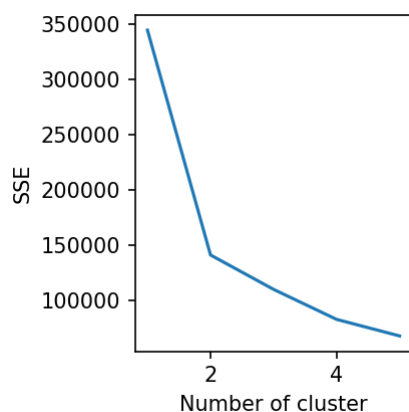


Support Vector Machines

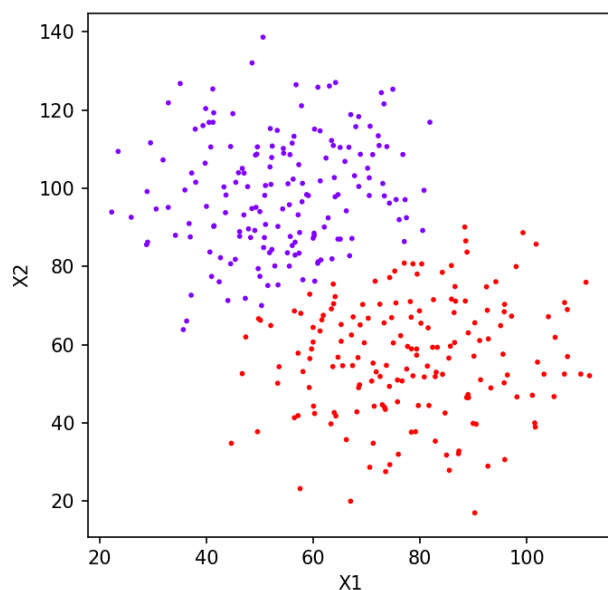
1. Used k-means clustering to determine the number of clusters. In order to determine the appropriate number of clusters to be supplied to the k-means clustering technique, used the elbow method (the value of k that significantly reduces the value of SSE).

Below are the required graphs:



Since the SSE is significantly reduced from $k=1$ to $k=2$, taking $k=2$ (number of clusters)

Plotting the scatter plot for the data points.



2. Scaled values present as a part of code
3. Applied SVM -> present as part of code
4. Best value set (C , γ and accuracy) [also present as part of code] -> [2^{-5} , 2^3 , 94.12 %]

This set is obtained after doing a loose grid search and then doing the fine grid search in the neighborhood of the best accuracy point obtained by the loose grid search.

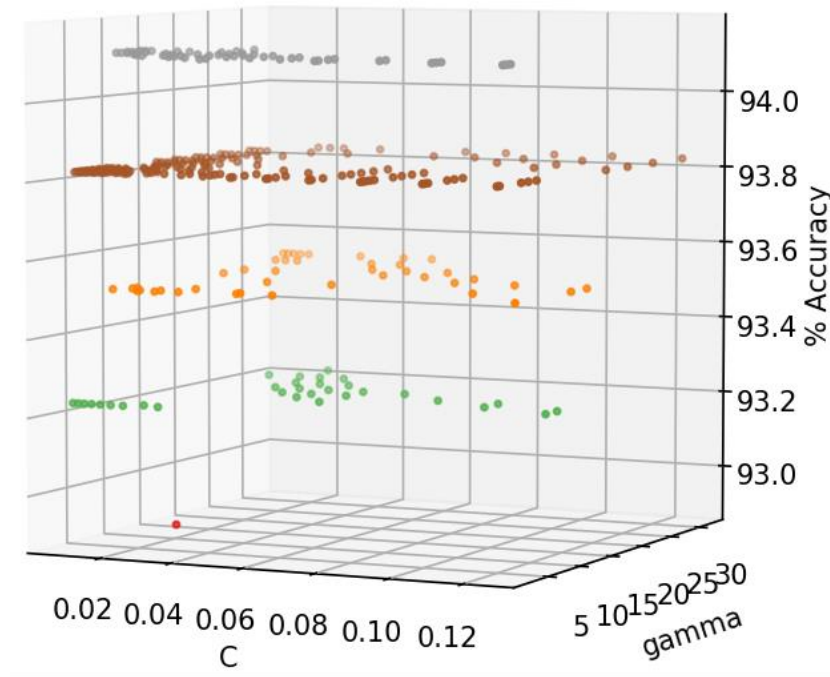
Loose grid Search:

These values are obtained by varying C from 2^{-5} to 2^{15} and γ from 2^{-15} to 2^3 and finding the accuracies for each value of C and γ using the k -cross validation method where $k = 5$. In this, the entire dataset is divided into 5 groups and of these 4 groups are taken as training sets and 1 group is taken as the validation set. Then the accuracy is computed. This is then repeated in a circular way where each group behaves as the validation set and the other 4 groups behave as the training set and accuracies are obtained. Then the average of all the 5 accuracies is taken to obtain the mean accuracy. Of all the values of C and γ , the one which have the highest accuracy will be used in the model that will finally be used to find the target attribute of the actual test data.

Fine Grid Search:

The best value point obtained from the Loose Grid search is $[C, \gamma \text{ and accuracy}] \rightarrow [2^{-5}, 2^3, 94.12\%]$. Doing a loose grid search by varying C from 2^{-3} to 2^{-7} in multiples of $2^{-0.25}$ and γ from 2^1 to 2^5 in multiples of $2^{0.25}$. After finding the highest accuracy point (in this case multiple points), the highest accuracy still remains the same as the value obtained by the Loose Grid Search of 94.12% and is shared by many neighborhood points. One of those points is the original point $[2^{-5}, 2^3, 94.12\%]$.

5. Present as part of the code
6. Present as part of the code
7. Present as part of the code
8. 3D Scatter Plot for Fine Grid Search showing regions of same accuracy (there are 5 different regions of same accuracy):



The points that are away from the user's view are shown in light shades of the same color and should not be confused for a different color.

Final Result Comments:

From the scatter plot shown at the top, it is clear that there are 2 separate clusters and a line of maximum separation can be drawn to separate the 2 clusters (thus approximately all points are correctly classified with their corresponding labels) giving a high accuracy of 94.12%