Computer Assignment: KNN

Submitted by- Vandit Sharma **Roll No.-** 17EC10060

Aim-

- To implement a function to return K-Nearest Neighbours
- To predict the class of the Test data-set for each k value and each distance metric.
- To compute the accuracy and plot a bar chart to compare the performance of hyperparameters.

Case-I: Euclidean Distance as distance metric

Results-

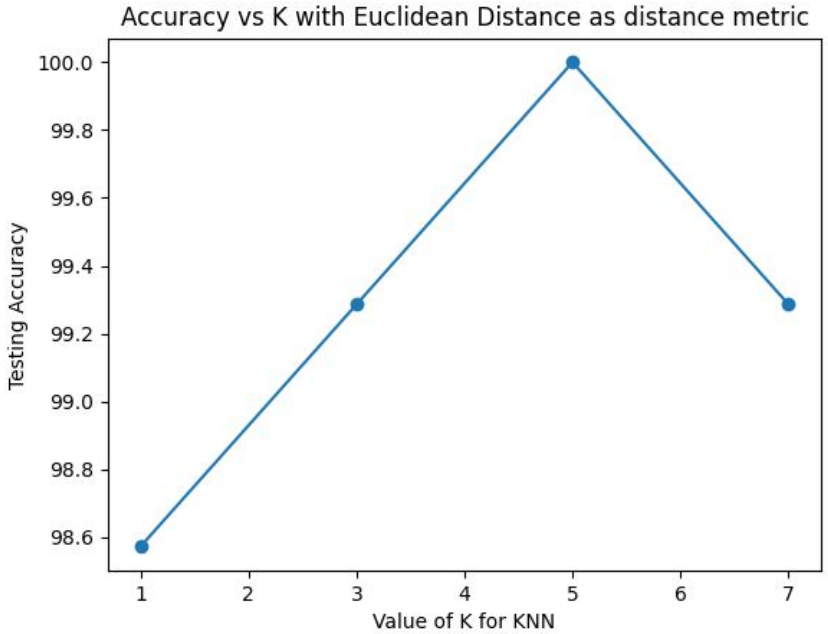

Figure 1: Accuracy vs K with Euclidean Distance as distance metric

(k = 1, 3, 5, 7)

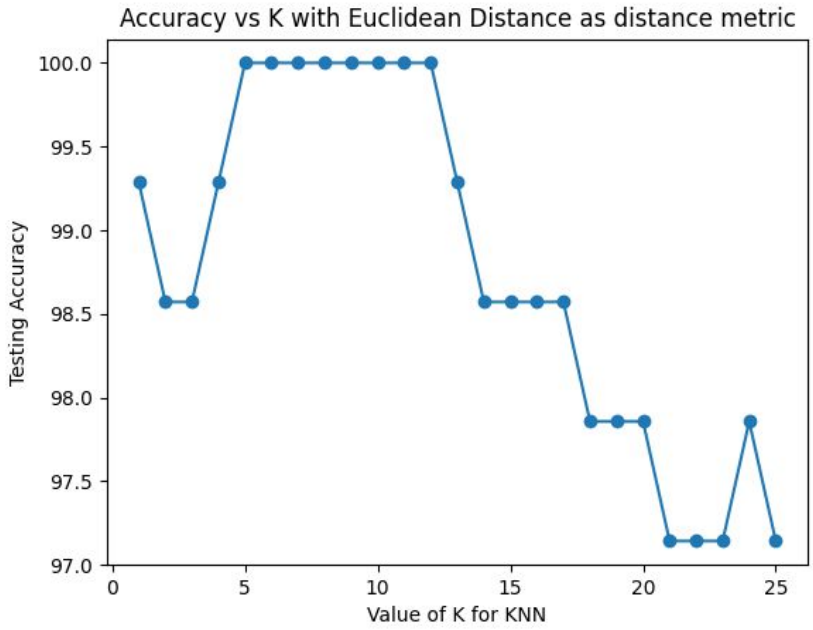

Figure 2: Accuracy vs K with Euclidean Distance as distance metric

(k ∈ [1, 25])

Observations-

- In Figure 1, a maximum accuracy of 100% was achieved at K = 5.
- There was a variation in accuracy with different values of K.
- Upon running the program several times, it was found that the best accuracies were achieved when K belonged to the range [4, 13] approximately.
- For larger values of K (say 20), the accuracy kept decreasing, though wasn't seriously impacted.
- An excellent accuracy of over 98% was achieved for the given four values of K (1, 3, 5, 7)

Case-II: Normalized Euclidean Distance as distance metric

Results-

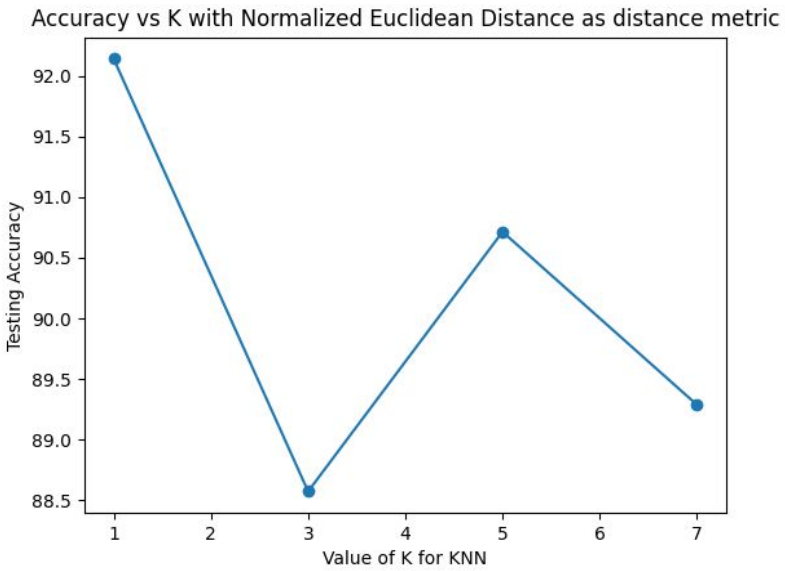

Figure 3: Accuracy vs K with Normalized Euclidean Distance as distance metric

(k = 1, 3, 5, 7)

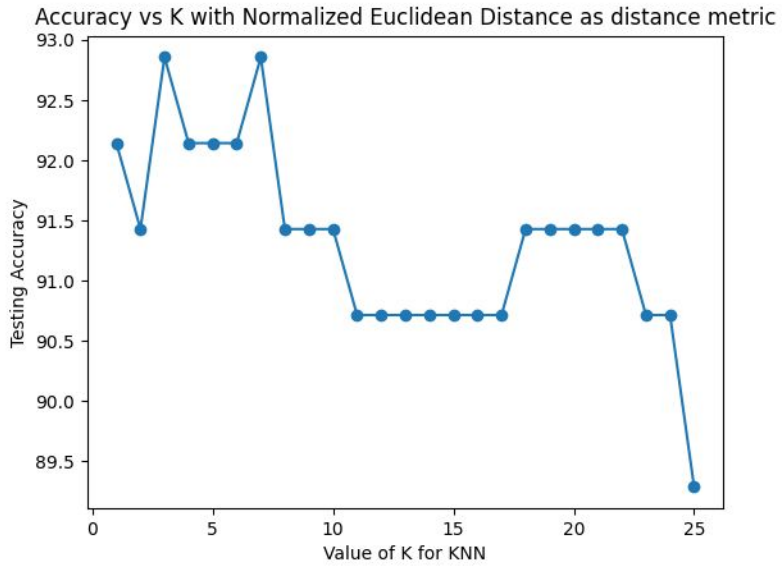

Figure 4: Accuracy vs K with Normalized Euclidean Distance as distance metric

(k ∈ [1, 25])

Observations-

- In Figure 3, a maximum accuracy of 92.14% was achieved at K = 1.
- There was a variation in accuracy with different values of K.
- For larger values of K (say beyond 20), the accuracy kept decreasing and a significant drop in accuracy was observed compared to lesser values of K.
- A decent accuracy around 90% was achieved for all four values of K.
- The accuracy was, in general, lower than that achieved using Euclidean distance as distance metric

Case-III: Cosine Similarity as distance metric

Results-

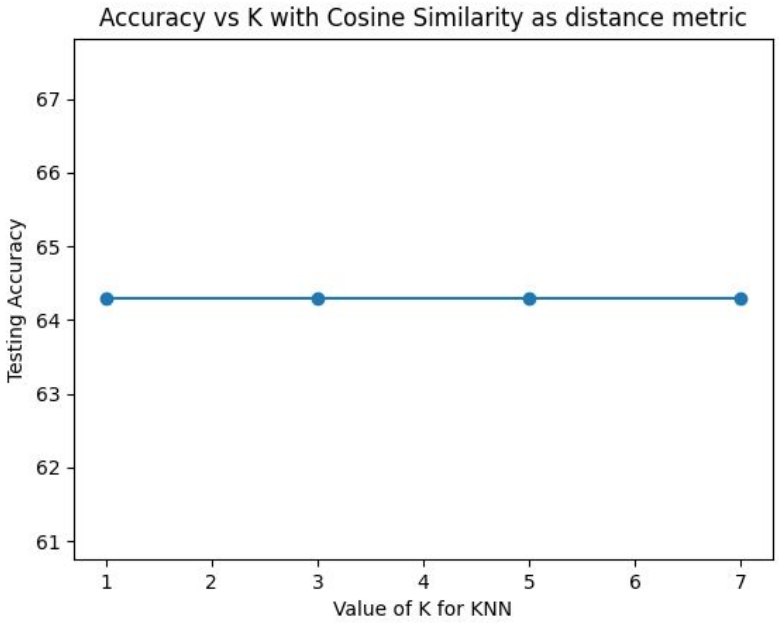

Figure 5: Accuracy vs K with Cosine Similarity as distance metric

(k = 1, 3, 5, 7)

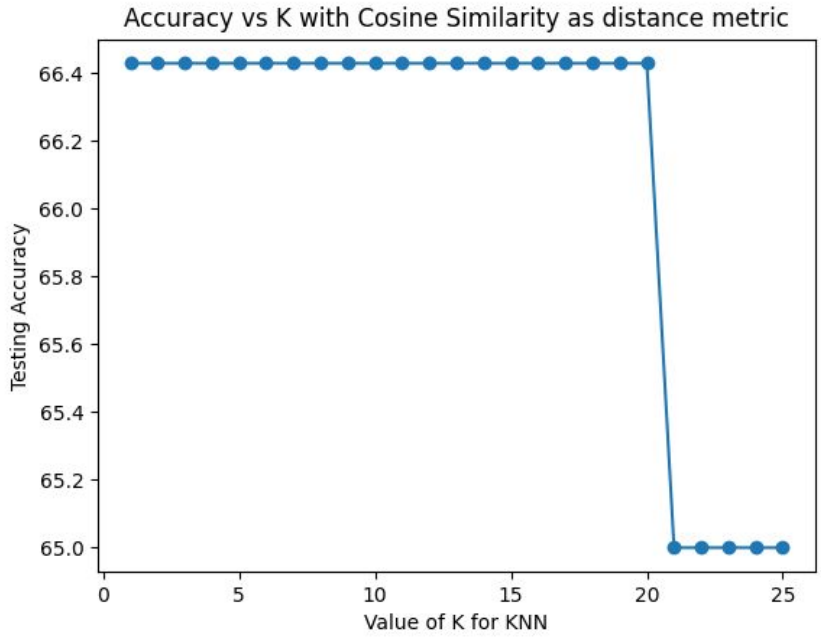

Figure 6: Accuracy vs K with Cosine Similarity as distance metric

($k \in [1, 25]$)

Observations-

- In Figure 5, a maximum accuracy of 64.28% was achieved at K = 1, 3, 5, 7.
- There was a variation in accuracy with different values of K, however, it did not change much for lesser values of K (less than 15)
- Upon running the program several times, it was found that lesser values of K consistently had better accuracies than larger values, and usually this was reelected by a sudden drop in the plot.
- The accuracies achieved in this case were, in general, significantly poorer than the other two cases.