CS 557 STATISTICAL PATTERN RECOGNITION AND LEARNING
FALL 2016
ASSIGNMENT 4

DUE: October 23, 2016.
PROBLEM

1. Read the datasets in the given OCR files. This is a subset of OCR dataset taken from:
http://cmp.felk.cvut.cz/cmp/software/stprtool/index.html

Each image is a row, however, the data has been scrambled to hide the identity of each image.

2. Apply LDA on this dataset for multiple classes. Make a scatter plot of this dataset after the transformation in 2D, giving different symbols to each class. After applying LDA, use your nearest neighbor MAP classifier to classify all points and report the balanced accuracy/error rate.

3. Ignore the labels and apply MDS on this dataset for multiple classes. Make a scatter plot of this dataset after the transformation, giving different symbols to each class. After applying MDS, use your nearest neighbor MAP classifier to classify all points and report the accuracy/error rate.
Repeat the same experiment by applying MDS separately to each class and then apply the nearest neighbor classifier to classify all points. What are your observations?

4. Ignore the labels and apply LLE (embed in the 2D space) on this dataset for multiple classes and different values of k. You have to try at least 5 different values of k and make a scatter plot of at least two of them. Apply the nearest neighbor MAP classifier to classify all points for all the different values of k and report the accuracy/error rate for all of them.
Repeat this experiment by applying LLE separately to each class and then apply the classifier to classify all points. What are your observatios.

5. Submit the results on the test set at the following website (this part is mandatory for the assignment to be graded):
https://inclass.kaggle.com/c/scrambled-ocr

NOTE:
- When you make the figures, make sure you make a legend and label all axis. You can use the plot command to make a scatter plot. For example plot(x,y,'*g') will make a scatter plot in green color using the symbol '*'.
- Use pdist2, eigs, sort routines of Matlab. AVOID THE USE OF LOOPS

TO SUBMIT
1. Softcopy of all code and report on slate
2. **Hard copy** of a report which is **not more than two pages** long that describes all the results of your experiments AND YOUR CONCLUSION and COMMENTS ON THE RESULTS. In the results you have to include the results on the training set and also your public leaderboard scores for the test set.