Sara Ahmed
Sec: 1
B.N.: 27

Lab6 - Logistic Regression

1. Write the variable pairs that are not correlated at all to each other.

- Price and Income
- Price and Age

2. Are there any highly correlated variables in this dataset?

No

the correlation is between Age and Income = 0.09612083

3. How many categories are there for the Price variable?

Three categories that are represented by two indicator variables (dummy variables).

4. Why is it divided into two entries only in the model?

It is divided into two entries only in the model because the third category can be represented using the two variables

as.factor(Price)20	as.factor(Price)30	category
0	1	as.factor(Price)30
1	0	as.factor(Price)20
0	0	Third category->10

5.1. Write the value of AUC.

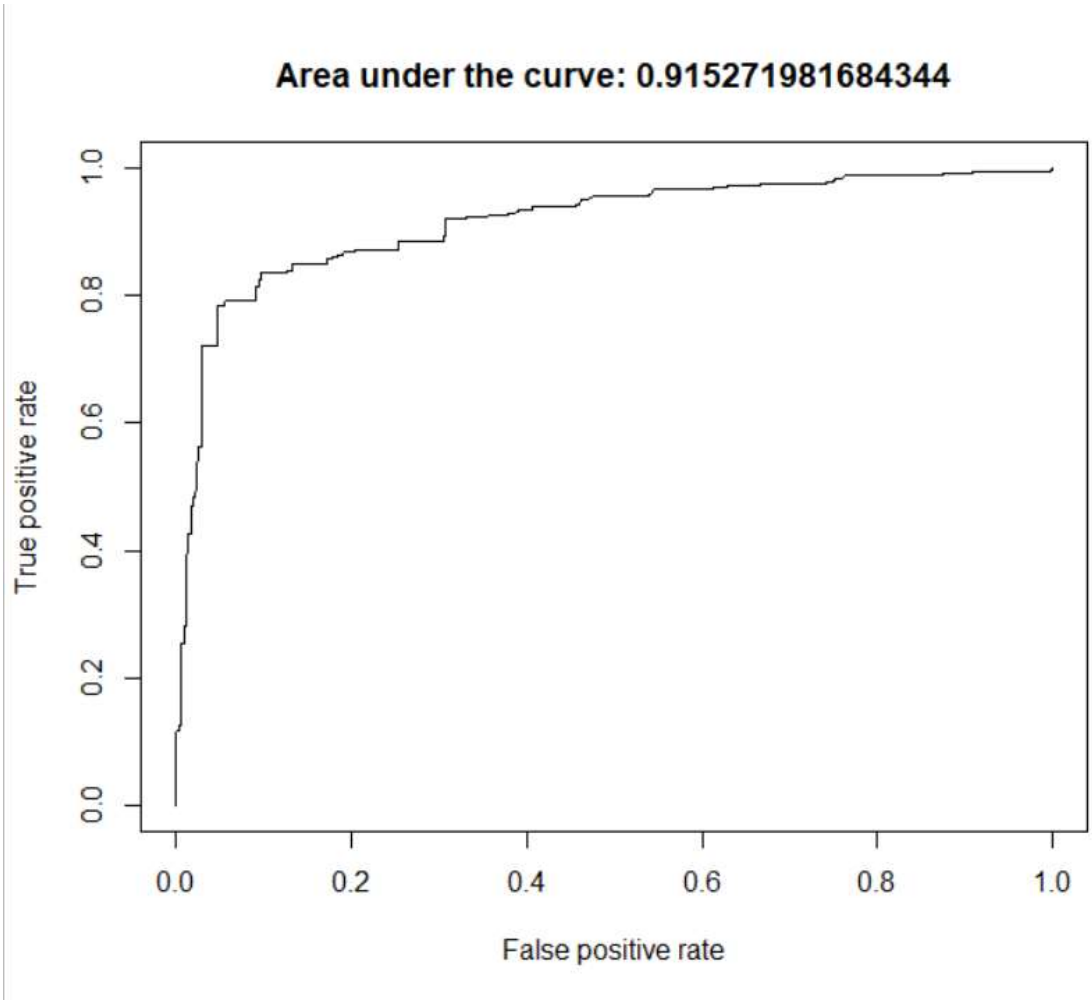
0.915272

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

1

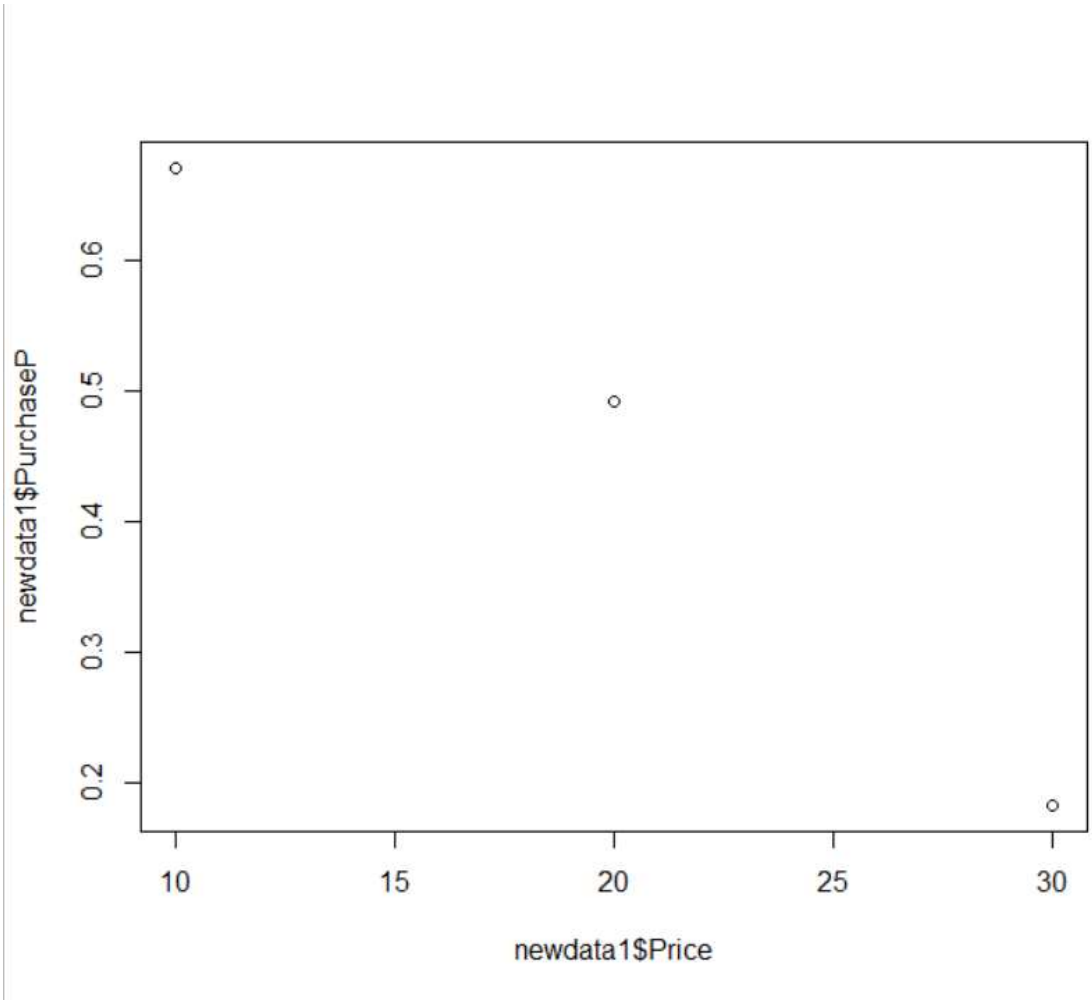
6. What does each point in the ROC graph represent? In other words, what is the value that changes and drives TPR and FPR to change too from one point to another in the graph?

The value of the threshold above which the data point is predicted as positive and below which the data point is predicted as negative.



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

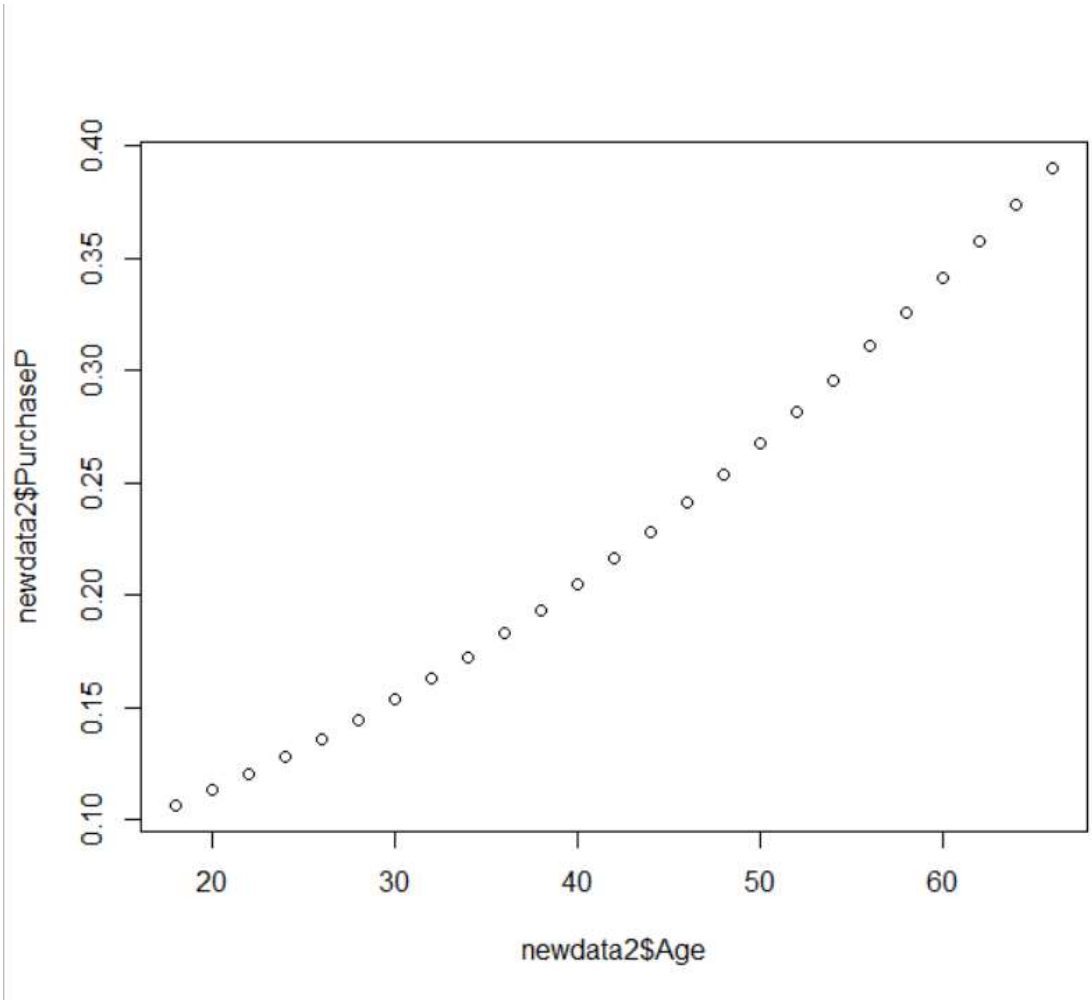
As the price increases the predicted probability of purchase decreases.



-ve correlation

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

As the age increases the predicted probability increases.



+ve correlation

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

The relation is like a sigmoid function.