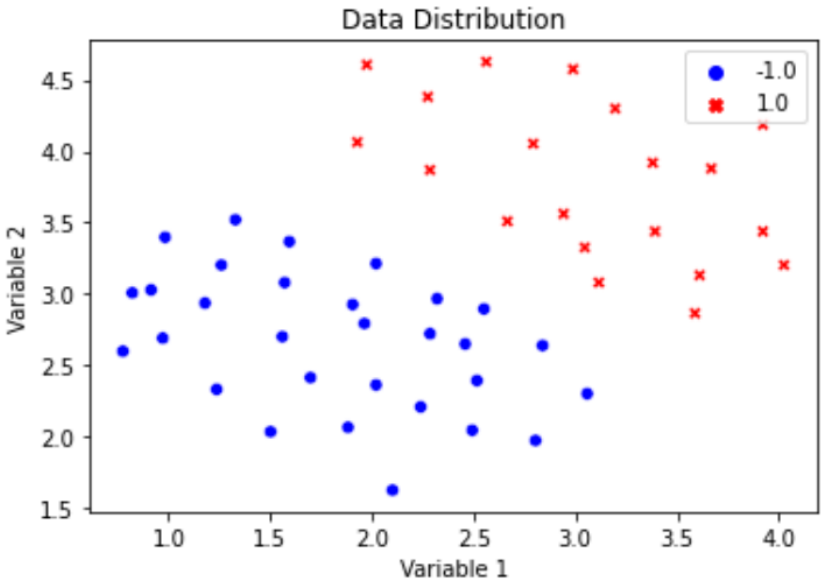
CLL788 Assignment 3

Rishav Kumar Rajak 2018CH70302

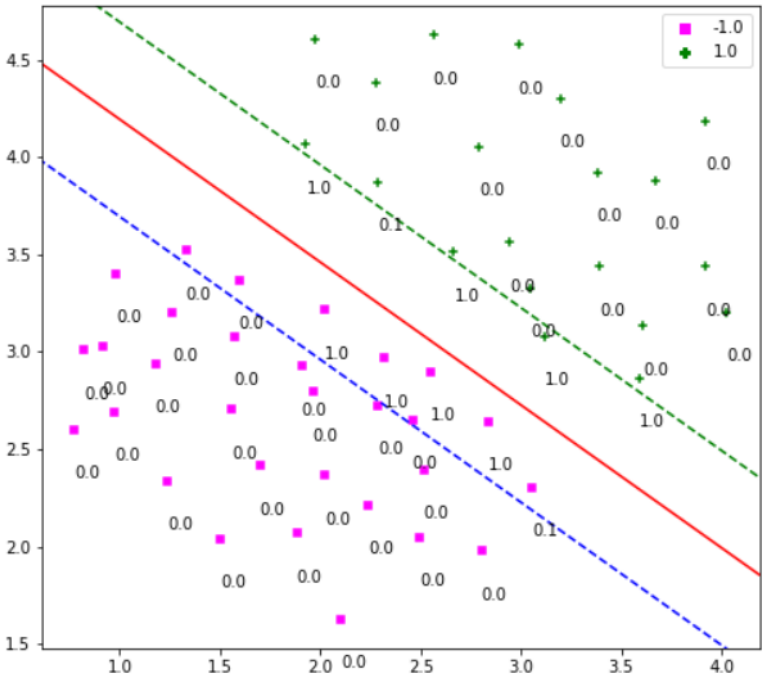
1)



From the above plot, we can see that the data points are grouped into two distinct spaces in the 2D coordinate system. One group is close to the origin and has low values for both the features (Variable 1 and Variable 2).

2) a) SVM Decision Boundary with margin

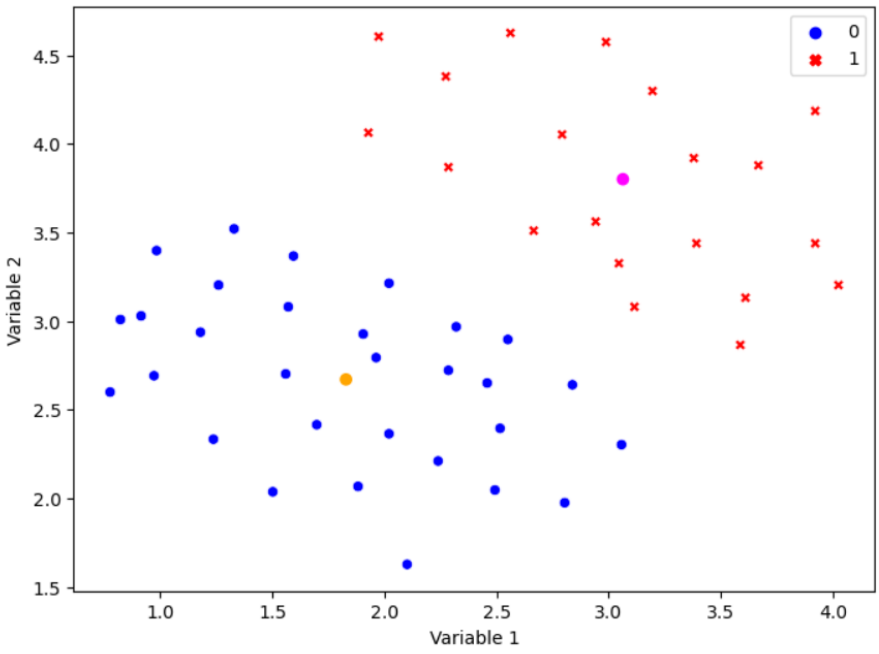
Weights and Bias: ([1.46838317, 2.00128397]), -9.860366657974234)



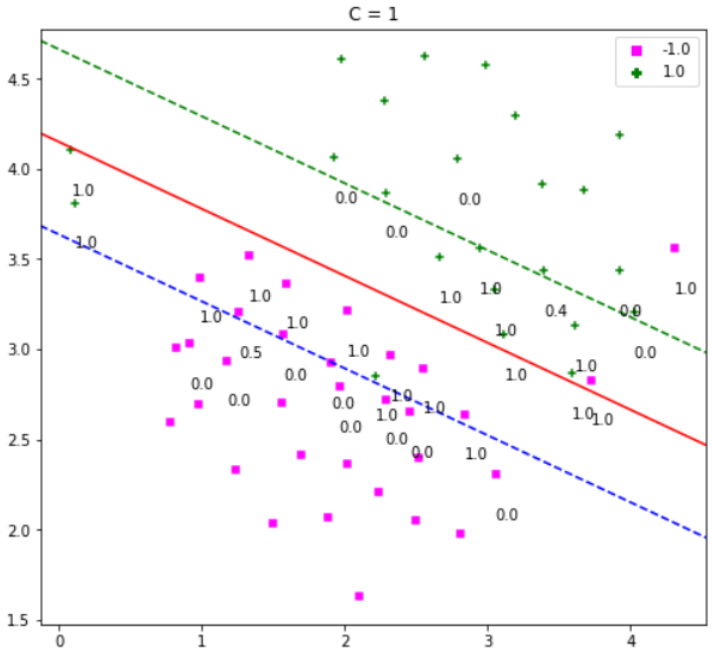
2) b) KMeans

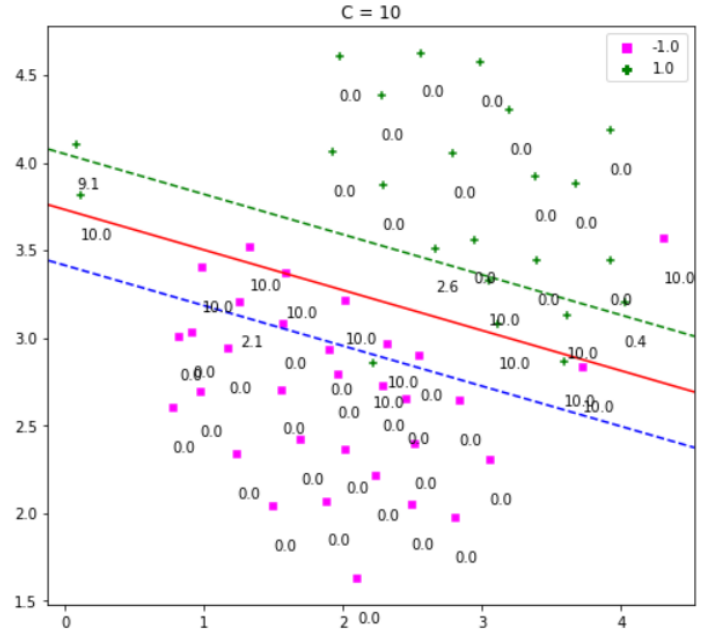
Here I applied the K-Means algorithm to find out the two clusters. These clusters contain datapoints, whose characteristics are similar and by that, it can be understood that, between two clusters the degree of similarity is really less and that is why there are two clusters to group them in the first place. The K Means algorithm groups the data points on the basis minimum euclidean distance between them. The quetion says that we need to make two clusters, which means that the value of K = 2. So, the input parameters to this algorithm is the value of K and the dataset (which needs to be clustered). Now for every cluster, we have a cluster center, popularly known as a centroid.

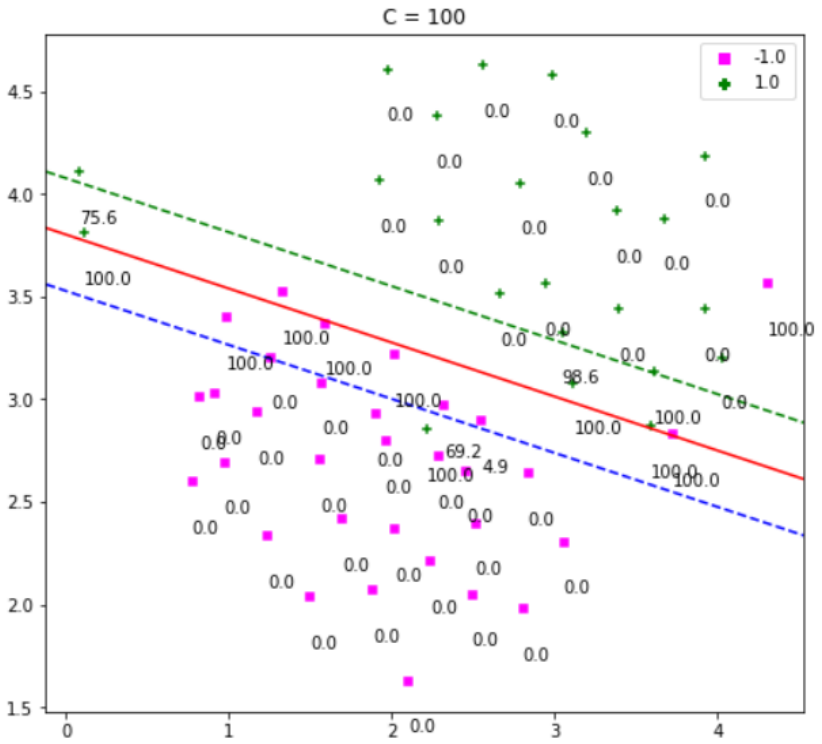
Cluster Centres: array ([1.828335, 2.67121], [3.063565, 3.800865])



3) SVM with modified optimization problem. As you can see from the graph that there are some misclassifications within the margin. To account for this, we introduce a slack variable. C is the constant that controls the trade-off. With increase in C, the gap between the margin decreases.







Changing Kernel

Below is the plot with different kernel (Linear, RBF, Polynomial with degree 3)