Bogazici University

CMPE 462 - Machine Learning

Project 2 Report

TEAM **F2E**++

Enes Turan Özcan - 2015400003

Emirhan Saraç - 2019700213

Muhammet Furkan Gök - 2014400039

May 18, 2020

1 Hard Margin SVM

In this part we were asked to train a hard-margin SVM with given data. In order to train a hard-margin SVM with LIBSVM library we needed to decide on the value of the coefficient C.

As we discussed in the lectures, increasing the value of the coefficient C means we care more about violating the margin, which gets us closer to the hard margin SVM. Therefore, we picked 999999999 for C and trained the SVM. Here are the accuracies with test and training data:

	Percentage	Rate
Training Data	100%	150/150
Test Data	76.6667%	92/120

The train accuracy actually gives a clue about the correctness of the separation of data points for hard margin.

2 Soft Margin SVM

In this part we were asked to train a soft margin SVM with different values of the parameter C and with different kernel functions.

After investigating on this issue, we concluded that the performances of kernel methods are so close to each other for some C values, however for some values of C performances of kernel methods differ a lot.

Furthermore, we noticed that increasing the value of C for a fixed kernel method may increase the test accuracy up until some point. However, in most of the cases an increase in test accuracy resulted in a decrease in training accuracy.

2.1 Different Values of Parameter C with Fixed Kernel

Below figures display the accuracy of SVMs with a fixed kernel method and increasing coefficient C on both test and training data. For kernel 0, 1 and 2 we can conclude that the test accuracy increases with the increasing C value. For kernel 3, the accuracy behaves in a decreasing manner with increasing C in general. This behavior is also observed for train accuracy.

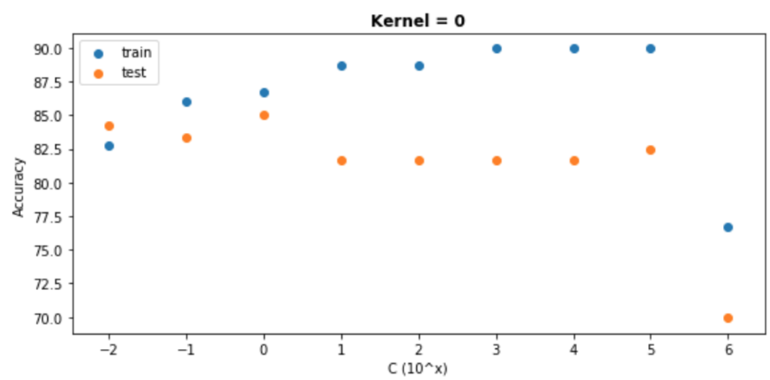


Figure 1: Kernel 0

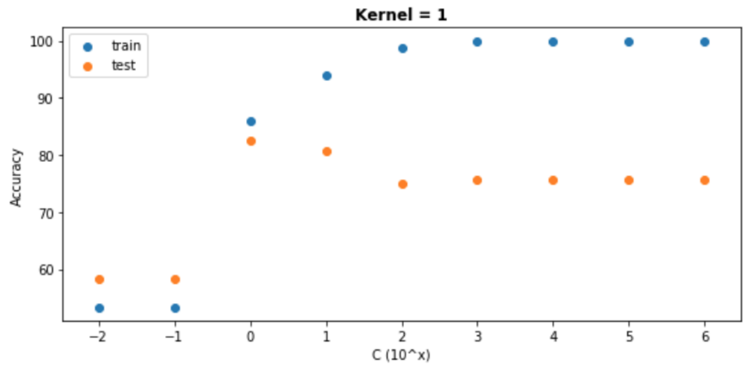


Figure 2: Kernel 1

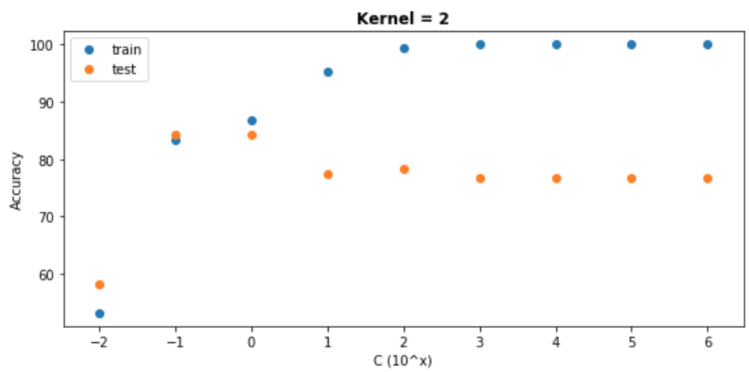


Figure 3: Kernel 2

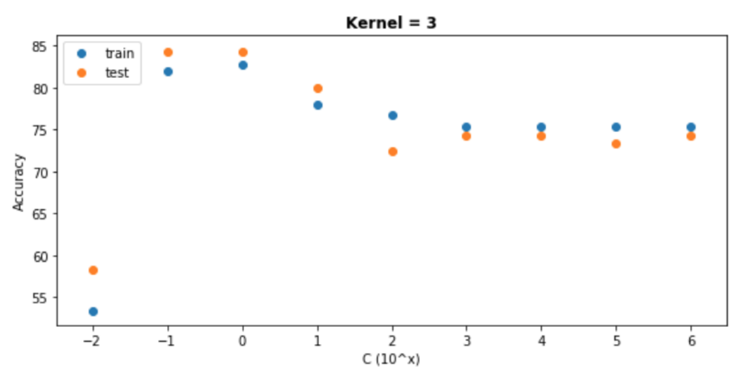
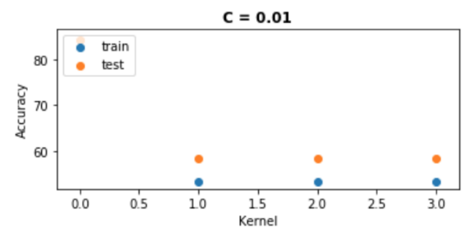
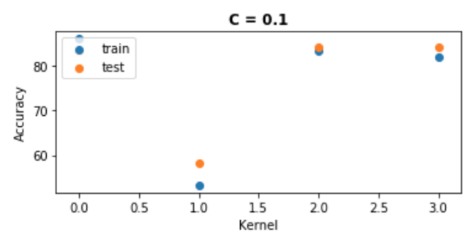


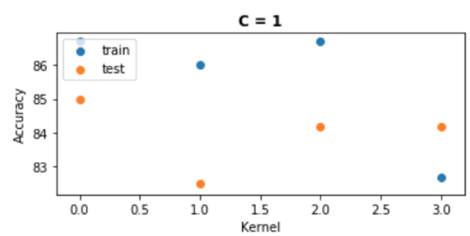
Figure 4: Kernel 3

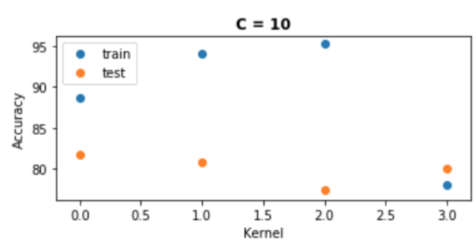
2.2 Different Kernels with Fixed Value of C

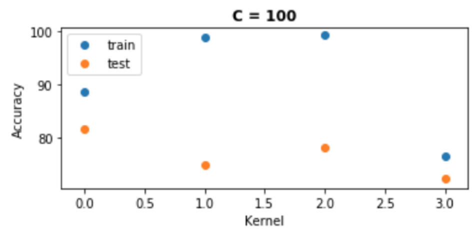
Below figures display the accuracy of SVMs with a fixed value of parameter C with 4 different kernel methods. It's really hard to derive a general conclusion for all grahps here but in general we say, test accuracy is higher for C values greater than 10. Also, there is a slight drop in accuracy values for kernel 1 compared to others.

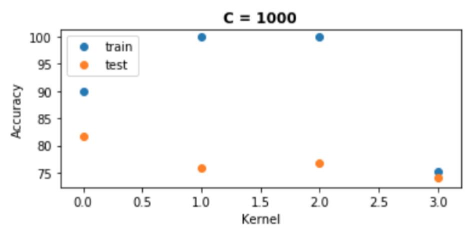


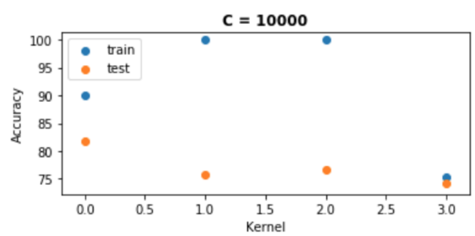


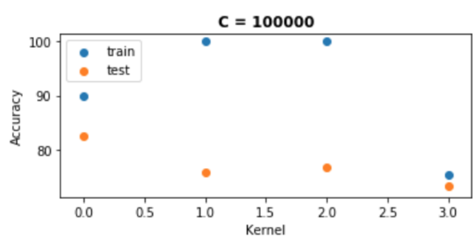


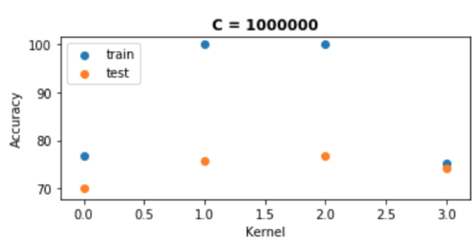



C = 100
train
test
Accuracy
Kernel


C = 1000
train
test
Accuracy
Kernel

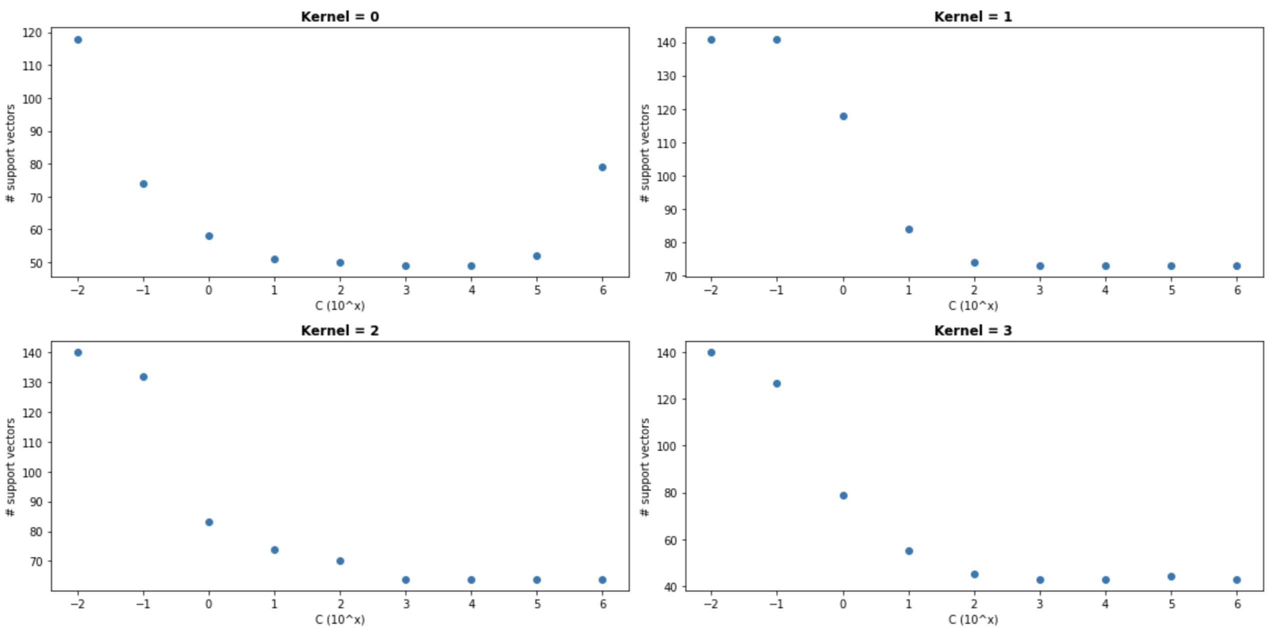

C = 10000
train
test
Accuracy
Kernel


C = 100000
train
test
Accuracy
Kernel


C = 1000000
train
test
Accuracy
Kernel

3 Number of Support Vectors with Increasing C

When we observed the number of support vectors for increasing C values, we concluded that it is decreased or kept the same value when C is increased in all kernels. The only outlier value was for kernel 0 and the largest C value in which it was expected to decrease or stay the same but it increased. Thus the results were compatible with the theory. As C incerases, the number of support vectors either decreases or stays unchanged



4 Removing Data Points from Training Set

As we discussed in the lectures, support vectors are data points supporting the margin of the seperating line, indeed they are the points giving the shape to the seperating line. This made us think that removing a support vector causes changes in the hyperplane. And we also said that, in the lectures, non-support vector data points has nothing to do with the margin as they are not close to the seperating line.

In order to examine the changes in the hyperplane, we defined a function taking 3 paramaters (train_label, train_data, coef) and returning the coefficients of w. Then, we called this function 4 times by removing a nun-support vector data point in each and observed the differences in the hyperplane. We noticed that removing a non-support vector data points has either no effect or a minor effect on the hyperplane (We may interpret this little difference as a noise).

In order to see the effect of removing a support vector from training data, we called the same function 4 times by removing a support vector in each and observed the differences in the hyperplane. This time we observed that removing a support vector from training data causes changes in the hyperplane, as we discussed in the lectures.

5 Bonus

In order to use CVXOPT library, we need to obtain give 4 parameters that we obtain from data and its labels. Those 4 variables are P, q, G, and h. We have used toy data from slides as $X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix}$ and $Y = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix}$. What we expect from QP solver to find coefficients as $b = -1, w_1 = 1$, and $w_2 = -1$ Then, we have obtained parameters as $b = -1.00000001, w_1 = 1.00000001$, and $w_2 = -1.00000001$ which is so close to the result we expect and it is due to the noise from CVXOPT QP solver.