Task 10: KNN – Handwritten Digit Classification

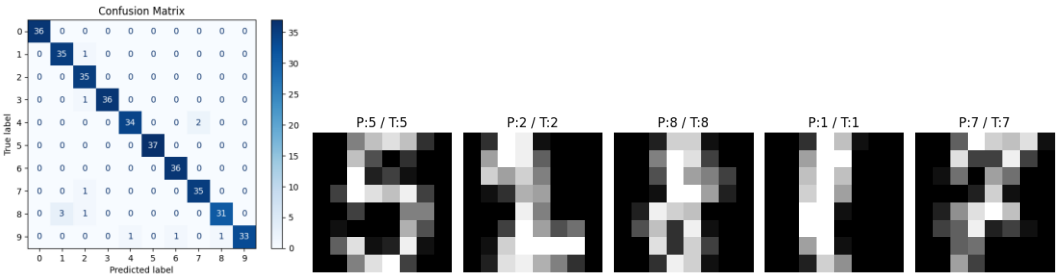
Report

Objective

The objective of this task is to develop a machine learning model that can correctly classify handwritten digits (0–9) using the K-Nearest Neighbors (KNN) algorithm. The task also aims to study how different values of the hyper-parameter *K* affect the classification performance and to identify the best K value for the given dataset.

Methodology

In this task, the handwritten digits dataset from sklearn.datasets was used, where each 8×8 grayscale image is represented as 64 numerical features and labeled 0–9. The dataset was explored by visualizing sample images to understand its structure. The dataset was then split into training and test sets (80:20) with stratification to maintain class distribution. The Features was scaled using StandardScaler since KNN is distance-based. Multiple KNN models were trained with different neighbor values (K = 3, 5, 7, 9) and evaluated using test accuracy. The best K was selected based on the highest accuracy, and a final model was trained with this parameter. Model performance was further assessed using a confusion matrix, and sample predictions were compared with true labels for visual verification.



Conclusion:

This task successfully demonstrates the use of the K-Nearest Neighbors algorithm for handwritten digit classification. The results show that proper data preprocessing, especially feature scaling, plays a crucial role in improving the performance of distance-based models. The experiment also confirms that the choice of the number of neighbors significantly influences model accuracy. By selecting an optimal K value and evaluating the model using accuracy and a confusion matrix, a reliable and interpretable digit classification model was obtained. Overall, the KNN approach proved to be effective for small and well-structured datasets such as the sklearn digits dataset.