CS489/689- HW 3

In HW3, we are going to write codes for a linear regression model from scratch to solve classification problems.

Classification problem using MNIST (hand-written digit data)

Download the MNIST data below:

MNIST_15_15.csv: http://mkang.faculty.unlv.edu/teaching/CS489_689/HW3/MNIST_15_15.csv

MNIST_LABEL.csv: http://mkang.faculty.unlv.edu/teaching/CS489_689/HW3/MNIST_LABEL.csv

There are two files: MNIST_15_15.csv and MNIST_LABEL.csv. The former file contains hand-written digit image data (n = 335, p=15*15 pixel values) and the latter has the corresponding label of digit 5 or 6. Normalize the data (by min-max normalization, i.e. divide by 255) and train a **linear model for classification** (use a threshold of 0.5). You must use 10-fold cross-validation. For inverse of (X'X), use pseudo inverse, otherwise it may cause an error. Show a table of TPR, FPR, and accuracy for each experiment of 10-fold CV and compute the average accuracy.

Submission:

You must submit the followings to WebCampus:

1. MS word file
   - Describe what you did for the homework assignment.
   - Must include a table of TPR and FPR, accuracy of 10-fold CV, and the average accuracy of the ten experiments.

2. Source code file(s)
   - Must be well organized (comments, indentation, …)
   - You need to upload the "original python file (*.py)" after changing to "*.py.txt". For example, "*.py" to "*.py.txt"

You must submit the files SEPERATELY. DO NOT compress into a ZIP file. If you fail to provide all required information or files, you may be given zero score without grading.

Rubric:

- If used a library for linear regression instead of writing code from scratch, zero will be given.
- You can use any functions or libraries other than library of linear regression. E.g., libraries for cross-validation are okay.
- No TPR, FPR, and accuracy on the word file will deduct at least 50 points.

Deadline:

You must submit HW3 by **Tuesday, March 23, 2021**. Late assignments will not be accepted.