CSE 5523. Homework 3. Due Apr 26.

Problem 1.
Apply (1) decision trees, (2) bagged and (3) boosted decision trees to the digit datasets from Homework 2. (You may use the standard libraries of Matlab or download Matlab code from the Web.) Use appropriate cross-validation on the training set. Compare performance.

Problem 2.
1. Implement PCA and apply it to the digit data, reducing the dimension to two. Visualize the data after dimensionality reduction using colors for different classes.
2. Produce pictures of "eigendigits" for the dataset combining both classes and for each class separately. Observations?

Problem 3.
Apply k-means clustering to the digits data set for $k = 2, 5, 10, 50$. How well does it identify the different digits? (Note that clustering is unsupervised – how do you compare classification and clustering results?)