Cairo University
Faculty of Engineering
Computer Engineering Department

CMPN451
Spring 2020

Data Mining, Big Data and Analytics.

Lab 6 – Logistic Regression

Note 1: At the end of the lab, you should submit a document containing **all plots** as well as full and clear answers to **non-coding questions.** The answers and plots should be in the same order as questions.

Note 2: The grade of the lab will be based on the submitted document at the end of the lab.

Requirements:

1.	Write the variable pairs that are not correlated at all to each other.
2.	Are there any highly correlated variables in this dataset?
3.	How many categories are there for the **Price** variable?
4.	Why is it divided into two entries only in the model?
5.	1. Write the value of AUC. 2. What is the maximum value of AUC (ideal case)? **Note:** For this part, you may need to search and read about the **ROC curve**.
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?
7.	How is the predicted probability affected by changing only **Price** holding all other variables constant?
8.	How is the predicted probability affected by changing only **Age** holding all other variables constant?
9.	How is the predicted probability affected by changing only **Income** holding all other variables constant?