

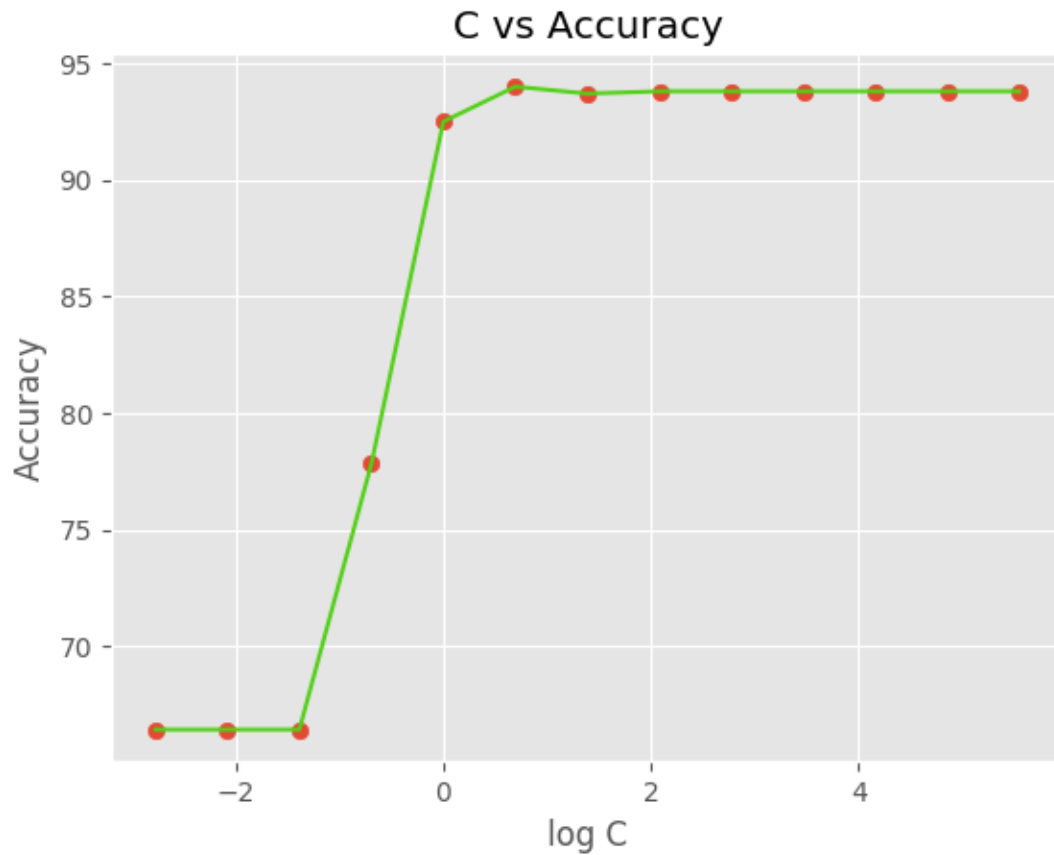
CSE 5526

Programming Assignment 3

Pragya Arora

arora.170

Linear SVM



- The maximum accuracy achieved is 94.006 at $C = 2$
- From the above graph we can see that
 - After attaining the best accuracy there is a slight drop in accuracy, after which the accuracy becomes constant
 - There is a sharp increase in accuracy at $C=0.25$ to 1
- Higher value of C indicates that focus is more on reducing the misclassification and not on increasing the margin of separation. Thus, we can see that the test error reduces when training misclassification is low.

RBF Kernel SVM

1. Cross Validation Results

C/ α	$\alpha = 2^{-4}$	$\alpha = 2^{-3}$	$\alpha = 2^{-2}$	$\alpha = 2^{-1}$	$\alpha = 2^0$	$\alpha = 2^1$	$\alpha = 2^2$	$\alpha = 2^3$	$\alpha = 2^4$	$\alpha = 2^5$	$\alpha = 2^6$	$\alpha = 2^7$	$\alpha = 2^8$
C = 2^{-4}	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4
C = 2^{-3}	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.4	68.5	68.7	68.5	68.4	68.4
C = 2^{-2}	68.4	68.4	68.4	68.4	68.4	68.4	68.4	71.5	74.4	74	72.9	69.9	68.4
C = 2^{-1}	68.4	68.4	68.4	68.4	68.4	70.4	78.2	80	80.6	78.3	75.8	73.2	69.6
C = 2^0	68.4	68.4	68.4	68.4	79.2	89.6	90.4	88.5	87	85.2	81.2	77	72.9
C = 2^1	68.4	68.3	69.1	86.2	92.8	93.5	92.8	91.9	90.4	87.7	84	79.5	74.1
C = 2^2	68.3	69.3	88.1	93.4	93.6	94	94	92.5	90.8	88	83.6	79.2	74.1
C = 2^3	69.7	89.2	93.9	94.2	94.5	94.3	93.9	92.3	90.4	87.3	83.7	78.9	74.1
C = 2^4	89.1	94.4	94.2	95	94.5	94.8	93.4	92.1	90.4	87.6	83.3	79.1	74.1
C = 2^5	95	94.7	94.8	95.1	94.7	94.2	93.5	92	90	87.4	83	79.1	74.1
C = 2^6	94.7	94.6	94.7	94.5	95	94.3	93	91.9	89.8	87.1	83	79.1	74.1
C = 2^7	94.9	94.4	94.5	94.6	95	94	92.6	91.5	89	86.4	83	79.1	74.1
C = 2^8	94.8	94.8	94.4	95.2	94.8	93	91.9	90.5	88.7	86.4	83	79.1	74.1

- After doing the cross-validation on half of the training set, the maximum accuracy of 95.2 is achieved at $C=2^8(256)$ and $\alpha = 2^{-1}(0.5)$
- From the table we can see that large value of C and small value of α gives better accuracy
- Along the diagonal of increasing C and decreasing α , high accuracies can be observed
- High values of α means a low standard deviation of gaussians centered at the support vectors. Thus, the influence of these gaussians is limited to only the support vectors and thus the model overfits the data. Even on increasing C (regularization parameter), the model doesn't stop overfitting
- The model behaves similar to a linear model when α is small. This is because when α is low, gaussian width of the kernels is high. Thus, all gaussians are influenced by all data points in the training set

2. Classification accuracy on test set

- Model is trained using $C=2^8(256)$ and $\alpha = 2^{-1}(0.5)$ on the entire training set
- The accuracy achieved by this model on the test dataset is 95.10