Homework Assignment 1: Introduction to ML and programming refresher

Due Friday, September 8th, 2023 at 11:59pm

Description

This problem set is really just to make sure you can use matlab/python to generate vectors, manipulate the values, produce some output, and submit the code along with the report. For Python programmers, please use **python3.**

What to submit

Create a folder named ps1_xxxx_LastName_FirstName,

1) ps1_matlab_LastName_FirstName if you are programming in MATLAB OR
2) ps1_python_LastName_FirstName if you are programming in Python

with the following structure and contents:

ps1_xxxx_LastName_FirstName /

- output/ - directory containing output images and other generated files

  Note: Output images must be stored with following mandatory naming convention:

  ps<problem set #>-<question #>-<part>-<counter>**.png**

  Example: ps1-3-c.png (see question 3-c)

  PNG format is easier to manipulate as it is lossless, hence we recommend using it over JPEG.

- ps1.m or ps1.py - your Matlab/python code for this problem set

- ps1_report.pdf - A PDF file that shows all your output for the problem set, including images labeled appropriately (by filename, e.g. ps1-1-a-1.png) so it is clear which section they are for and the small number of written responses necessary to answer some of the questions (as indicated). Also, for each main section, if it is not obvious how to run your code please provide brief but clear instructions (no need to include your entire code in the report).

- *.m or *.py - Any other supporting files, including Matlab function files, Python modules, etc.

Zip it as ps1_xxxx_LastName_FirstName.zip, and submit on canvas.

Guidelines

1. Include all the required images in the report to avoid penalty.
2. Include all the textual responses, outputs and data structure values (if asked) in the report.
3. Make sure you submit the correct (and working) version of the code.
4. Include your name and ID on the report.
5. Comment your code appropriately.
6. Please avoid late submission. Late submission is not acceptable.
7. Plagiarism is prohibited as outlined in the Pitt Guidelines on Academic Integrity.

Questions

1- **Conceptual question**: propose one new regression problem (one that we have not discussed in class) that you can solve with machine learning, and describe how you would go about solving it. Please give a short answer (text) to the following;

a. What features (x) would you use?
b. What would the labels (y) be?
c. How would you collect data?
d. Why might the problem turn out to be challenging?

2- Repeat question 1 for a classification problem.

3- Basic operations

a. Generate a 1000000x1 (one million by one) vector x of random numbers from a Gaussian (normal) distribution with mean of 1.5 and standard deviation of 0.6. Use Matlab's (or its numpy equivalent) `randn` function.
b. Generate a 1000000x1 (one million by one) vector z of random numbers from a uniform distribution between [-1 3]. Use Matlab's (or its numpy equivalent) `rand` function.

c. Use the function `hist` (or any other equivalent function) to plot the NORMALIZED histogram of your vectors x and z.
Output: Store the histograms images as ps1-3-c-1.png and ps1-3-c-2.png
Text response: Does the histogram for x look like a Gaussian distribution? Does the histogram for z look like a uniform distribution?

d. Add 1 to every value in x, by using a loop. To determine how many times to loop, use Matlab's `size (or python's shape)` function. Time this operation (i.e., find how long does it take to execute this operation) and print the number in the code. (Use Matlab's/python's documentation to find out how to time operations.)
Text response: the execution time

e. Now add 1 to every value in the original random vector x, without using a loop. Time this operation, print the time and write it down.

Text response: the execution time. Which way is more efficient to add a constant to a long vector?

f. Define vector y whose elements are the positive numbers in z that are less than 1.5. How many elements did you retrieve? Rerun your code two times (you can comment your code for parts c – e) and compare the number of retrieved elements. Is there any difference? How to explain that?

4- Linear algebra

a. Define the matrix $A = \begin{bmatrix} 2 & 1 & 3 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$ in matlab. **Without using loops**, find the minimum value in each column, the maximum value in each row, the smallest value in A, the sum of each row in A, the sum of all elements in A, and then compute a matrix B whose size is the same as A but its elements are the square of the corresponding elements in A.

b. Use Matlab or python functions and vector/matrix representations to solve the system of linear equations: 2x + y + 3z = 1; 2x + 6y + 8z = 3; 6x + 8y + 18z = 5. (Don't use the symbolic toolbox) Text response: the resultant solution

c. Analytically compute (by hand; show your steps) and print the L1-norm and L2-norm for each of the following two vectors: x1 = [0.5 0 -1.5] and x2 = [1 -1 0]. Verify your answer against the Matlab/nUMPY function `norm`.

5- Write a function `function [B] = sum_sq_ROW(A)` that take matrix A as an input, and returs a vector B that has the same number of columns as A. Each column of output vector, B, is the summation of the squared elements of the corresponding column in A. Test your function using two different matrices, of two different sizes, and print the input and output matrices in your report.