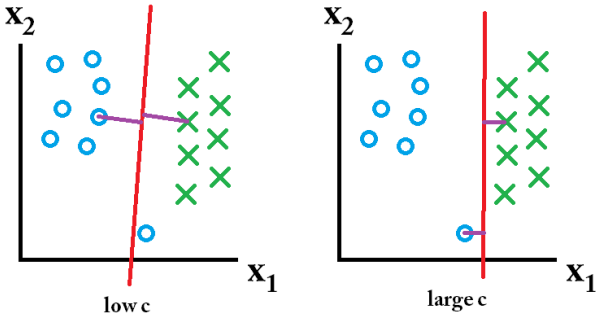
Analysis
Homework 5: SVM by Rohan Chandra

Part 1:

The following table list 5 values of 'C' for 2 kernels using all 50,000 data points. The accuracy decreases with increasing C in the case of linear kernel, and the accuracy increases with increasing C for thr RBF kernel.

Reason:
The C parameter trades off between a hyperplane with the largest minimum margin, and a hyperplane that correctly separates as many instances as possible. For large values of C, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. Conversely, a very small value of C will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.

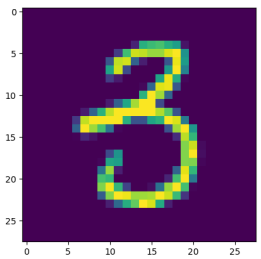
In a linear kernel, a larger margin width is desirable as it reduces overfitting. Hence a low C gives a higher accuracy on the test set.



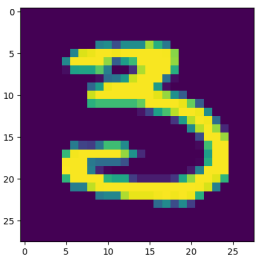
Kernel	C	Accuracy
Linear	0.2	0.97
Linear	0.4	0.97
Linear	0.6	0.96
Linear	0.8	0.96
Linear	1.0	0.96
RBF	0.2	0.964
RBF	0.4	0.967
RBF	0.6	0.967
RBF	0.8	0.968
RBF	1.0	0.969

Part 2:

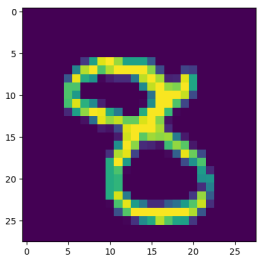
Support Vectors are those training data points that lie on the support margin and are closest to the separating hyperplane. Since they are closest to the separating hyperplane, practically speaking, these are the examples that are most likely to be confused with the oppposite class. That being said, consider the following support vectors generated using C= 0.2 and using all data points with a linear kernel.



(a) The upper and lower tails curl inwards to close the loop, horizontal bar in the middle more prominant



(b) The upper and lower tails curl inwards to close the loop, upper loop converges



(c) Paet of lower loop is erased or missing on the left side. Resembles the left open loop of a '3'