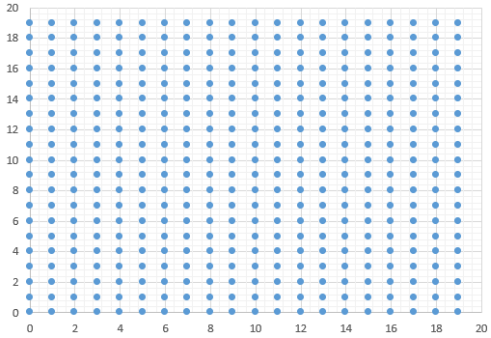
CSE 6363 - Machine Learning
Fall 2024
Due Date: November 23, 2024, 11:59 PM

Data Set: Use the dataset given with the project.
Assignment Algorithm: Logistic Regression with gradient descent
Submission: Submit both a jupyter notebook and .py version of your program. The jupyter notebook version should display all your graphs.
The Parking lot before the earthquake. The blue dots represent parking spots where cars can be parked.



There are **400** parking spots in the **Town Square drive-in movie theater**, a vast grid stretching from 0,0 to 19,19 of parking spots. The night is unsettlingly still, as the clock inches past midnight, but disaster has already struck. A violent magnitude 8 earthquake ripped through the heart of the city just moments ago. The ground beneath the once serene theater has shattered, splitting the lot into two treacherous halves.

No power. No light. Only chaos.

For over an hour, no sound has come from many of the vehicles scattered across the lot. Their signals, extinguished by the unforgiving darkness, have left only a chilling assumption — the occupants are injured, or worse, dead. Their records show a cold, **unyielding 0** in the third column of the **P3input2024_pre.txt** file.

Now, using the X and y coordinates of the parked cars and the times of the last received signals, you must uncover the mystery. The crack in the open-air theater floor — it can be found – even in the pitch black darkness, mathematically mapped to reveal the precise fault line that divides the lot. But more crucially, your model holds the key to identifying which cars, still trapped in the darkness, might hold survivors. The fate of those lost in the night rests in your hands. The cars parked to the north of the split have survived but those to the south have not – the data indicates.

The data records include X, y, label where label indicates a zero (meaning no signal received) or the time at which signal is received from the car occupants.
The Fire marshal organizing emergency response efforts has asked two teams to assume that equation of the crack could be based on: **$Y= a_1 + a_2X + a_3X^2+ a_4\sin(X)$**

The question remains—could there be multiple equations representing the unpredictable path of the crack caused by the earthquake? The teams must act fast.

Logistic Regression:

Q1 Consider the classification task using logistic regression.

a) Implement a logistic regression learner to solve the problem for the 2-dimensional data provided to find the decision boundary where a clean rift has occurred due to the earthquake. The first two columns in the data file **P3input2024_pre.txt** are the input features while the last column represents the time at which the car occupants signaled that they are alive.
Prepare the file to generate **P3input2024.txt. Split the data into 80:20 ratio to create P3input2024_train.txt (80% data) and P3input2024_test.txt** (20% data). Use 200,000 iterations and a learning rate of 0.1 for each of the stated gradient descents. Learn the parameters(thetas) using the following with the training data:

a. Batch gradient descent
b. Mini batch gradient descent **(batch size of 20)**
c. Stochastic gradient descent

b) Observe the **time taken** to complete 300,000 iterations using the three types of gradient descent algorithm with a learning rate of 0.1.
c) State the accuracy of predictions for all the test datapoints as you use the three types of gradient descent with 300,000 iterations and a learning rate of 0.1.

Q2 Implement a program to plot the Logistic Regression **decision boundary** and the **data points.** Display the data points using **two different color** to indicate their different classes.

Q3 Implement the logistic regression learner using **batch gradient descent.** Plot the loss on the Y-axis and number of iterations on the X-axis. Plot, mark, and save the graph in files for 200,000 iterations and a learning rates of:

a. 0.00004
b. 0.00006
c. 0.00008
d. 0.0001
e. List the thetas (parameters for this function)

Some rules to follow:

1. **Handwrite, sign, and date (with date of submission) a copy of the Honor Code (shown below) and share the image as part of your project; a handwritten, signed, and dated (with the date of submission) copy of the Honor Code must be included with every project and exam submission. (Failing to include will cost 20 points)**
2. **Students are required to NOT share their project questions and solutions even after the semester is over or even after graduation. However, they can show their projects during their interviews. They are also required to not discuss the solution with others or use anyone else's solution. Any violation of the policy will result in a 0 for this project for all students concerned. Violations can cause students to lose their degree.**

HONOR CODE

I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.
I promise that I will submit only work that I personally create or that I contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code
I will not participate in any form of cheating/sharing the questions/solutions.