CSCI 59000: Machine Learning

Assignment 3

Assigned: March 27, 2018

Due: April 10, 2018

Download the liblinear package (http://www.csie.ntu.edu.tw/~cjlin/liblinear/) and install it on your machine. Liblinear has stand-alone executables but you can also use the Matlab interface provided in the package.

The following training options are available. You will use option –s 5.

-s type : set type of solver (default 1)

for multi-class classification

0 -- L2-regularized logistic regression (primal)

1 -- L2-regularized L2-loss support vector classification (dual)

2 -- L2-regularized L2-loss support vector classification (primal)

3 -- L2-regularized L1-loss support vector classification (dual)

5 -- L1-regularized L2-loss support vector classification

6 -- L1-regularized logistic regression

7 -- L2-regularized logistic regression (dual)

Use the most recent version of the competition data set for the following experiments. First split the competition data into two as train (all_ids=1,3,5,7…) and test (all_ids=2,4,6,8, ..). Normalize each feature to have range [0 1] or zero mean, unit variance.

Part 1. Using s=5 and all of the features fill out the following table. F1 score is computed on the test data.

C	F1 score	Training time	# of support vectors	# of nonzero coefficients in w (averaged over all w's)
0.01				
.1				
1				
10				

Independently investigate the correlation (if any) between the following pairs of variables.

a. F1 score vs. # of support vectors
b. F1 score vs. # of nonzero coefficients in w
c. F1 score vs. C
d. C vs. training time
e. C vs. # of support vectors
f. C vs. # of nonzero coefficients in w

Part 2. Repeat the experiment in Part 1 this time using every tenth channel as a feature. Discuss items a through f from Part 1.

Part 3. For the best configuration from Part 2 create an ensemble of SVMs (by random subsampling of 25 features) and use the mode of the predictions as your final predictions. Evaluate the performance of the ensemble for M=1, 5, 10, 20, and 50 weak learners. Plot F1 score vs M graph.