CS 445/545

Machine Learning

Winter 2016

Homework 1: Perceptrons

Due Tuesday, January 19, 2016

For this homework you will write code to implement perceptron learning and all-pairs classification, and use these to train perceptrons to do a letter recognition task. Please write your own perceptron and all-pairs code; don't use existing code written by others. You may use whatever programming language you prefer.

The dataset for this task is the **Letter Recognition** data set from the UCI machine-learning repository: http://archive.ics.uci.edu/ml/datasets/Letter+Recognition

Inputs: Your perceptrons will have 16 inputs (the attributes used in the letter-recognition data) plus one bias input (always +1), and, likewise, 17 weights, and a single output unit.

Task: Implement code to perform a multi-class classification (i.e., recognize a given letter) using the all-pairs method described in class, with perceptrons.

Data: Split the data into approximately half for training and half for testing. Your split should be such that each letter has about the same number of instances in the training set as in the test set.

Preprocessing: Scale each data value to be between 0 and 1 (i.e., divide each value by 15, which is the maximum value in the original data). This will help keep the weights from getting too large.

Initial weights: Your perceptrons should start off with small ($-1 < w < 1$) random positive and negative weights.

Experiment: Using a learning rate $\eta = 0.2$, train your 325 perceptrons on the training data. E.g., for your "A" vs. "B" perceptron, you should use the "A" examples and "B" examples from the training set.

For each perceptron, repeat cycling through the training data until the accuracy on the training data has stopped improving. When you have trained all 325 perceptrons, run the all-pairs algorithm on the test data to determine the accuracy and to create a confusion matrix.

Report: Your report should be a one paragraph description of the experiment, including:

- Accuracy (fraction of correct classifications) on the test set and
- Confusion matrix on the test set

This homework's report is very short—just a paragraph!

Here is what you need to turn in:

- Your report, with all the information requested above.
- Your well-commented code.

How to turn it in (read carefully!):

- Send these items in electronic format to mm@pdx.edu by 2pm on the due date. No hard copy please!
- The report should be in pdf format and the code should be in plain-text format.
- Put "MACHINE LEARNING HW 1" in the subject line.

If there are any questions on this assignment, don't hesitate to ask me or e-mail the class mailing list.

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 have 5% of the grade subtracted for each day the assignment is late.