

ENGR 421 / DASC 521: Introduction to Machine Learning
Homework 06: One-Versus-All Support Vector Classification

Deadline: December 20, 2021, 11:59 PM

In this homework, you will implement the one-versus-all support vector classification algorithm in Python. Here are the steps you need to follow:

1. Read Section 14.9 from the textbook.
2. You are given a multivariate classification data set, which contains 5000 clothing images of size 28 pixels \times 28 pixels (i.e., 784 pixels). These images are from five distinct classes, namely, T-shirt, Dress, Coat, Shirt, and Bag. The figure below shows five sample clothing images from each class. You are given two data files:
 - a. `hw06_images.csv`: clothing images,
 - b. `hw06_labels.csv`: corresponding image labels (1: T-shirt, 2: Dress, 3: Coat, 4: Shirt, 5: Bag).



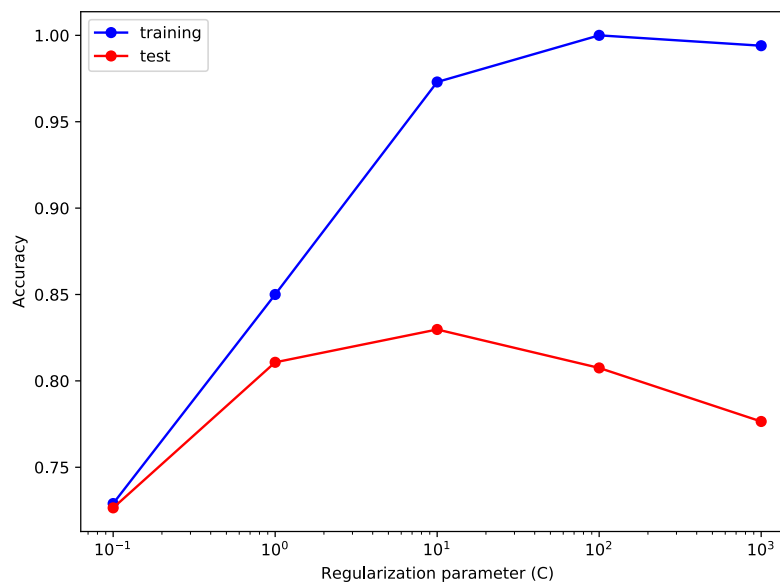
3. Divide the data set into two parts by assigning the first 1000 images to the training set and the remaining 4000 images to the test set.
4. Implement the one-versus-all support vector classification algorithm using the binary classification code given in “Lab08: Kernel Machines”. The Gaussian kernel should be used as the similarity metric between data points.
5. Calculate the confusion matrix for the data points in your training set using the one-versus-all support vector classification algorithm trained using $C = 10$ and $s = 10$. Your confusion matrix should be the following matrix.

y_train	1	2	3	4	5
y_predicted					
1	207	1	0	9	0
2	2	199	1	1	0
3	0	1	204	6	0
4	0	1	4	185	1
5	0	0	0	0	178

6. Calculate the confusion matrix for the data points in your test set using the one-versus-all support vector classification algorithm trained using $C = 10$ and $s = 10$. Your confusion matrix should be the following matrix.

y_test	1	2	3	4	5
y_predicted					
1	641	23	3	137	9
2	43	714	27	40	4
3	4	39	666	90	10
4	100	32	69	541	16
5	12	2	6	15	757

7. Learn one-versus-all support vector classification algorithms by setting the regularization parameter C to 0.1, 1, 10, 100 and 1000, and the kernel width parameter s to 10. Draw the classification accuracy for training and set data points as a function of C . Your figure should be similar to the following figure.



What to submit: You need to submit your source code in a single file (.py file) and a short report explaining your approach (.doc, .docx, or .pdf file). You will put these two files in a single

zip file named as ***STUDENTID.zip***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the zip file you created to Blackboard. Please follow the exact style mentioned and do not send a zip file named as ***STUDENTID.zip***. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.