CSE 6363 - Machine Learning

Fall 2024

Due Date: October 12, 2024, 11:59 PM

Data Set: Use a parking lot dataset given with the project in **P2input2024.txt The dataset is comprised of a pair of values:** 24-hour time in decimal notation and corresponding open parking spots at that time.

Assignment Algorithm: Linear Regression

You need to design an AI based predictive model for a parking lot. The overnight parking lot opens its doors at midnight and allows parking until noon. As per the data provided to you by the parking lot owner, so far the lot has stayed open only for 12 hours each day. The owner wants to become more profitable. You now have only the data provided by the owner and no other additional information. Your AI-based solution needs to answer the prediction problems.

Consider a simple fitting problem in the **time domain** where we look to find the best fit of the data using linear regression with a set of linear, polynomial, and sinusoidal terms. The following equation represents the equation you will use to relate the first and the second column in the data from the data file. The first column takes a value of time "**t**" ranging from 0 to 12 hours in increments of 0.25 and predicts "y", the number of available parking slots (second column) at any given time- ranging from 0 to 100 during the 12 hours on a weekday. Time "**t**" is represented in hours and starts from 0 at midnight and each whole number represents a full hour. For example, 3.25 represents 3:15 AM while 13.75 would represent 1:45 PM. All cars have to leave latest by 11:00 PM and the lot opens again at midnight for new cars.

$y(t)=a0 - (a1*t + a2*t^2 + a3*t^3) - b\sin(c*t)$ where $c = \pi/8$ The constant term is the full capacity of the parking lot.

The polynomial models a general decrease in availability over time, peaking when parking demand is highest and ebbing when the demand is lower. Sinusoidal term models the daily cyclic demand for parking and accounts for peak hours during busy neighboring activity times and non-peak time reductions in demand for parking spots. This means that this prediction problem can be solved using linear regression as the function is linear in terms of the parameters **Θ = (** a0, a1, a2, a3, b). In other words, Θ is a vector of a0, a1, a2, a3, and b.

Q1 Implement a linear regression learner to solve this best fit problem for 1 dimensional data. You can use scikit-learn and other libraries. Compute the MSE.

Q2 Implement Linear Regression from scratch.

a. Implement a linear regression learner to solve this best fit problem for 1 dimensional data in Python using only NumPy, math, matplotlib libraries. Compute the MSE.
b. If the garage stays open beyond noon, at what time will it become fully utilized (no parking spots available)? Use the learnt model to predict this time.
c. Plot the t and the y_pred (predicted number of available spots) till the parking lot gets full. Additionally plot with it the scatter plot for the original data from P2input2024.txt. It shows the real and the predicted data and a clear visual on any discrepancy between them.

Q3 In Q2 you used linear regression to compute the values of a0, a1, a2, a3, a4, and b. Vary the coefficients as follows and find the MSE to find the best set of coefficients for the most accurate prediction.

a. Try a0: 100 a1: 0,1 a2:0.2, 0.4 a3: 0, 4 b:20, 40 and compute MSE for each set of parameters.
b. Sort and list the sets of parameters by MSE so that the parameters with the least MSE value are listed first and so on.
c. Use the 4 sets of parameters with the least amount of MSE and plot the t (X-axis) and y-pred (Y-axis).
d. If the owner makes the parking lot bigger to increases the parking spots to 200 (a0=200 now), at what time will the 200 parking spots be occupied if you were to use a0=200 and the other parameters learnt in Q3.a and Q3.b. These are the set of parameters of the model that have the least MSE.
e. Plot the t and y in Q3.d to indicate when the 200-spot parking lot will be full.

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.