MATH 689 / CSI 747

Midterm Exam. Submit no later than October 23, 2012.

Implement your own Support Vector Machine for the hand written digits recognition using AMPL.

1) Use the discretized hand written digits data created earlier.
2) Train the primal soft-margin SVM (the one that incorporates a non-separable case) to detect 3s and 6s only. Calculate the error rate for testing examples.
3) Train the dual soft-margin SVM optimization problem. Use scalar products as kernels. Compare the results with those from (2).
4) Train the dual polynomial kernel machine with $\alpha = 0.0156$, $\beta = 0$, $d = 3$ to detect 3s and 6s only. Calculate the error rate for testing examples.
5) Train the dual radial basis function machine $\gamma = 0.0521$ to detect 3s and 6s again. Calculate the error rate for testing examples.
6) Chose the machine with the better error rate for the main experiment.
7) Run 10 SVMs to train to detect a particular digit (e.g. 2) against the rest digits (e.g. 0, 1, 3, 4, 5, 6, 7, 8, 9). In the training use the value $y = +1$ for a particular digit and $y = -1$ for the rest of them. Obtain 10 different separating hyperplanes $h_0,\ldots,h_9$ that separate each $0,1,\ldots,9$ from the rest digits.
8) While testing the digits you may find out that a particular digit may be classified not uniquely. For example, some tested digit can be on the positive side of h_3, h_5 and h_8, meaning that this digit can be classified as 3, 5 or 8. To resolve the ambiguity classify this digit as the one that corresponds to the hyperplane with the maximum classification number

$$\sum_{i=1}^{l} y_i \alpha_i^* K(x_i, x) - b$$

9) Use $C = 100$ as the penalty parameter. Increase if necessary.
10) Document all the experiments you are doing.

In the file digits.zip you will find 20 files:

10 for training: train0.txt, train1.txt,…,train9.txt
10 for testing: test0.txt, test1.txt,…,test9.txt

Each training file has 93 digits, so the total number of the training digits is 930.
Each testing file has 41 digits, so the total number of the testing digits is 410.