

CS 445/545
Machine Learning
Winter, 2009

**Homework 1: Perceptron Learning, PCA, and Comparing
Learning Algorithms**

Due Wednesday Jan. 21, 5:00pm.

For this homework you will write Matlab code to implement the perceptron learning algorithm, and use it to train perceptrons on two data sets: (1) a set of handwritten characters; and (2) the same set transformed by principal components analysis.

For this assignment, you need to do the following:

1. Download the data from our class website. The tarball contains three files: (1) optdigits.names (a description of the data set); (2) optdigits.tra (the training set); and (3) optdigits.tes (the test set).
2. Use Matlab to run principal components analysis on this same data set to create a transformed data set, optdigits.tra_pca and optdigits.tes_pca.
3. Using Matlab, implement the perceptron learning algorithm to train 10 perceptrons, one to classify each digit (“one against all”). For each perceptron i , use the algorithm for training described in class, where the training data is labeled +1 if the class is digit i , -1 otherwise. Train one collection of 10 perceptrons for each data set (original and PCA-transformed). The learning rate η should be set to 0.2. Each perceptron should be trained for many epochs over the training data, stopping the training only when the percent error on the training set is less than or equal to 5%.
4. Once a collection of 10 perceptrons is trained, in order to classify a new instance \mathbf{x} , run each of the 10 perceptrons on \mathbf{x} . The class is the perceptron with maximum output activation (before performing the `sgn` function).
5. Do a paired-t test to determine the percent confidence that the mean accuracy of the perceptron trained on the original data is less than (or greater than) the mean accuracy of the perceptron trained on the PCA-transformed data set. Use the algorithm described in class, with ten-fold cross-validation.
6. Repeat the above steps for the PCA-transformed data, but this time use only the first four principal components to represent each instance in the training and test sets. Again, do a paired-t test to determine the percent confidence that the mean accuracy of the perceptron trained on the original data is less than (or greater than) the mean

accuracy of the perceptron trained on the reduced-dimensionality-PCA-transformed data set.

Here is what you need to turn in:

- Your Matlab code for perceptron learning algorithm and paired t-test.
- The results from steps 5 and 6 of error on transformed test data minus error on original test data, for each of the ten folds of cross-validation.
- Three plots: Choose one of the digit classes (e.g., “3”) and plot the percent error on the training set as a function of epoch number during training for (1) the original training data, (2) the PCA-transformed data, and (3) the reduced-dimensionality-PCA-transformed data set.
- A write-up with summary of paired-t results and summary statement on the outcome of your experiments.

Send these items in **electronic format** to karan@pdx.edu by the due date. No hard-copy please!

This assignment has several parts to it, so start early. If there are any questions on this assignment, don’t hesitate to ask!

Policy on late homework: If you are having trouble completing the assignment on time for any reason, please see me **before** the due date to find out if you can get an extension. Any homework turned in late without an extension from me will be graded down on grade (e.g., “A” to “B”).