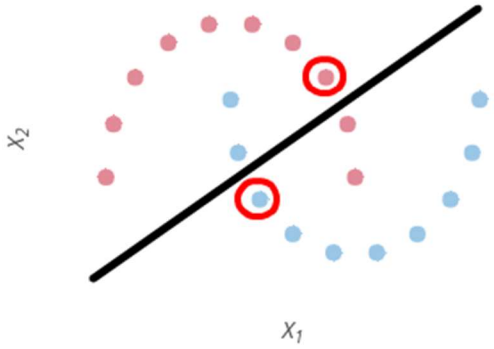
Homework 2
COSC 6342: Machine Learning
University of Houston
Department of Computer Science
Sent on: Sep. 21, 2018; Due: Oct. 8, 2018

Names: Akshit Tandon & Ansh Dedha

Support Vector Machines

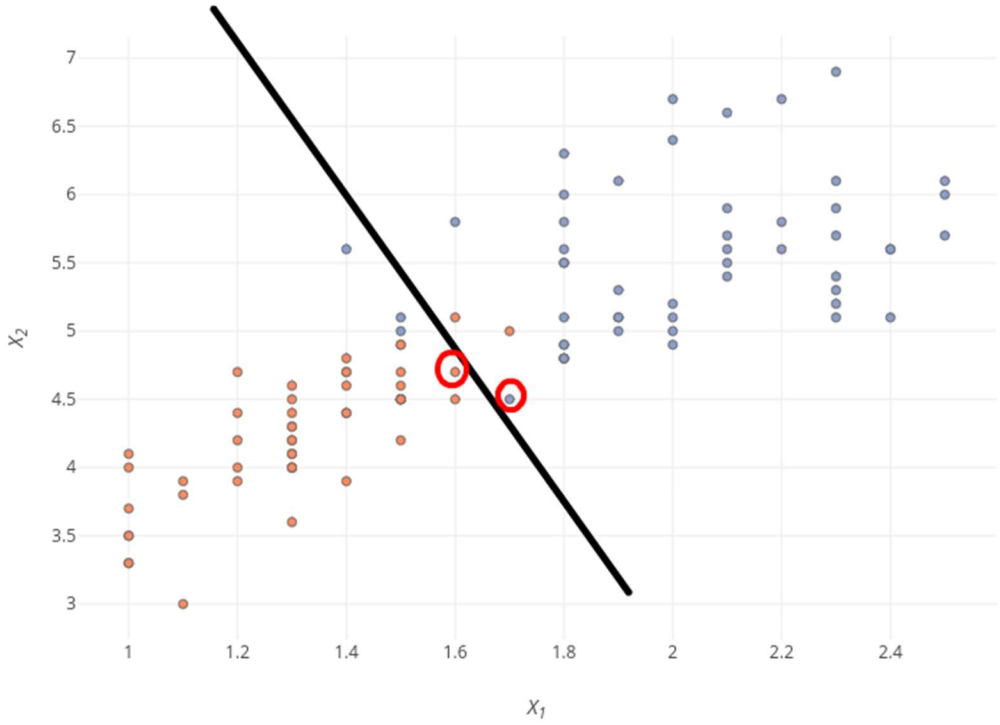
Question 1 (10 pts): In an informal way, circle the support vectors a linear SVM may choose for the data in Figure 1 and 2.

Figure 1: Determine Linear SVM hyperplane



c

Figure 2: Support Vectors



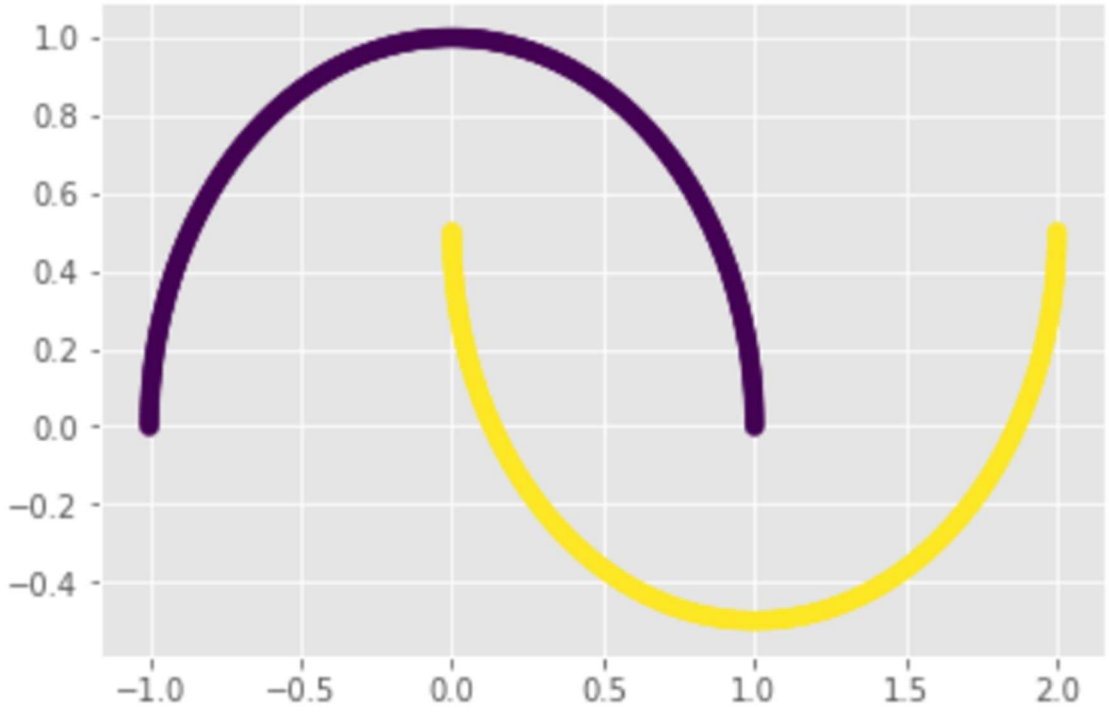
Question 2 (10 pts): Briefly explain the purpose of the kernel function.
The kernel functions are used to transform the data into the required form. It is the generalized function of the different dimensions of data. It enables operation in higher dimensions.
In other words, the kernel functions modify the training data that is not linearly separable to a form that can be separated linearly in some other higher dimension.

Question 3 (10 pts): Suppose your dataset has a large number of features. What effect, if any, would feature selection have on an SVM?
Although SVMs with kernels do not deal with the feature space directly, selecting the right features helps in finding the correct classification. The features should be selected in such a way that any prominent information or feature is not discarded. It is best if our dataset could be described using less number of features but if not, appropriate number of features be selected such that it can describe our dataset in appropriate manner.

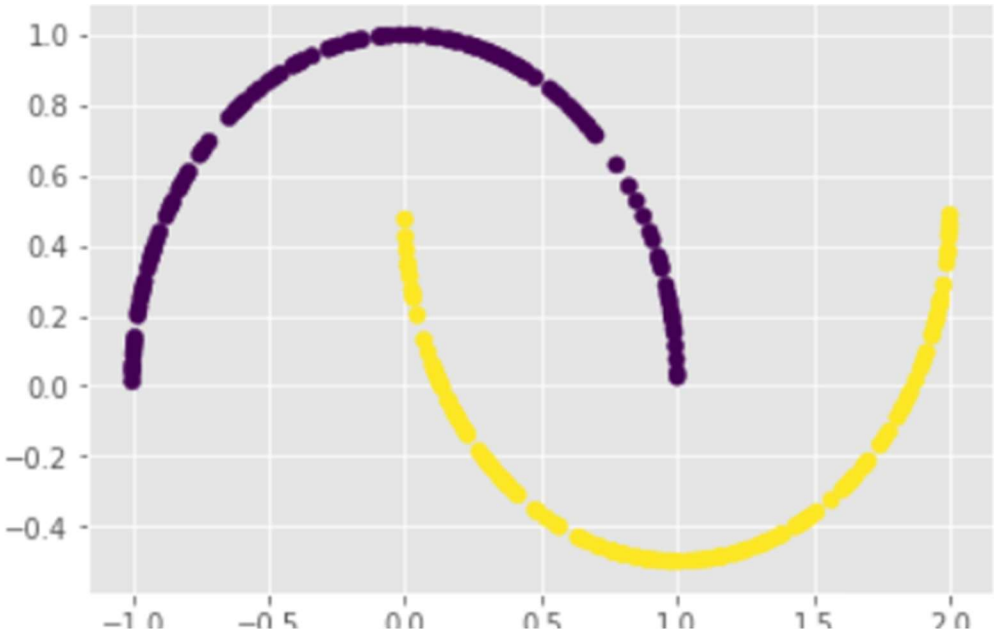
Question 4 (10 pts): What is the effect of raising or lowering λ hyper-parameter in an SVM
λ is the class-weight. Increasing the value of λ for a class increases the margin between that class and the hyperplane. This results in more number of examples being classified as that class. For higher λ for class A, there can be more number of examples that might be miss-classified as class A. Conversely, if the value of λ is too low for a class A, some examples that actually belong to that class may be classified as another class.

Question 5 (70 pts): Train an SVM with linear and rbf kernels on the moons dataset, as well as one other of your choice, taken or generated from sklearn datasets. Optimize the hyper-parameters and settings to try to achieve the best accuracy on the validation set and report your results. Then run your model on the test dataset. Report the accuracy. Graph and attach the classifications on the test set.

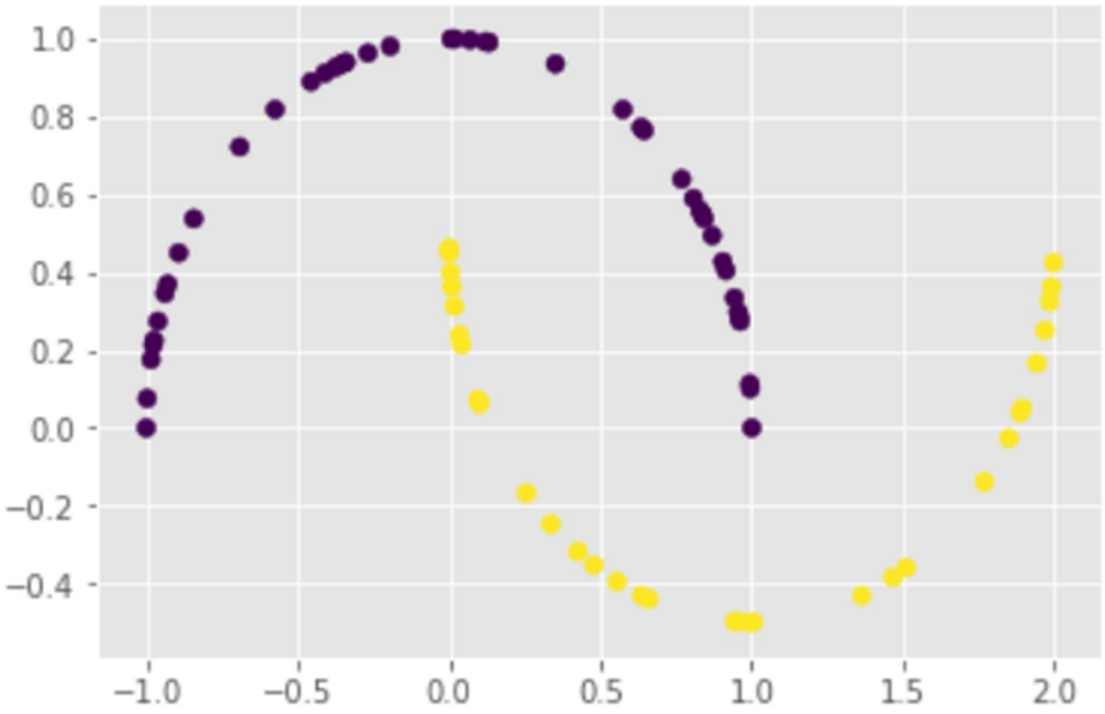
Please refer to 'Code-1' file for source code (using Moon dataset). In Linear kernel we only need 'c' hyper parameter. In the validation set it was found that the most optimized value for c=1.9 where the accuracy was 90.66%. The accuracy for the same in Testing dataset was 88%.
Now for 'rbf' kernel we require two parameters 'c' and 'gamma'. In the validation set it was found that the most optimized value for c=0.5 and gamma=0.6, where the accuracy was 100%. The accuracy for the same in Testing dataset was also 100%.



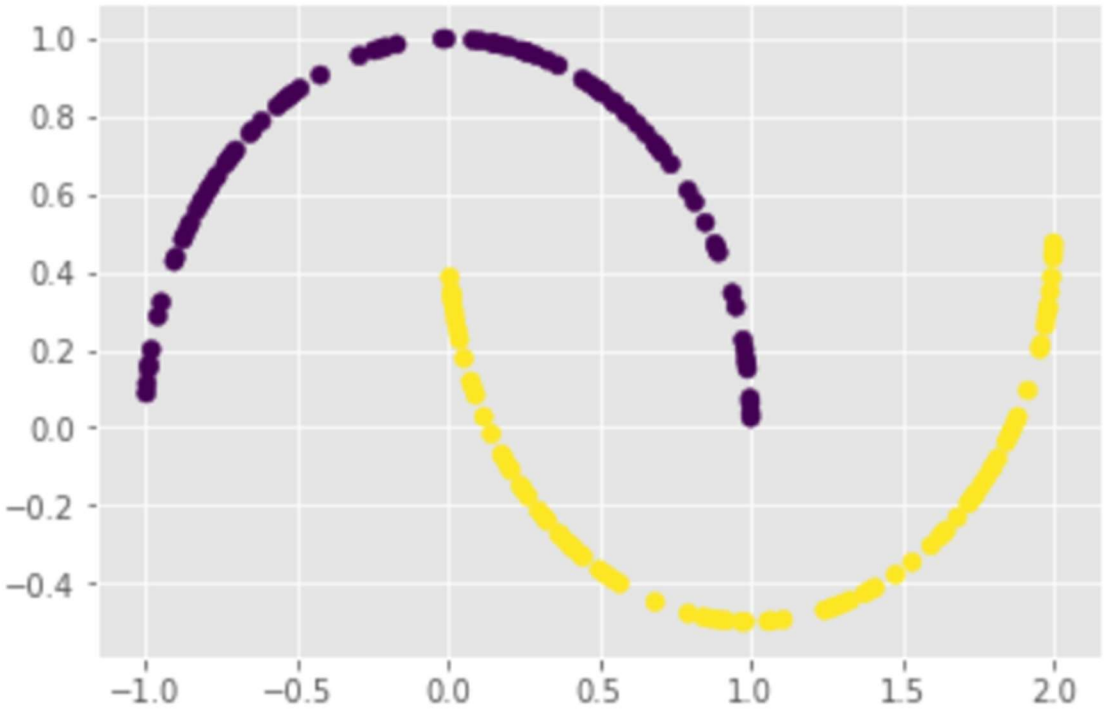
Original Dataset



Training Dataset



Validation Dataset



Testing Dataset

Please refer to 'Code-2' file for source code (using Iris dataset). In Linear kernel we only need 'c' hyper parameter. In the validation set it was found that the most optimized value for c=1.2 where the accuracy was 95.65%. The accuracy for the same in Testing dataset was 96.15%.

Now for 'rbf' kernel we require two parameters 'c' and 'gamma'. In the validation set it was found that the most optimized value for c=1.7 and gamma=1.6, where the accuracy was 91.30%. The accuracy for the same in Testing dataset was also 92.30%.

Testing Dataset