ELC ACTIVITY – HAND RECOGNITION USING MNIST DATASET WITH KNN ALGORITHM

K-Nearest Neighbor (KNN) is a machine learning algorithm used for classification and regression problems. It is a non-parametric algorithm, meaning it does not make any assumptions about the underlying data distribution.

In the KNN algorithm, the idea is to predict the class or value of a new data point based on its proximity to other similar data points in the training set. The prediction is made by finding the K nearest data points to the new point, and using their classes or values to determine the class or value of the new point.

The working of the KNN algorithm we followed is as follows:

Data collection: A labeled dataset provided with the Activity.

Model training: The KNN algorithm does not have a training phase in the traditional sense, as it does not learn any parameters or coefficients. Instead, it simply stores the training set.

As we per assignment we trained it for 5 values of k

At k = 3, accuracy is around 76.4%

At k = 5, accuracy is around 80.1%

At k = 7, accuracy is around 80.6%

At k = 9, accuracy is around 81.02%

At k = 11, accuracy is around 81.1%

Prediction: When a new data point is encountered, the algorithm calculates the distances between the new point and all the points in the training set. The K nearest data points are then found, and the class or value of the new point is determined based on the majority class or mean value of these K nearest neighbors.

One of the main advantages of the KNN algorithm is its simplicity and ease of implementation. It can be used for both binary and multi-class problems, and does not require a lot of computational resources. Additionally, KNN can handle missing data and non-linear relationships between features.

However, KNN also has some limitations. It can be computationally expensive for large datasets, as the distance calculations need to be done for each new data point. The algorithm is also sensitive to the choice of K, as a small change in K can result in a large change in the predictions. Furthermore, KNN can be affected by the curse of dimensionality, meaning that the performance of the algorithm may degrade in high-dimensional spaces.

In conclusion, the K-Nearest Neighbor algorithm is a simple and effective tool for classification and regression problems. Its ease of implementation and ability to handle non-linear relationships make it a popular choice for many applications. However, its limitations need to be taken into account and alternative algorithms may need to be considered, d