CS697A – Topic in Computer Science – Machine Learning
Spring 2020

Assignment 4 (12.5 points)
Due date : July 28 Tuesday, 2020 at 11:00pm

PURPOSE:

To understand how linear discrimination and neural networks work.

To understand AdaBoost algorithm and classifier diversity

WHERE TO SUBMIT ASSIGNMENTS

Please submit through the class Blackboard site. Submit a zip file (studentID_HW4.zip) of the Jupyter Python notebook you used for your homework and also all your other files including the input files and directories.

POLICY:

Collaboration in the form of discussions is acceptable, but you should write your own answer/code by yourself. Cheating is highly discouraged for it could mean a zero or negative grade from the homework. If a question is not clear, please let me know (via email, during office hour or in class).

DATA:
Read:
https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits

Download the dataset from:
https://archive.ics.uci.edu/ml/machine-learning-databases/optdigits/
test set: optdigits.tes
training set: optdigits.tra

IMPORTANT NOTE: For this homework, **use only classes 6 and 9.** Delete all the other classes from the original training and test datasets. Rename class 6 as “0” and class 9 as “1” so that your algorithms work.

Questions:

Q1 [6pts]: Implement the **logistic regression (i.e. 1 layer neural network with a single sigmoidal output) algorithm** yourself in Python, use adaptive learning rate and momentum for training. Train and test 10 times, each time, start from different random initial weights and use a random subset of 80% of the training data and also start with a different initial learning rate (e.g. 0.0001, 0.005, 0.001, 0.01 etc.) and momentum (0.9, 0.95, 0.99). Report the total training and test errors for each of the 10 runs. Report also the initial learning rate and the momentum you used.

Q2 [2.5pts]: Using the 10 runs in Q1, for each feature compute feature importance as
a) Fa: the average (over 10 runs) increase in the test error when the weight for that feature is set to 0.
For each feature, compute also,
b) Fb: the variance of each feature on the test set and
c) Fc: the mutual information between the feature column and the label column.
Plot x=Fa and y=Fb, Fc
Comment on how your feature importance Fa the other measurements Fb and Fc are related.

Q3 [4pts]: Use the scikit-learn neural network implementation to train a neural network and test it using the same instances as in Q1. Decrease the test error as much as you can through selection of:
-different number of hidden layers and units,
-L1 or L2 regularization/weight decay,
-different optimization algorithms,
-feature selection