Lab 4

Logistic Regression

Muhammad Ahmad Hesham Sec: 2 BN: 16

Muhammad Alaa Abdel-Khaleq Sec: 2 BN: 22

## Q1) Write the variable pairs that are not correlated at all to each other.

(Price, Age) and (Price, Income)

## Q2) Are there any highly correlated variables in this dataset?

No because all pair have a low correlation value

Q3) How many categories are there for the Price variable?

There are 3 categories for price (10, 20, 30)

Q4) Why it is divided into two entries only in model?

Because it is 3 categories so it is divided into n-1 variables which is 2

Q5.1) Write the value of AUC.

Area under the curve = 0.91527

Chart

Description automatically generated

Q5.2) What is the maximum value of AUC (ideal case)?

The maximum AUC is 1 so it maximizes the TPR and minimizes FPR

Q6) What does each point in the ROC graph represent? In other words, what is the value that changes and drives TPR and FPR to change too from one point to another in the graph?

The threshold of classification either positive or negative so a good model when move from left to right the TPR approach values near 1 with only a small change in FPR.

Q7) How is the predicted probability affected by changing only Price holding all other variables constant?

It is getting lower because as the price increase the probability of purchasing something more expensive is getting lower

Q8) How is the predicted probability affected by changing only Age holding all other variables constant?

It is getting higher which also make sense because when a client is old.

Chart, scatter chart

Description automatically generated

Q9) How is the predicted probability affected by changing only Income holding all other variables constant?

It is getting higher as the income increase and this is making sense because when the income is getting higher a client would have more money and would purchase with a high probability

Chart, scatter chart

Description automatically generated