OBAFEMI AWOLOWO UNIVERSITY	EEE 526 Machine Learning and Intelligent Control	Dr. Tokunbo Ogunfunmi
Final Exam (In-Class, Open-Book, Online Exam)		

Name:

KOMOLAFE ELISHA AYOBAMI

I will not give or get unpermitted aid in this exam. and I will report any honor code violations observed by me.

Signature:

← Your signature means you really

Grades:

	Possible Score	**Your Score**
Problem 1	100	
Problem 2	100	
Problem 3	100	
TOTAL	**300**	

- The Exam has a duration of up to 3 hours. Please upload your solutions (Jupiter notebook is preferred and PDF) on Google Classroom.
- You are allowed to ***use class notes and the texts*** prescribed for the course. All Python program should be written in Jupiter notebook.
- You cannot use Google search on the internet during the test and all work turned in must be your own. We expect strict adherence to the Obafemi Awolowo University Honor Code

Problem 1 Investigating the Space Shuttle "Challenger" Accident using Logistic Regression [100pts.]

On January 28, 1986, the Space Shuttle "Challenger" exploded and disintegrated over the Florida skies about 90 seconds after take-off from Cape Canaveral, Florida. (See YouTube video and President Reagan's, speech). It was on a mission to the International Space Station. The first elementary school teacher to go to space was on board. All seven astronauts on board perished in addition to the science experiments, satellites, etc. on board.

O-rings in the booster rockets used in space launching play an important part in preventing rockets from exploding. Probabilities of O-ring failures are thought to be related to temperature[1].

Each flight has six O-rings that could be potentially damaged in a particular flight. The data from 23 flights are given in Table 1 below can also be found in the attached data file SpaceShuttle.csv.

For each flight, we have the number of O-rings damaged and the temperature of the launch.

Index	# Damaged	Temperature
1	2	53
2	1	57
3	1	58
4	1	63
5	0	66
6	0	67
7	0	67
8	0	67
9	0	68
10	0	69
11	0	70
12	0	70
13	1	70
14	1	70
15	0	72
16	0	73
17	0	75
18	2	75
19	0	76
20	0	78
21	0	79
22	0	81
23	0	76

[1] A detailed discussion of the background of the problem is found in The Flight of the Space Shuttle Challenger in Chatterjee and Simonoff (1995, pp. 33-35).

(a) Fit a **logistic regression model** connecting the probability of an O-ring failure with temperature.
What are the coefficients ? Interpret the coefficients.
What is the Prob(Failure) / Prob(No Failure) for each degree increase in temperature ?
[30pts]

(b) The data for Flight 18 that was launched when the launch temperature was 75 degrees Fahrenheit was thought to be problematic, and was deleted.
Fit a logistic regression model to the **reduced data set**.
What are the coefficients ? Interpret the coefficients.
What is the Prob(Failure) / Prob(No Failure) for each degree increase in temperature ?
[30pts]

(c) From the fitted model in (b), **find the probability** of an O-ring failure when the temperature at launch was 31 degrees Fahrenheit. This was the temperature forecast for the day of the launching of the fatal Challenger flight on January 28, 1986.
[20pts]

(d) Would you have advised the launching of the Space Shuttle on that particular day?
[20pts]

Problem 2 Multiple Linear Regression [100pts.]

The scores of exams in one OAU EEE 526 class is shown in the dataset ExamScores.csv
This dataset contains 25 individual student scores on all exams that are considered carefully by the professor in assigning a final grade.
Here are the data columns:

X1 = Exam #1 Score,
X2 = Exam #2 Score
X3 = Exam #3 Score
X4 = Final Exam Score

The dataset is contained in the file "ExamScores.csv".

(a) Produce some **numerical and graphical summaries** of the ExamScores data. Does there appear to be any patterns? **[10pts]**
(b) Use a Python program to fit a **multiple linear regression** model to the data. **[20pts]**
(c) Which predictor(s) are significant (if any) ? **[10pts]**
(d) Determine whether **multicollinearity** is a potential problem in this model? **[10pts]**
(e) Devise a hypothesis and state the relevant null and alternative **hypothesis**. **[10pts]**
(f) What is the conclusion you can draw based on the linear regression results ? Is your hypothesis rejected ? **[20pts]**
(g) Create **one forecast (prediction)** using your model. Comment on how the results from your model compare with observed results. **[20pts]**

Problem 3 Short Answer Questions [100pts.]

- In what cases would you use a Logistic Regression model instead of a Linear Regression model ? [20pts]

- Explain why we need re-sampling methods ? [10pts]

    - Give three examples of re-sampling methods ? [5pts]

    - Which re-sampling method would you use for an application and why ? [5pts]

- Explain what is a Support Vector Machine (SVM). [20pts]

- Differentiate between K-Means Clustering and K-Nearest Neighbor (KNN) [20pts]

- Differentiate between Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) [20pts]