need multi-layer neural networks to implement this



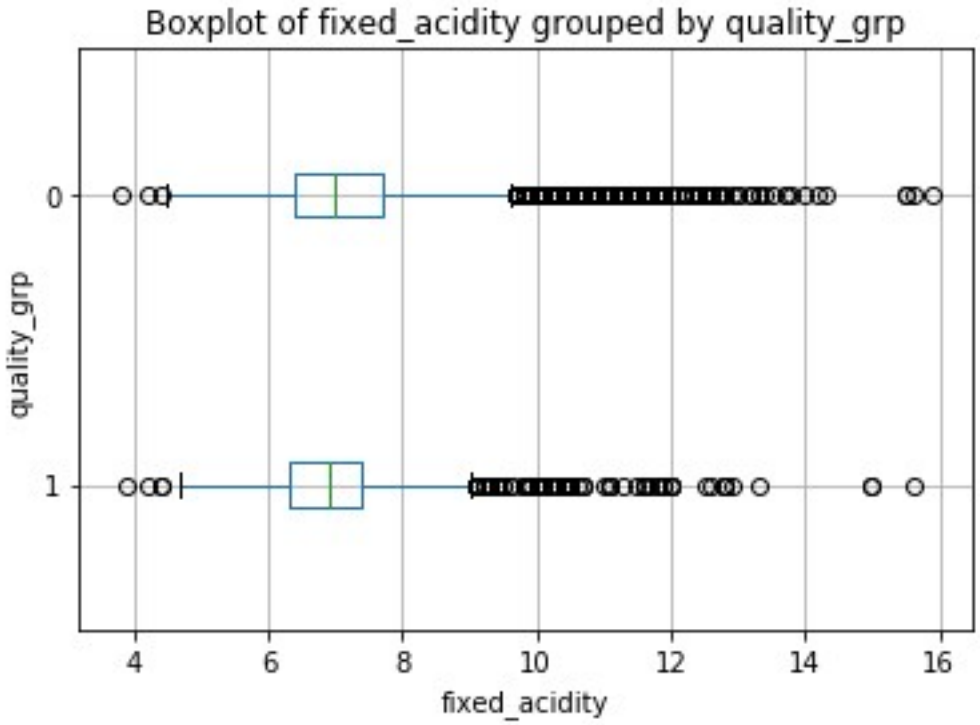
The minimum integers for parameters that will implement XOR gate are:

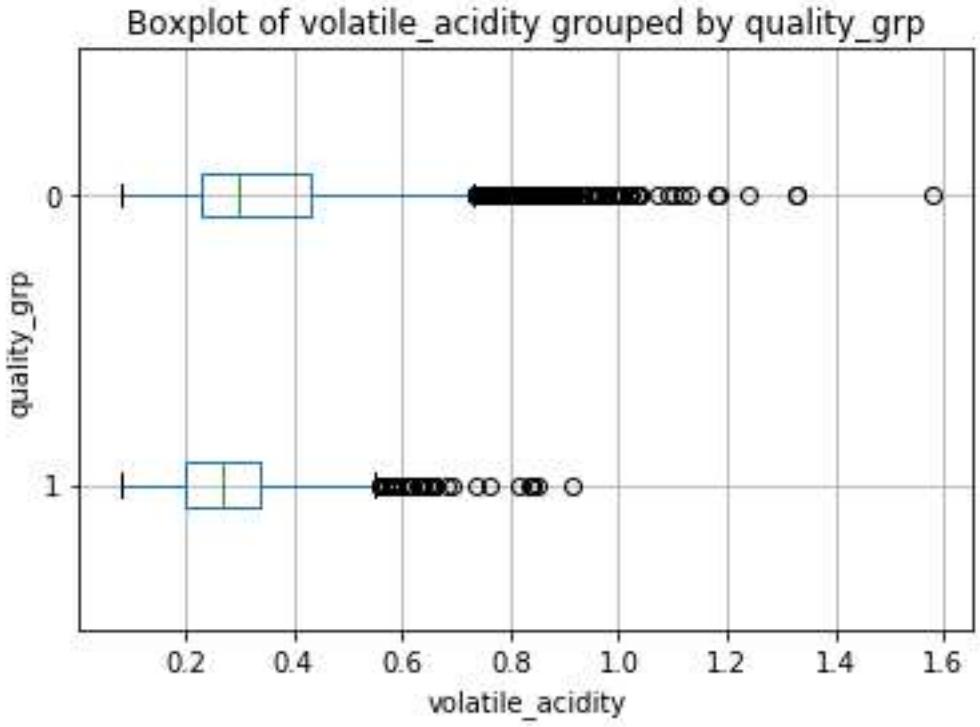
$w1 = 1$, $w2 = -1$, $w3 = -1$, $w4 = 1$

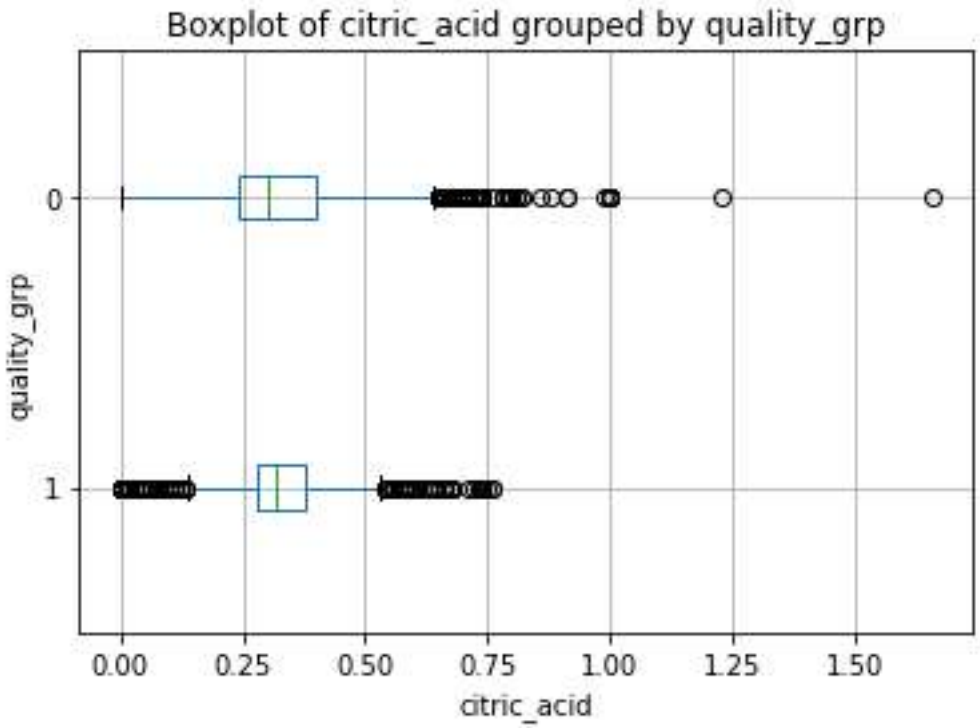
$w5 = 1$, $w6 = 1$, $c = 1$

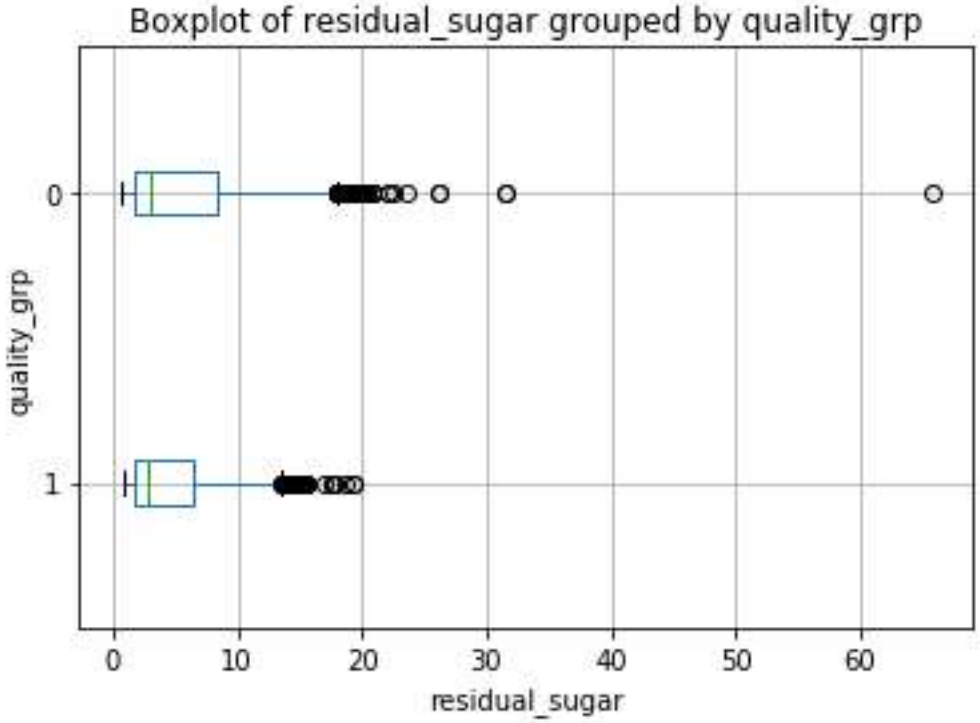
Question 2 (60 points)

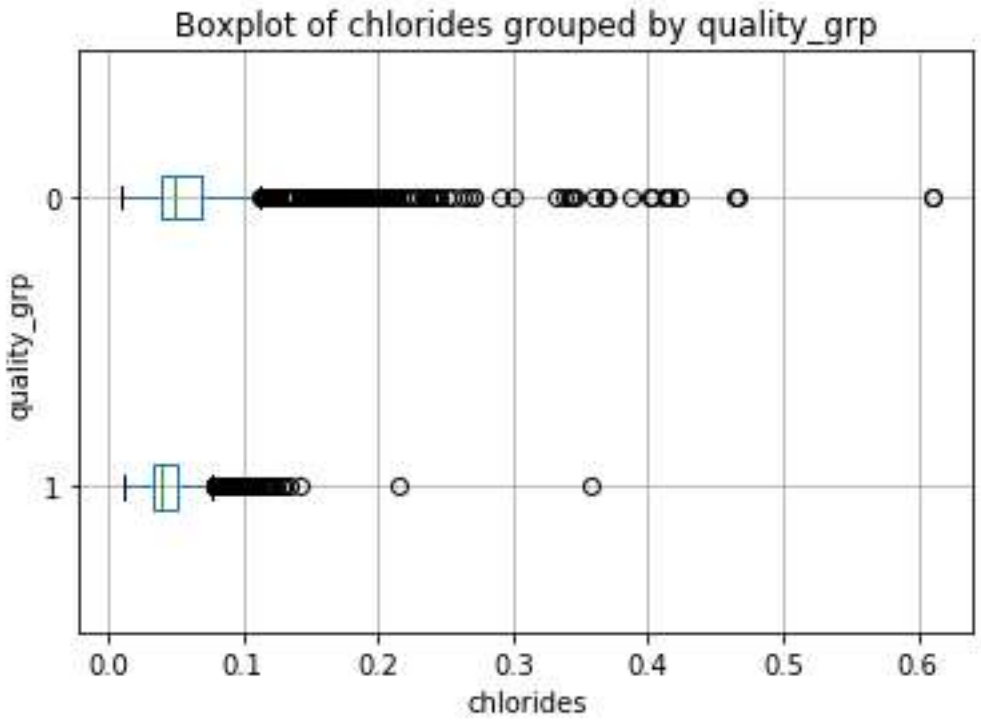
a) (10 points). Generate a horizontal box-plot for each input attribute, grouped by the target variable quality_grp.

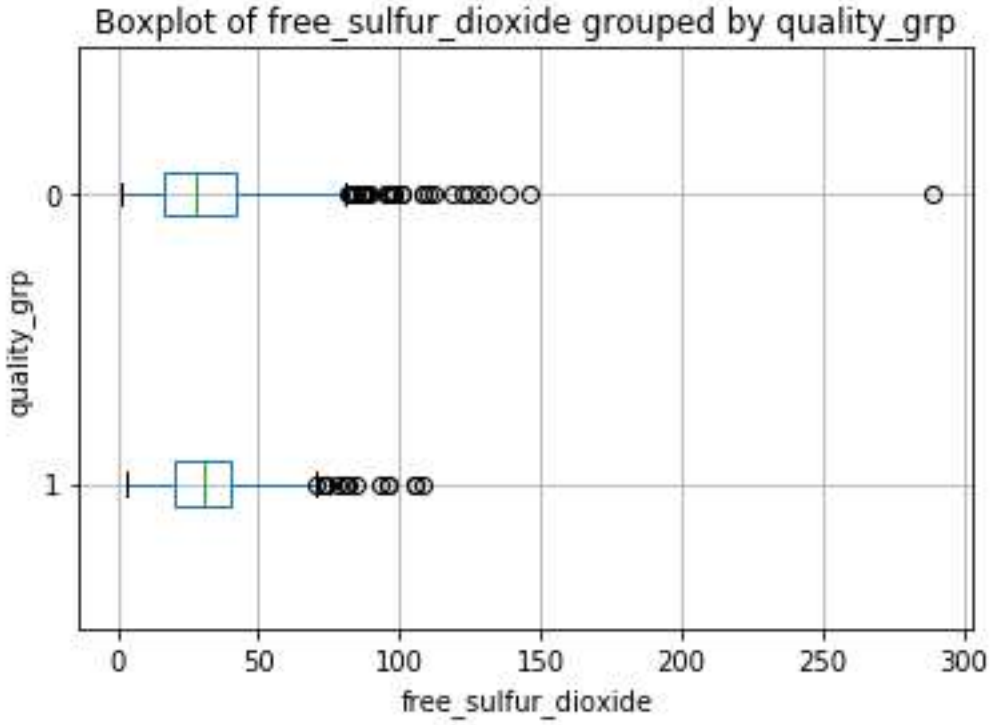


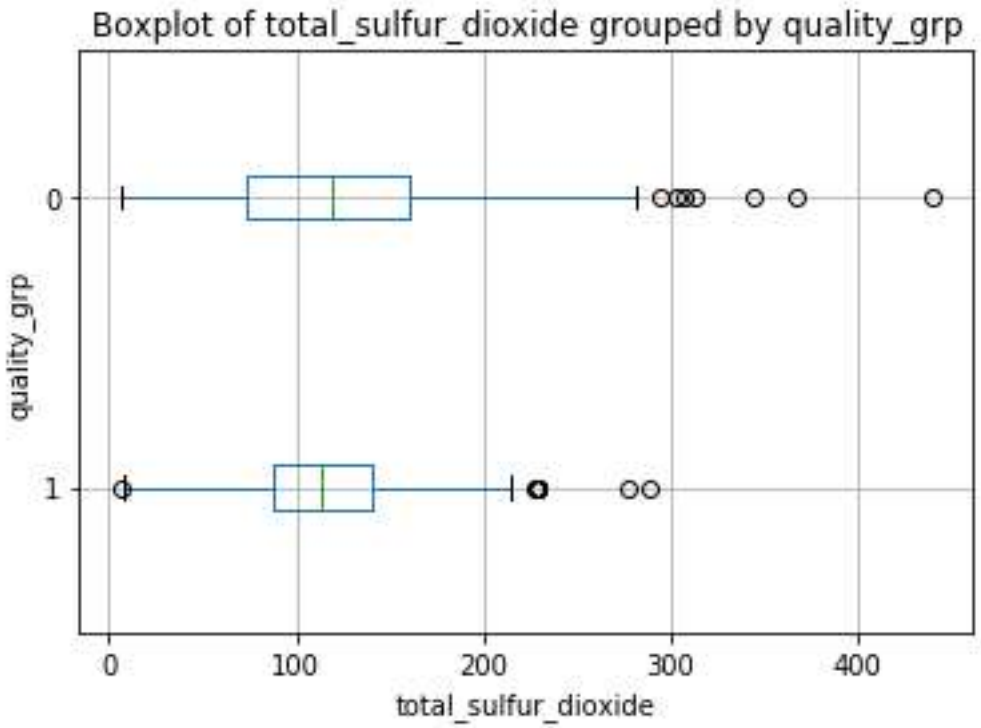


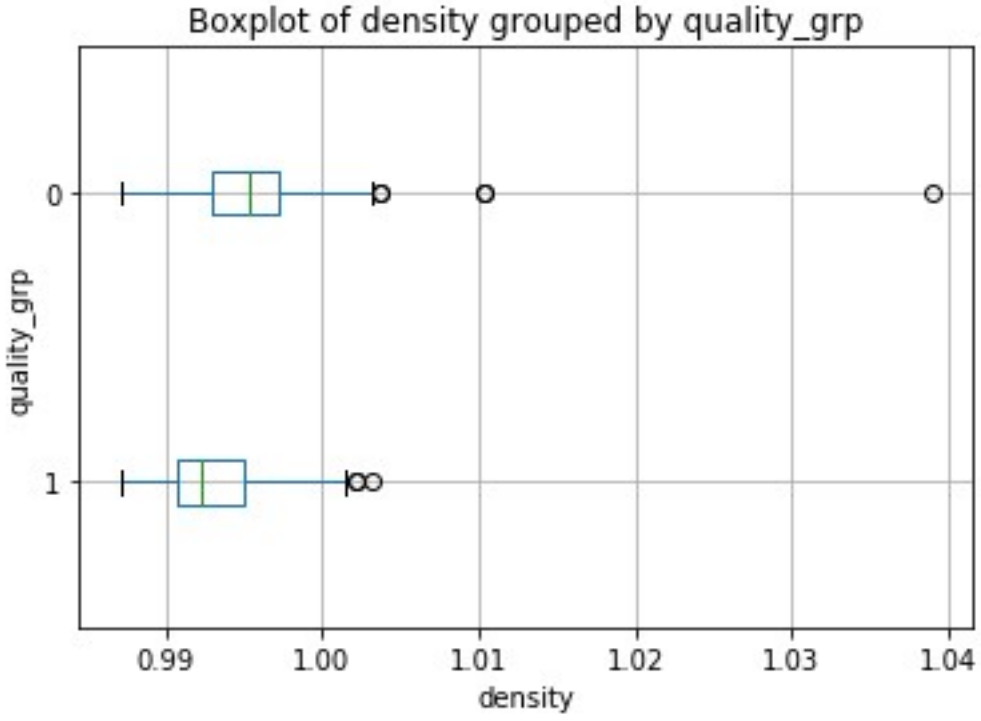

Boxplot of citric_acid grouped by quality_grp
quality_grp
0
1
0.00
0.25
0.50
0.75
1.00
1.25
1.50
citric_acid

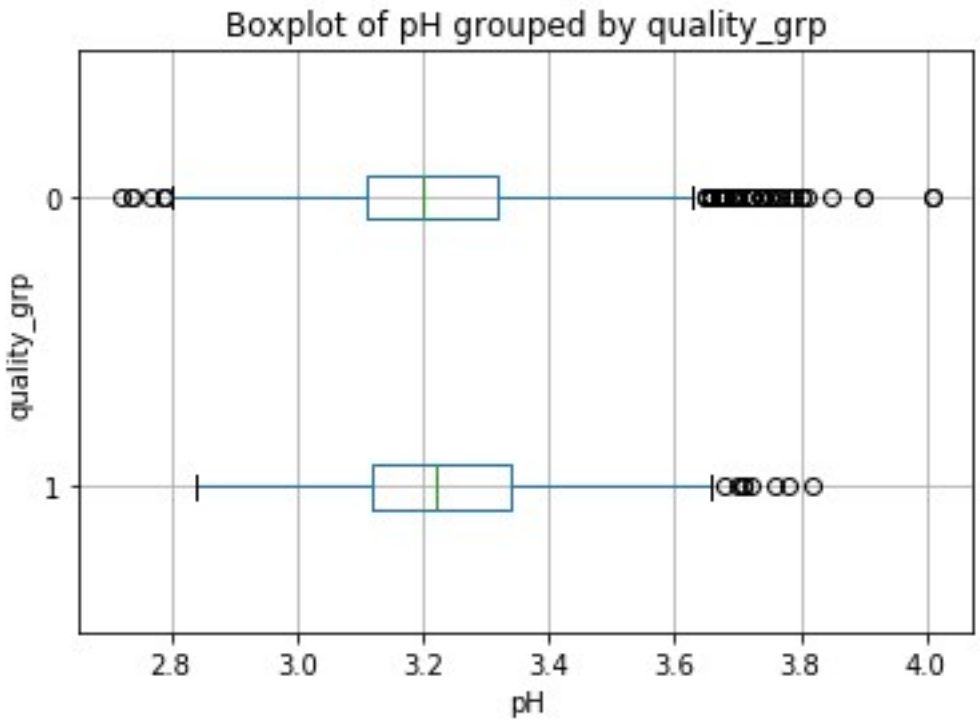

Boxplot of residual_sugar grouped by quality_grp
quality_grp
0
1
0
10
20
30
40
50
60
residual_sugar

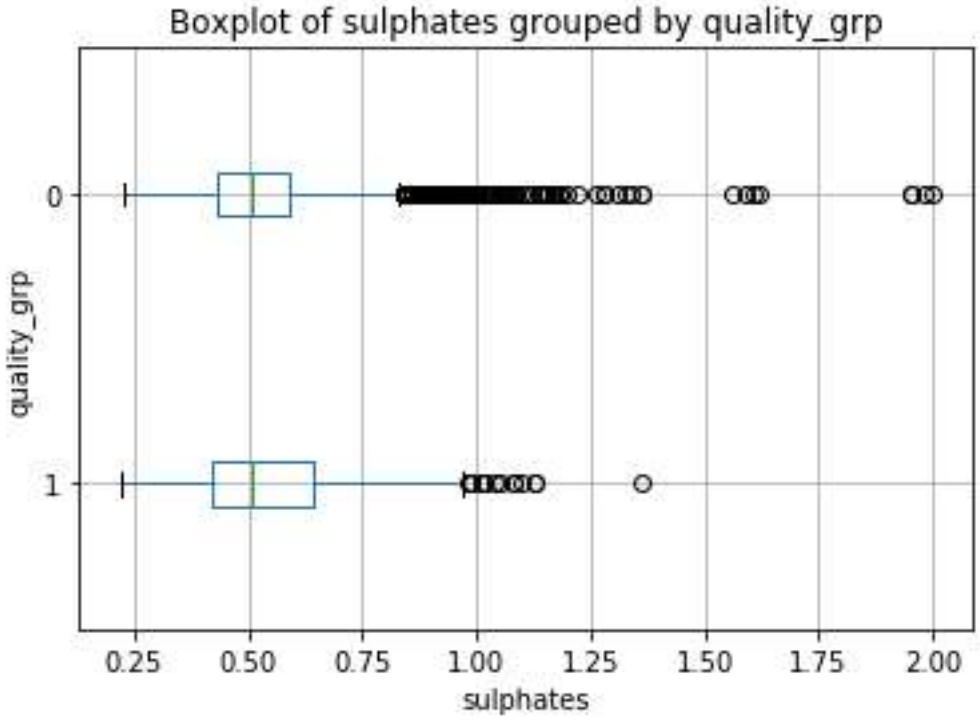

Boxplot of chlorides grouped by quality_grp
quality_grp
0
1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
chlorides

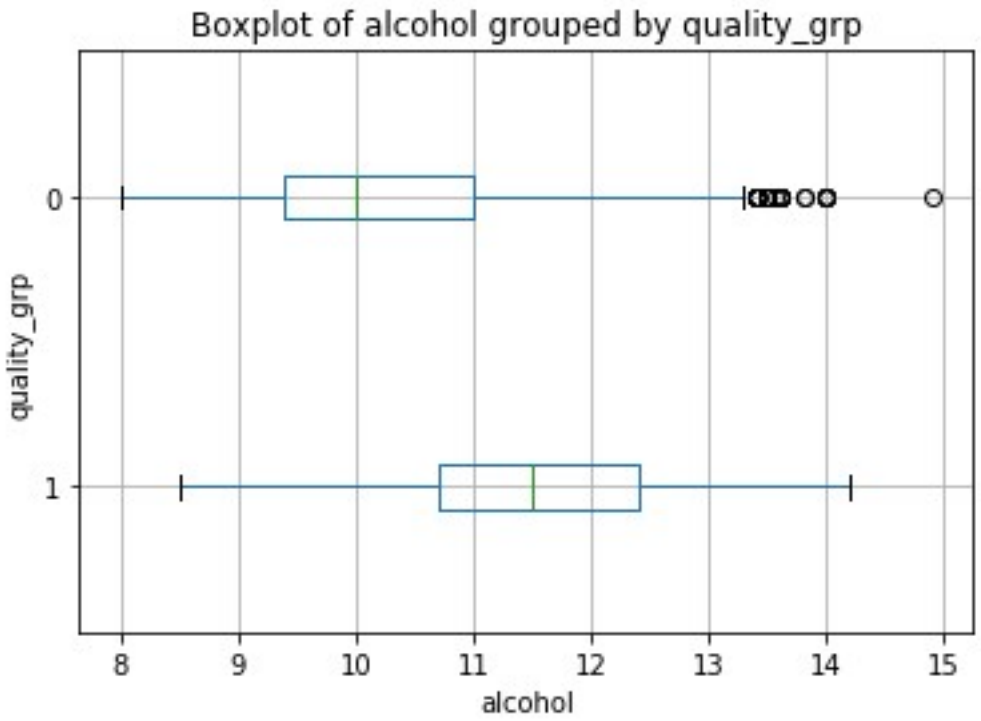

Boxplot of free_sulfur_dioxide grouped by quality_grp
quality_grp
0
1
0
50
100
150
200
250
300
free_sulfur_dioxide


Boxplot of total_sulfur_dioxide grouped by quality_grp
quality_grp
0
1
0
100
200
300
400
total_sulfur_dioxide


Boxplot of density grouped by quality_grp
quality_grp
0
1
0.99
1.00
1.01
1.02
1.03
1.04
density


Boxplot of pH grouped by quality_grp
quality_grp
0
1
2.8
3.0
3.2
3.4
3.6
3.8
4.0
pH


Boxplot of sulphates grouped by quality_grp
quality_grp
0
1
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
sulphates



b) (10 points). The scipy.stats module has the ttest_ind function for comparing two independent samples using the Student's *t* test. Use this function to calculate the two-sided *p*-value of the Student's *t* test. The group variable is the target variable quality_grp. List the names of the input attribute, their *t* statistics, and their two-sided *p*-values. The rows are in descending order of the two-sided *p*-values.

Index	lab	statistic	p_value
5	free_sulfur_dioxide	1.19023	0.233999
8	pH	2.26945	0.0232738
9	sulphates	2.73935	0.00617293
0	fixed_acidity	-3.98992	6.6823e-05
6	total_sulfur_dioxide	-4.13378	3.61338e-05
2	citric_acid	4.39428	1.12906e-05
3	residual_sugar	-5.1678	2.43932e-07
1	volatile_acidity	-12.37	9.32718e-35
4	chlorides	-13.2123	2.39099e-39
7	density	-23.0914	1.89026e-113
10	alcohol	34.1838	1.17967e-235

c) (10 points). Perform the Support Vector Machine analysis using the svm.LinearSVC function. The random state value is specified to 20181111. The maximum number of iterations is specified to 10000. When the algorithm does not converge, we may need to remove some variables. You will first use all eleven input attributes. If the algorithm does not converge, you will remove the variable which has the highest *p*-value. If that does not help the algorithm converge, then the variable which has next highest *p*-value is removed, and so on. What input attributes are retained such that the algorithm can converge for the first time?

Input attributes that are retained such that the algorithm can converge for the first time:

['volatile_acidity', 'chlorides', 'density', 'alcohol']

d) (5 points). What is the Mean Accuracy of your model in (c)?

Mean Accuracy of the model in (c) = 0.8123749422810528

e) (5 points). What is the hyperplane? You need to present the hyperplane in this format $w_0 + w_1x_1 + \cdots + w_px_p$. Include only the attributes that you use in (c).

In geometry, a hyperplane is a subspace whose dimension is one less than that of its ambient space. If a space is 3-dimensional then its hyperplanes are the 2-dimensional planes, while if the space is 2-dimensional, its hyperplanes are the 1-dimensional lines. This notion can be used in

any general space in which the concept of the dimension of a subspace is defined. In machine learning, hyperplanes are a key tool to create support vector machines.

Let $\mathbf{w}^t = (w_1, \dots, w_p)$ be a vector of scalars and at least one of them is not zero.

Let $\mathbf{x}^t = (x_1, \dots, x_p)$ be a vector in the $\mathbb{R}^p$ the *p*-dimensional space of real numbers.

The set $\{\mathbf{x} \in \mathbb{R}^p : \mathbf{w}^t\mathbf{x} = c\}$ for a constant *c* is a hyperplane. In other words, a hyperplane is a subspace of $\mathbb{R}^p$.

a. $\mathbf{w}^t\mathbf{x} = \sum_{j=1}^{p} w_j x_j$ is the inner product of the vectors $\mathbf{w}$ and $\mathbf{x}$.

If *p* = 1, then this point $x_1 = c/w_1$ is a hyperplane.

If *p* = 2, then this line $w_1x_1 + w_2x_2 = c$ is a hyperplane when both $w_1 \neq 0$ and $w_2 \neq 0$.

b. If $w_1 = 0$ or $w_2 = 0$ (but not both), then this hyperplane reduces to a point.

If *p* = 3, then a plane $w_1x_1 + w_2x_2 + w_3x_3 = c$ is a hyperplane when $w_1 \neq 0$, $w_2 \neq 0$, and $w_2 \neq 0$.

c. If one of the scalars is zero, then this hyperplane reduces to a line.

d. If two of the scalars are zero, then this hyperplane reduces to a point.

In general, a hyperplane lies in a subspace which is at least one dimension less than that of $\mathbb{R}^p$.

Intercept = [-2.11236983]

Weight Coefficients = [[-0.83531774 -0.65083327 -0.9685864 0.26397236]]

X1 = values of volatile_acidity

X2 = values of chlorides

X3 = density

X4 = alcohol

Hyperplane = (-2.11236983) + (-0.83531774)(volatile_acidity) + (-0.65083327)(chlorides) + (-0.9685864)(density) + (0.26397236)(alcohol)

f) (10 points). When the attributes are at their overall means, what will be the predicted category for quality_group? List the attributes' overall means with your answer.

['volatile_acidity', 'chlorides', 'density', 'alcohol']

Overall means:

volatile_acidity	chlorides	density	alcohol
0.339666	0.05603386	0.99469663	10.49180083

Quality_group = 0

g) (5 points). When the attributes are at their overall 25th percentiles, what will be the predicted category for quality_group? List the attributes' overall 25th percentiles with your answer.

['volatile_acidity', 'chlorides', 'density', 'alcohol']

Overall means:

volatile_acidity	chlorides	density	alcohol
0.23	0.038	0.99234	9.5

Quality_group = 0

h) (5 points). When the attributes are at their overall 75th percentiles, what will be the predicted category for quality_group? List the attributes' overall 75th percentiles with your answer.

['volatile_acidity', 'chlorides', 'density', 'alcohol']

Overall means:

volatile_acidity	chlorides	density	alcohol
0.4	0.065	0.99699	11.3

Quality_group = 0