

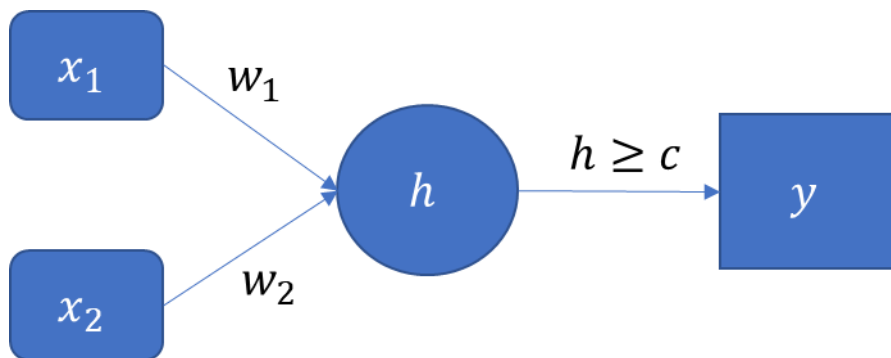
Diego Martin Crespo A20432558

Question 1 (40 points)

Logical operators (i.e., NOT, AND, OR, XOR, etc.) are the building blocks of any computational device. Logical functions return only two possible values, TRUE or FALSE, based on the truth or false values of their input values. For example, the operator AND returns TRUE only when all the input values are TRUE. If at least one of the input values is FALSE, then it returns FALSE. If we denote TRUE by 1 and FALSE by 0, then the logical AND function can be represented by the following table:

x_1	0	0	1	1
x_2	0	1	0	1
$x_1 \text{ AND } x_2$	0	0	0	1

This function can be implemented by a perceptron with two binary inputs:



The activation function for the Output layer has this form: $y = \varphi(h) = 1$ if $h \geq c$. Otherwise, $\varphi(h) = 0$.

- (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
 - $C = 2$
 - $W1 = 1$
 - $W2 = 1$
- (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:

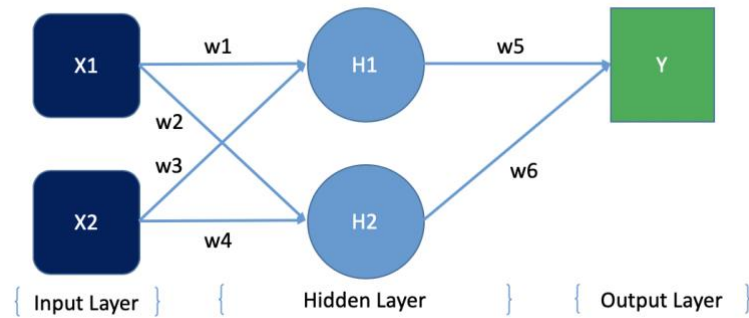
x_1	0	0	1	1
x_2	0	1	0	1
$x_1 \text{ OR } x_2$	0	1	1	1

- $C = 1$
 - $W1 = 1$
 - $W2 = 1$
- (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 six 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.

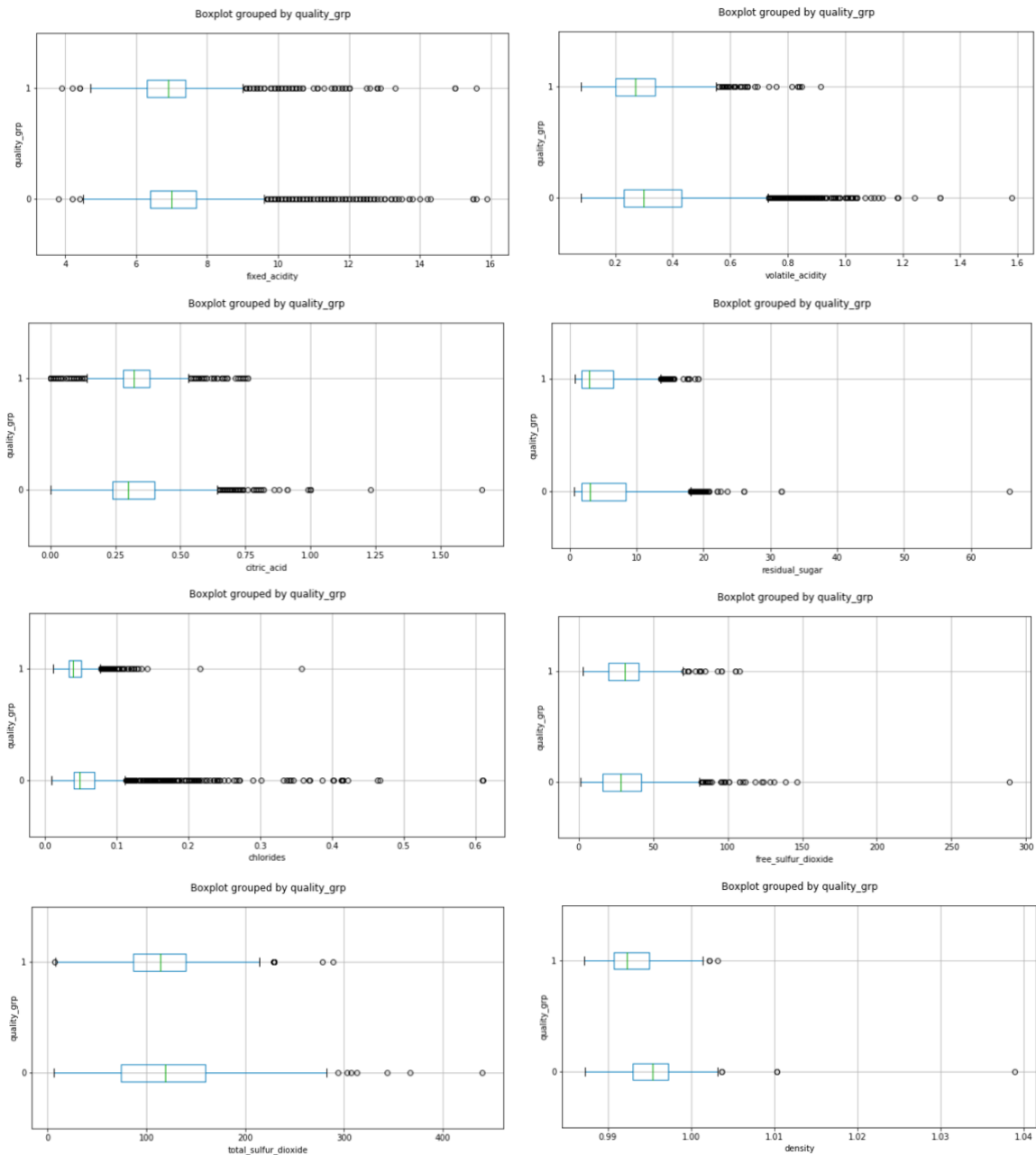


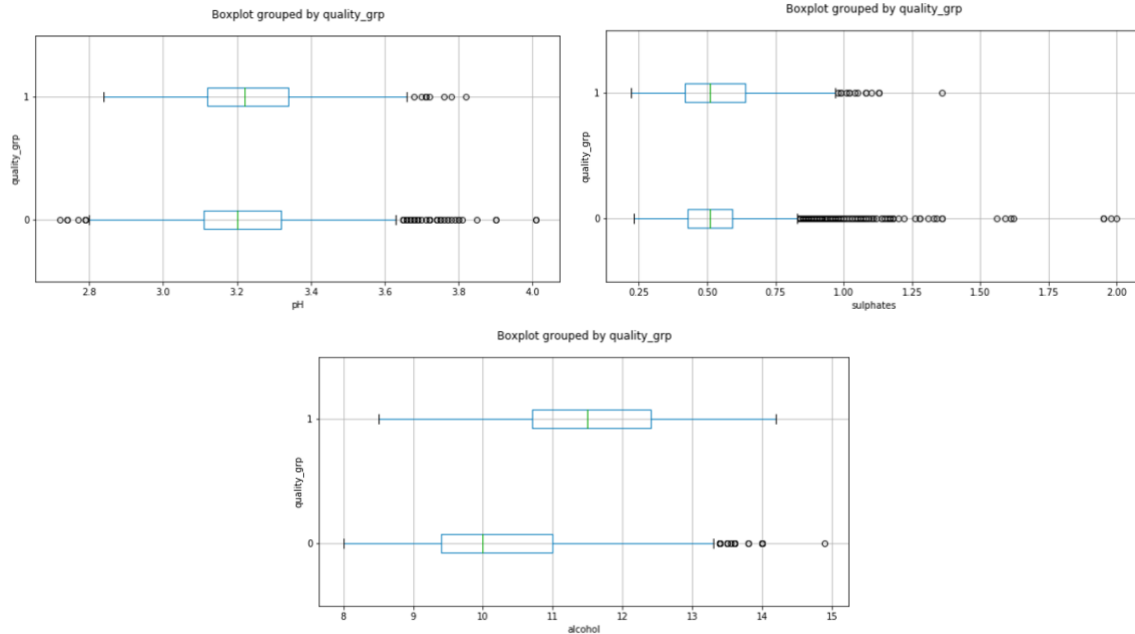
- $W_1 = 1$
- $W_2 = 1$
- $W_3 = 1$
- $W_4 = 1$
- $W_5 = -1$
- $W_6 = 1$
- $C_1 = 2$ (threshold h_1)
- $C_2 = 1$ (threshold h_2)

Question 2 (60 points)

The Center for Machine Learning and Intelligent Systems at the University of California, Irvine manages the Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>). We are going to analyze the Wine Quality dataset which is made available in the WineQuality.csv. The target variable is `quality_grp` which is binary. Its values are 0 or 1. The input attributes are: (1) `fixed_acidity`, (2) `volatile_acidity`, (3) `citric_acid`, (4) `residual_sugar`, (5) `chlorides`, (6) `free_sulfur_dioxide`, (7) `total_sulfur_dioxide`, (8) `density`, (9) pH, (10) sulphates, and (11) alcohol.

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





- 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.

	Statistic	Pvalue
free_sulfur_dioxide	-1.190230528243533	0.23399929911798764
pH	-2.2694471336134794	0.023273793852869526
sulphates	-2.739347339370487	0.006172931956782703
citric_acid	-4.394284280466727	1.129060811556467e-05
total_sulfur_dioxide	4.133779288121243	3.613375403719096e-05
fixed_acidity	3.9899182765257244	6.6822962414701e-05
residual_sugar	5.167800398741869	2.4393222593072283e-07
volatile_acidity	12.370043858067653	9.327180422012432e-35
chlorides	13.212272683932138	2.3909892435647046e-39
density	23.0914289659322	1.8902613849883943e-113
alcohol	-34.18376617265789	1.179668101759697e-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?

The algorithm will converge for first time leaving the last four attributes of the table in b): volatile_acidity, chlorides, density and alcohol.

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

0.8123749422810528

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

- The intercept term= -2.11236956
- The weights of the coefficients= -0.83531774, -0.65083345, -0.96858641 and 0.26397233.
- Therefore, the hyperplane:

-2.11236956 = -0.83531774*X1 - 0.65083345*X2 - 0.96858641*X3 + 0.26397233*X4
(With X1=volatile_acidity, X2=chlorides, X3=density, and X4=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.

Index	volatile_acidity	chlorides	density	alcohol
0	0.342268914	0.0569906437	0.994913088	10.3712502
1	0.285520134	0.0361308725	0.99019396	12.9994966

Predicted category is 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.

Index	alcohol	chlorides	density	volatile_acidity
0	9.5	0.039	0.9927	0.23
1	12.7	0.029	0.9892225	0.22

Predicted category is 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.

Index	alcohol	chlorides	density	volatile_acidity
0	11.2	0.066	0.99707	0.41
1	13.3	0.039	0.99098	0.34

Predicted category is 0.