CENG499 - Introduction to Machine Learning
HW02 Report
Anıl Utku Ilgın
2260073

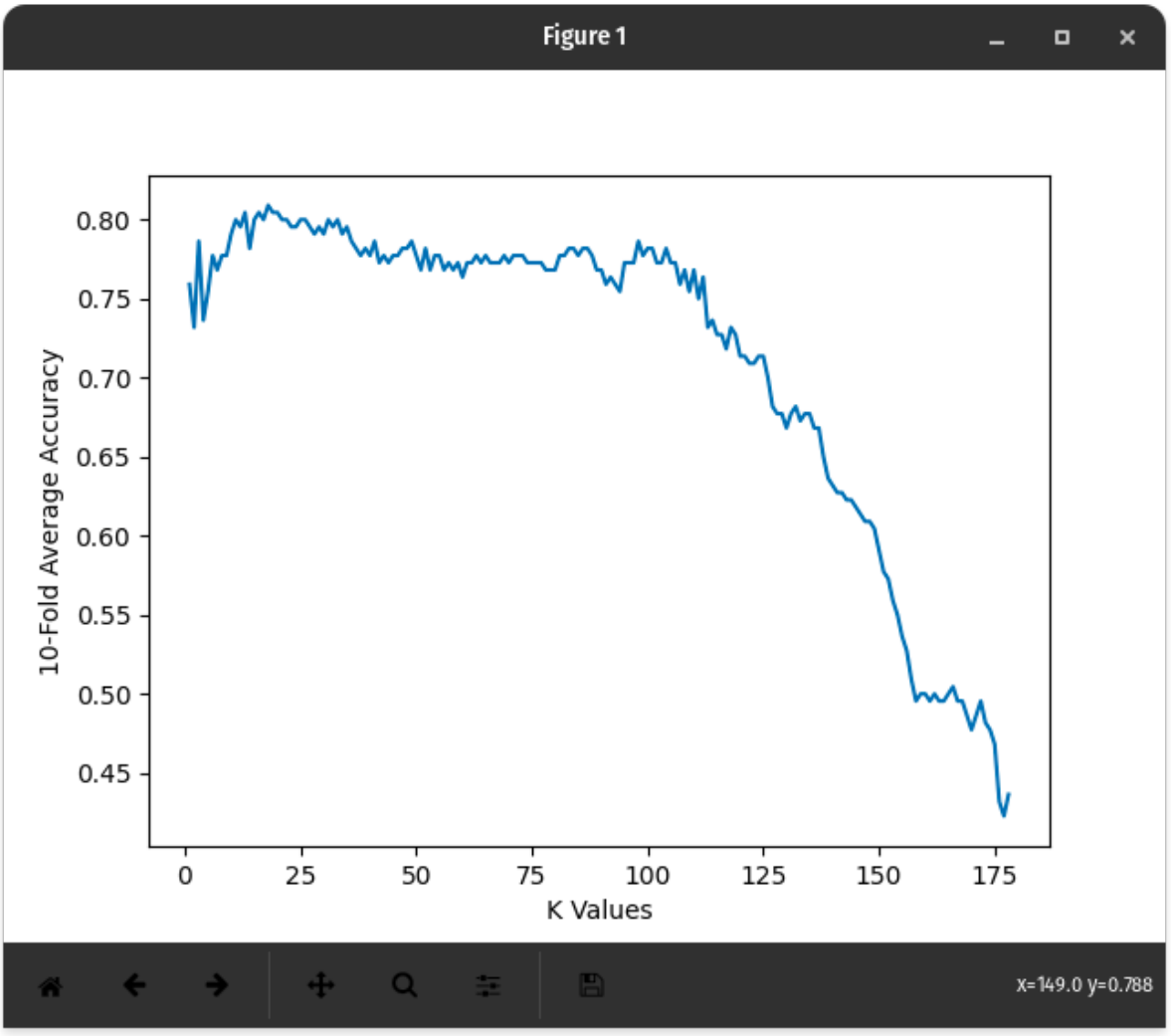


Figure 1 : KNN plot for 10-fold Cross Validation for L1

Comment:
As the K values increase, first accuracy increases to a point where K is the optimum choice for the dataset. After that point, accuracy starts to decrease since there are multiple data points that start to influence the labels of our data, even when they shouldn't be related because their distance is too much compared to the more nearest neighbours. Hence, the accuracy drops because we are labeling wrong depending on more and more diversed, unrelated data.
Test Accuracy for Best Result : For K=18, Accuracy is 0.82

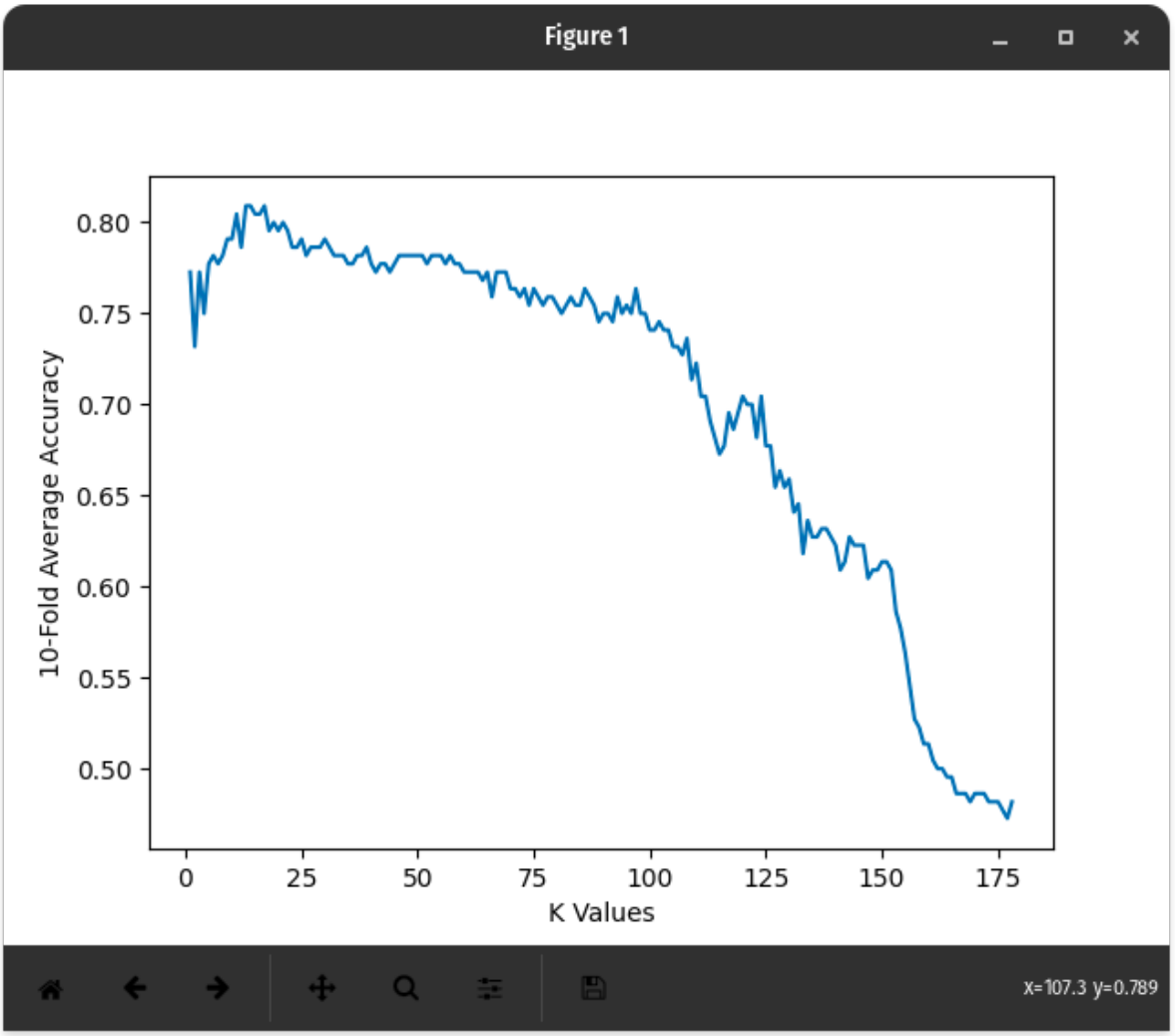
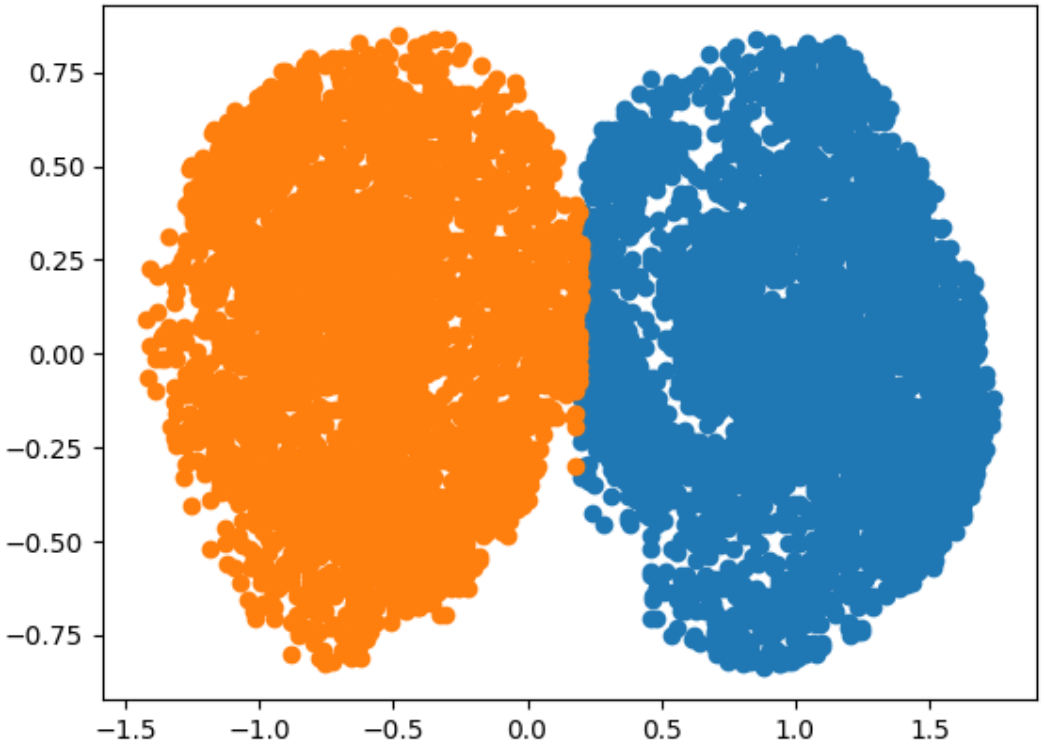


Figure 2: KNN plot for 10-fold Cross Validation for L2

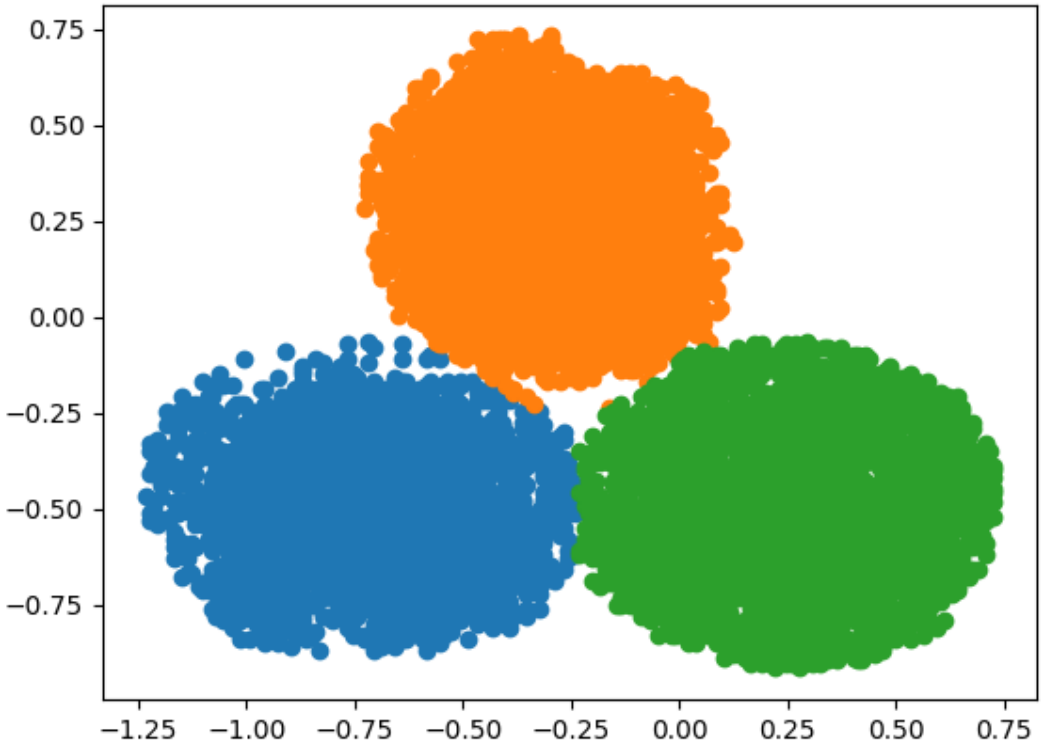
Comment:

As the K values increase, first accuracy increases to a point where K is the optimum choice for the dataset. After that point, accuracy starts to decrease since there are multiple data points that start to influence the labels of our data, even when they shouldn't be related because their distance is too much compared to the more nearest neighbours. Hence, the accuracy drops because we are labeling wrong depending on more and more diversed, unrelated data.

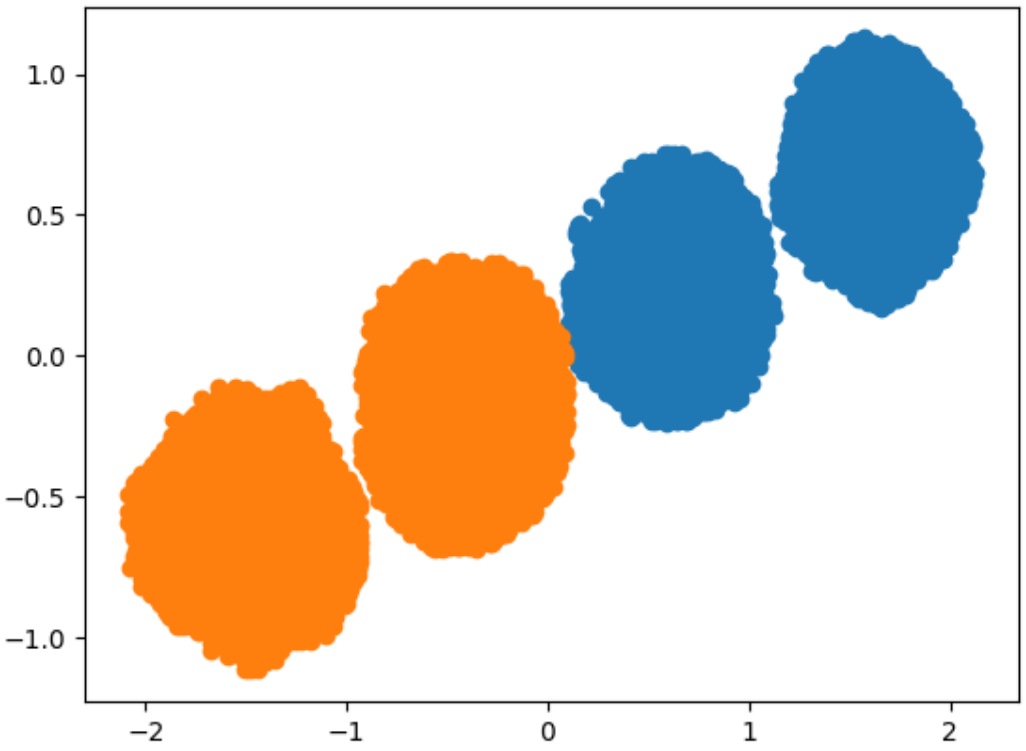
Test Accuracy for Best Result : For K=13, Accuracy is 0.81



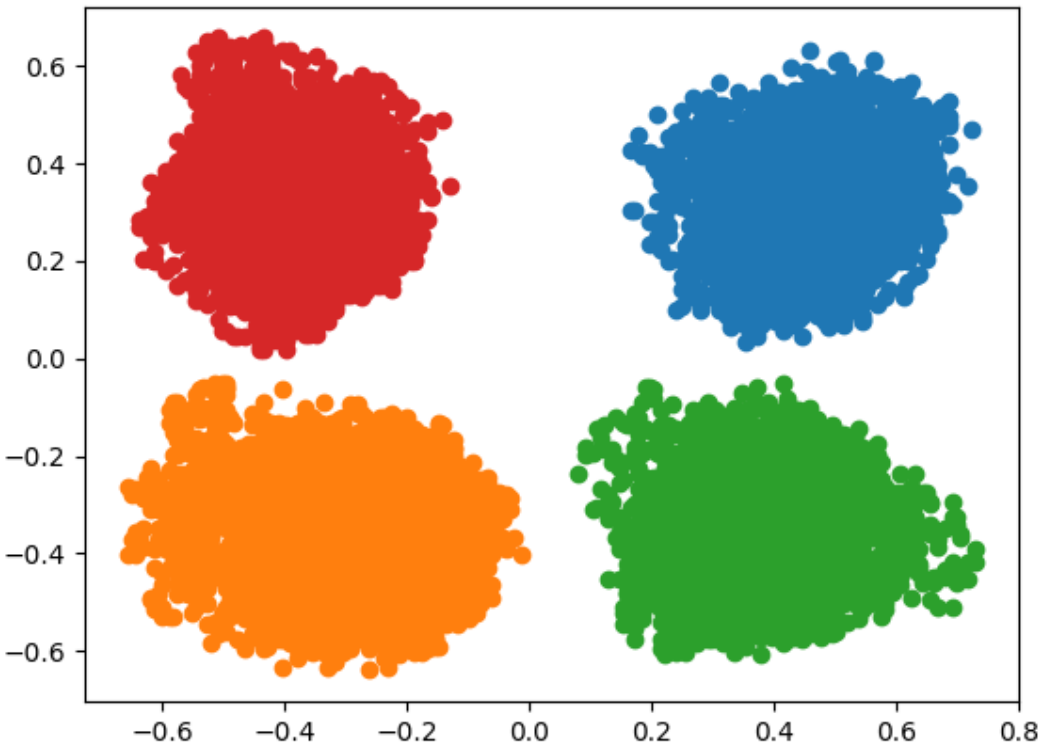
K-Means for Dataset 1 - K=2



K-Means for Dataset 2 K=3



K-Means for Dataset 3 K = 2



K-Means for Dataset 4 K = 4

I have run the loops for 10 times, yet they all resulted nearly similar since I've used Kmeans++ for selecting the centers.

-I couldn't comment and plot the rest of the codes I'm sorry for that.