Introduction to Machine Learning *CMSC 678, Fall 2017*

HW3

Please note that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. You'll need *mydefs.sty* and *notes.sty* which can be downloaded from the course page.

Gradient descent algorithm:

1. Implement the gradient descent algorithm for binary SVM.

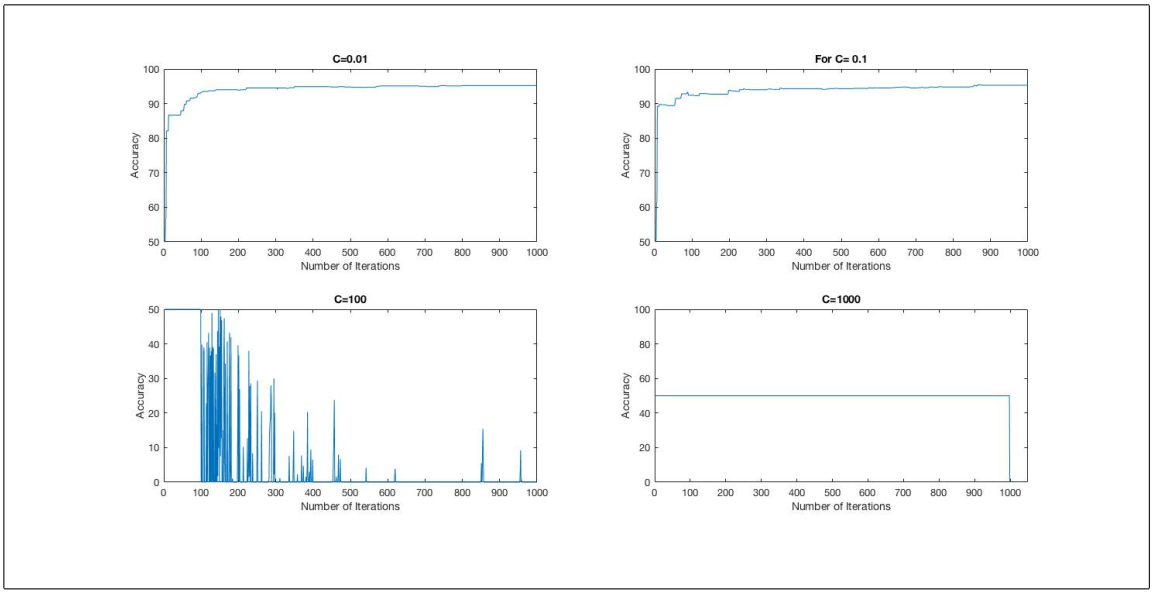
   The source for the training algorithm is SVM_Train.m

2. Similar to the previous homework, train and test it for classifying digits "1" and "6" in MNIST dataset.

   Accuracy for the digits 1 and 6 is 96.2% after 2000 Iterations

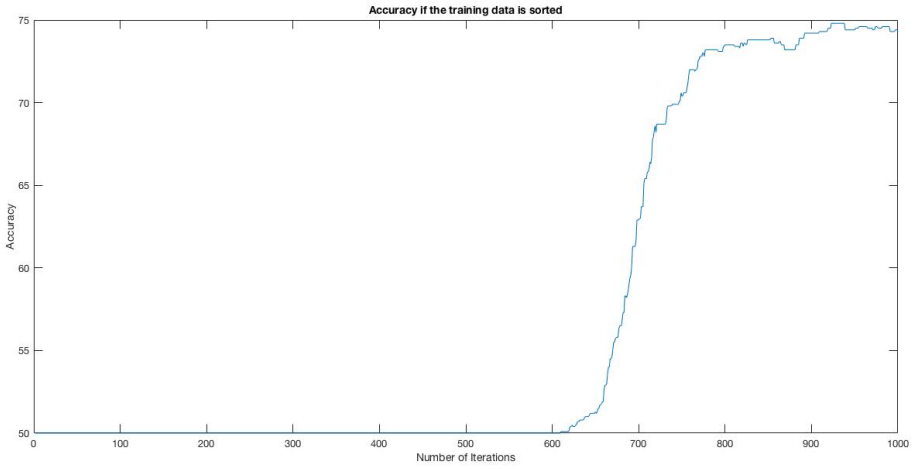
3. Plot the accuracy on the test set wrt. the number of iterations. Here, processing each data-point is considered one iteration. You don't need to use all training data if it takes a long time to run.

   (a) Play with the hyper-parameter C to see its effect in the accuracy and overfitting. For instance, try very large and very small values for it.



   (b) Choosing the learning rate may be a little tricky. One popular strategy is to reduce it promotional to $\frac{1}{t}$ where t is the iteration number.

4. Sort the data before training so that all "1"s appear before "6"s and plot the accuracy wrt iterations. Is this faster or slower in training?



5. Implement a function for 1-vs-all multi-class SVM that calls the binary SVM function.

Implemented in the HW3_Question5

6. Train and test it on all 10 digits in MNIST and report the confusion matrix as well as the average precision. $conf(i, j)$ is the number of images that are from category i and are classified into category j. You should normalize this so that all rows sum to one and then average accuracy is the average of its diagonal values.

CONFUSED MATRIX

956	0	2	1	0	1	11	1	8	0
0	1095	6	4	1	6	5	2	16	0
18	18	857	26	20	1	24	27	33	8
8	1	23	896	1	17	9	19	20	16
2	11	4	3	847	0	22	7	11	75
22	10	2	94	16	629	34	29	36	20
13	3	6	3	17	11	901	3	1	0
4	33	35	6	8	0	4	909	4	25
17	21	11	59	16	56	21	10	744	19
17	14	11	17	60	6	3	43	10	828

NORMALIZED CONFUSED MATRIX

0.9755	0	0.0174	0.0079	0.0020	0.0247	0.0136	0.0039	0.0175	0.0168
0	0.9648	0.0174	9.9010e-...	0.0112	0.0112	0.0031	0.0321	0.0216	0.0139
0.0020	0.0053	0.8304	0.0228	0.0041	0.0022	0.0063	0.0340	0.0113	0.0109
0.0010	0.0035	0.0252	0.8871	0.0031	0.1054	0.0031	0.0058	0.0606	0.0168
0	8.8106e-...	0.0194	9.9010e-...	0.8625	0.0179	0.0177	0.0078	0.0164	0.0595
0.0010	0.0053	9.6899e-...	0.0168	0	0.7052	0.0115	0	0.0575	0.0059
0.0112	0.0044	0.0233	0.0089	0.0224	0.0381	0.9405	0.0039	0.0216	0.0030
0.0010	0.0018	0.0262	0.0188	0.0071	0.0325	0.0031	0.8842	0.0103	0.0426
0.0082	0.0141	0.0320	0.0198	0.0112	0.0404	0.0010	0.0039	0.7639	0.0099
0	0	0.0078	0.0158	0.0764	0.0224	0	0.0243	0.0195	0.8206

The accuracy is 86.3472

7. Look at top mistakes and show images along with ground truth label and the predicted label to see if they make sense.