**CS 5402 Assignment 1 (PLA)**

**Due Sep 23 2016 11:00AM**

1. One linear separable data set ‘LinearSeparable.mat’ is provided, where dataset X contains 20 training samples of dimension 3 (one dimension is a constant, ‘1’) and Y is a row vector of 1 \* 20 containing the binary labels of each sample. Implement your PLA\_Pocket algorithm and compare it with the PLA algorithm on this linear separable dataset. Report your observation on the convergence and efficiency of the two algorithms on the linear separable dataset.  
Note: use load(‘LinearSeparable.mat’,’X’,’Y’) in Matlab to load the dataset.

2. One noisy data set ‘NonLinearSeparable.mat’ that is not linear separable is provided. Run your PLA\_Pocket and PLA algorithms on this dataset. Report your observation on the convergence and efficiency of the two algorithms on the nonlinear separable dataset.

3. A dataset of handwritten digits ‘zip.train’ is provided, where 7291 training samples are provided for 10 digits. The data structure of matrix ‘zip’ is 7291 \* 257 where each row is a data sample and the first column is label Y indicating the actual digit. Each digit was initially scanned as a 16\*16 grayscale image and then it was converted to a 1\*256 feature vector. We only use digit 1 and 5 in this assignment for the binary classification. You can use the following Matlab scripts to get the corresponding positive and negative samples:

load zip.train %7291 \* 257

posIdx = find(zip(:,1)==1);

negIdx = find(zip(:,1)==5);

posSamples = zip(posIdx,2:end); %digit 1: N\_positive \* 256

negSamples = zip(negIdx,2:end); %digit 5: N\_negtive \* 256

If you want to visualize a specific sample (e.g., sample 222 in the digit 5 samples), you can use

figure; imshow(reshape(negSamples(222,:),[16 16])',[]);

Now, run your PLA\_pocket algorithm on this dataset (digit 1 and 5) and obtain your best w\_pocket.

A testing dataset ‘zip.test’ is provided. Again, extract the digit 1 and 5 only. Apply your w\_pocket on the testing data samples to obtain the predicted labels. Compare them with the ground truth labels. Report your error rate.

4. A dataset of handwritten digits ‘features.train’ is provided, where 7291 training samples are provided for 10 digits. The data structure of matrix ‘features’ is 7291 \* 3 where each row is a data sample and the first column is label Y indicating the actual digit. The second and third columns represent the computed image features such as the mean intensity and asymmetry of the image.

Train your PLA-algorithm on this dataset and test it on dataset ‘features.test’. Compare your results with those from task 3.

In your codes, add enough comments so our GTA can understand it.

Upload your code and report to Canvas.