Introduction to Machine Learning	*RUJUTA PALANDE*

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

Figure 1: Accuracy plot with large C

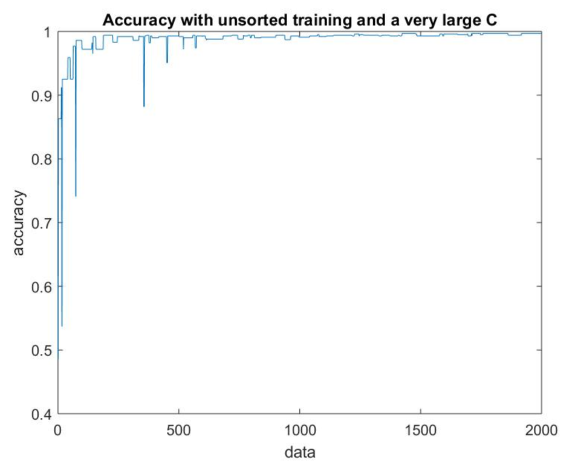
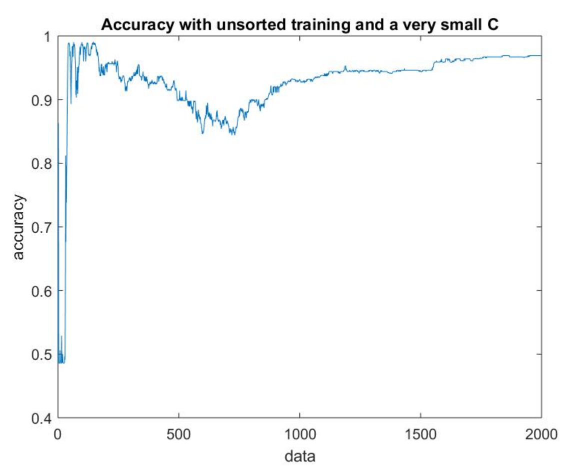


Figure 2: Accuracy plot with small C



(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?

Figure 3: Accuracy plot with large C and sorted training data

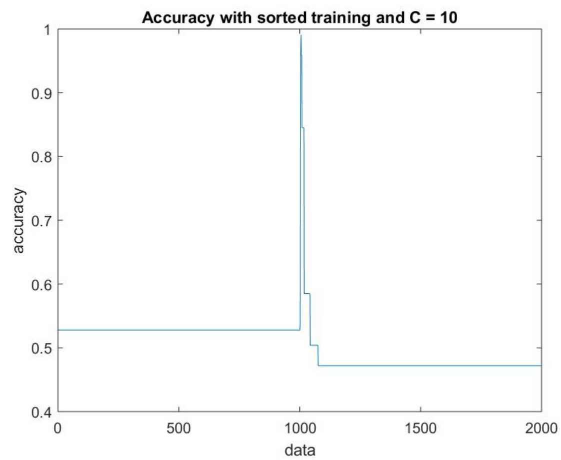
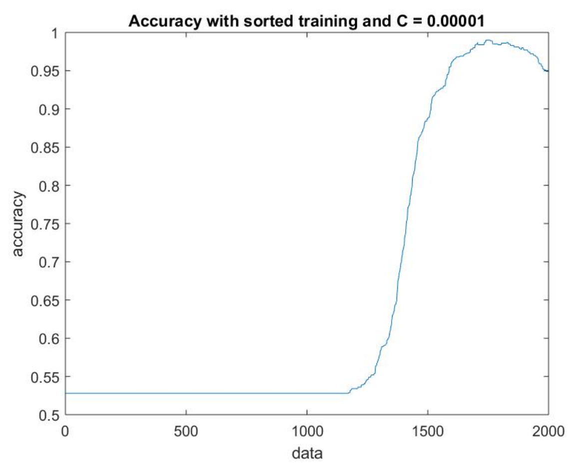


Figure 4: Accuracy plot with small C and sorted training data



`Answer : This model is slower in training. This is because the model first gets trained for 1 and when 6 appears, the accuracy rate drops and then it again re trains itself for 6.`

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

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.
`Answer : average accuracy : 0.8818`

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

Figure 5: Confusion matrix

10x10 double

	1	2	3	4	5	6	7	8	9	10
1	0.9998	0.0011	0.0021	0.0021	0	0.0011	0.0200	0.0042	0.0021	0
2	0	1.0000	0.0027	0.0018	0	0	0.0036	8.9046...	0.0018	0
3	0.1075	0.3077	0.9421	0.0221	0.0236	0	0.0442	0.0545	0.0177	0
4	0.0595	0.0790	0.0255	0.9938	0.0012	0.0024	0.0134	0.0413	0.0073	0.0036
5	0.0144	0.0874	0.0024	0	0.9948	0	0.0443	0.0120	0.0012	0.0192
6	0.6973	0.2415	0.0099	0.4780	0.0567	0.4386	0.1084	0.1380	0.0074	0.0222
7	0.1018	0.0407	0	0.0012	0.0072	0.0024	0.9939	0	0	0
8	0.0121	0.0924	0.0187	0.0011	0.0055	0	0.0011	0.9955	0.0022	0.0022
9	0.2085	0.6069	0.0188	0.2231	0.0480	0	0.0834	0.1189	0.7174	0.0063
10	0.0949	0.1310	0.0133	0.0266	0.3569	0	0.0057	0.5429	0.0019	0.7422

Figure 6: truth labels and predicted labels