Assignment#1(Intro to ML)
15 Points
Due Date: 31 Jan 2023

Instructions:
Answer all four questions.

1. Is it a good idea to stop mini-batch gradient descent immediately when the testing error goes up? Explain with suitable diagram. ---- (2 point)

2. Compute the **first iteration** of **mini-batch** gradient descent for Simple Linear Regression model for the following dataset. Assume the learning rate to be 0.01. Assume the number of observations in each mini-batch to be two. Assume initial value of $\Theta_0 = 0$ and $\Theta_1 = 0$ ----(3 points)

x	y
1	1
2	3
4	3
3	2
5	5

Hint: Use the below formulae.

repeat until convergence: {

$$\theta_0 := \theta_0 - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \sum_{i=1}^{m} \left((h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} \right)$$

}

3. **Data Analysis:** Even as a ML engineer you will have to perform some kind of data analysis.
Analyse the **crime.csv** dataset. The dataset shows the crimes happened in a city say "XYZ" over a period of time.
The dataset contains the following variables:

Year, Population, Murder, Forcible_Rape, Robbery, Aggravated_assault Burglary, Larceny_Theft, Vehicle_Theft.

Violent_Crime is the sum of Murder, Forcible_Rape, Robbery and Aggravated_assault
Property_Crime is the sum of Burglary, Larceny_Theft and Vehicle_Theft

Find the most dangerous and the safest decade to live in this country? ----(5 points)

4. **Data Preparation:** You are given a **student_records.csv** dataset. The dataset shows whether a student will receive a recommendation letter or not from the college based on several attributes.
   Data preparation is the process of cleaning and transforming a dataset to make predictions accurately through using ML algorithms. Perform data preparation on student_record.csv so that it can be used to later build a Logistic Regression model for predicting if a student will receive a recommendation or not? {Assume **student_records.csv** is your training dataset. Students are not required to do the model building, They are just to perform Data Preparation} ----(5 points)

Submission Format: In the DC Connect, post the following:

1. A word document with answers to question 1 and 2.
2. A ran jupyter notebook file for question 3. Use the markdown cell in jupyter notebook to explain the question and your answer to that question. Use the Code cell in jupyter notebook to write the python code.
   [Any submission other than the format of a notebook file(.ipynb) will be graded to zero.]
3. A ran jupyter notebook file for question 4. Use the markdown cell in jupyter notebook to explain your thought process and the reasons of why you have decided to perform specific data preparations. Use the Code cell in jupyter notebook to write the python code.

Assignment Rubric:

	Exceeds Expectations (9-10 points)	Meets Expectations (7–9 points)	Approaches Expectations (5–7 points)	Fails to meet Expectations (0-5 points)
Assignment Criteria	Assignment guidelines have been followed completely.	Assignment guidelines have been followed but 1 or 2 items missing.	Assignment guidelines have been followed more than two items.	Assignment guidelines have not been followed.
Organization of submitted documents	Assignment is exceptionally well organized.	Assignment is thoughtfully organized.	Some order/organization to submission but still some areas are unclear.	Assignment is disorganized and hard to follow.
Communication and Presentation of submitted documents	Superior communication	Reasonably understandable	Understandable with minor effort	Difficult to understand
Approach	The approach is well defined, clearly explained in detail and well formatted	The approach is well-defined, high-level explanation is given and well formatted	The approach has some flows, high level explanation is given and not well formatted	The approach has many flows, no explanation is given and not well formatted
Code	Code is run successfully, well documented, and complete	Code is run successfully but no documentation provided	Code has some flows and no documentation	Code is missing

Academic Integrity and Late submission:

Assignments are due by the due date announced in class and posted on DC Connect. At his or her own discretion, and depending on the nature of the assignment, each professor will provide a facility for the submission of late assignments up to a maximum of 72 hours after the assignment due date. All allowed late submissions will be assessed a penalty of 25% of the total possible grade for the

assignment. Assignments should be submitted on time, on a regular basis, to enable you to stay on track within the class.

Any violation of academic integrity will not be accepted and will be given a grade of zero (0) and reported. Find more information on academic integrity here

https://durhamcollege.ca/mydc/learning-resources/academic-integrity