Hadi Askari
Assignment 3 SVM Hand Digit Prediction

Question 1) Train a linear multi-class classification SVM with no kernel. Do not use a SVM package. Specify i) Your instance mapping function if any and ii) Your loss function (20 points)

Ans) To train a multiclass SVM I trained 10 different binary SVMs (one vs rest approach). One for each digit, the chosen class was assigned the value of 1 whereas all others were assigned -1. The SVM was soft-margin SVM and the output class with the highest score was assigned that class label while predicting. The mapping function was linear mapping as I did not use any kernel and the loss function was the hinge loss function. The implementation is present in the code file provided.

Question 2) Describe a method to estimate your performance using an empirical method. Compare this estimate with a well known theoretical bound. Explain why/if there is a difference. (5 points)

Ans) For the purposes of calculating and fine-tuning my model, I have used the 'train_test_split' method from sklearn to split the given training data into training data (90%) and validation data (10%) at random. Then:

- After building and training my model on the 90% of data, I made predictions for the 10% of training data which was hidden from the model.

- Finally, I have calculated accuracy based on correct predictions divided by correct + wrong predictions which led me to improve accuracy by modifying hyperparameters. The accuracy turned out to be **75.47%**

I will use “VC dimension”: to find a theoretical bound to the error in the given problem statement. VC dimension is given by:

$$R(\alpha) \leq Remp \ + \ \sqrt{\frac{hlog\frac{2l}{h} - \frac{n}{4}}{l}}$$

Where, h is the number of features in the given dataset plus 1. h = 9 in this case. l is the number of training examples (3372 since 90-10 split). N is a hyperparameter between 0 and 1 which I chose to be 0.5. Remp is 0.7547. Thus the theoretical bound becomes 1-0.0891=0.9109. Thus our result is within the theoretical error limits.

Another well known theoretical bound is PAC learning. The formula for that is:

$$m \geq \frac{1}{\epsilon}(ln|H| + ln\frac{1}{\delta})$$

Question 3) Submit your predictions on this test set, one prediction per line in the order given.

Ans) In zip folder.

Question 4) Implement both types of transfer learning SVM (hypothesis and instance transfer) to train 1 vs 7 (target problem) by transferring in 1 vs 9 (source

problem). We will use only 10 available 1 and 7 digits to create a more realistic transfer setting. Report your error estimate for the target problem with i) no transfer, ii) hypothesis transfer and iii) instance transfer. Which performs better? Why? (20 points)

Ans) I trained linear soft margin SVM (similar to question 1) on 1 vs 9 dataset.

- For no transfer I simply ran soft margin on the 10 1vs7 examples provided and analyzed the results on the test set.
- For Hypothesis Transfer I transferred over the weight vectors from the SVM and added them into the w in the constraints while training the 10 1vs7 examples. Then analyzed the results on the test set.

- For instance transfer I transferred the support vectors and their labels and added them to the constraints while training the 10 1vs7 examples. Then analyzed the results on the test set.

Instance transfer worked better for me since it taught the model which support vectors to focus on as they were more pivotal in helping the model in deciding the final decision boundary rather than the w vector from 1vs9.

	No Transfer	Hypothesis Transfer	Instance Transfer
Test Acc	68.568%	76.923%	78.912%

Question 5) Kernelize your approach for question 1). What Kernel do you use? What is your new error

estimate? Submit a new predictions file for the data set (20 points)

Ans) I used the polynomial kernel, with degree 2 to map the input space into a higher dimension and then performed the same procedure as before in the non-kernelized approach. My new error estimate was **92%.** The predictions are in the zipped folder.