Assignment#3 Report

Hamza 22p9222

1 Introduction

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning algorithm used for classification and regression. It classifies a data point based on the majority class of its k nearest neighbors. This report compares the implementation of KNN from scratch with Scikit-learn's built-in KNN classifier.

The dataset is split into training and testing sets for evaluation.

2 KNN Implementation from Scratch

The classifier uses the Euclidean distance formula to determine the nearest neighbors:

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{1}$$

2.1 Implementation Steps

1. Load and preprocess the dataset.
2. Compute the Euclidean distance between data points.
3. Sort the distances and select the k nearest neighbors.
4. Classify based on the majority class of selected neighbors.
5. Evaluate performance.

3 Accuracy Results

Model	**Accuracy**
KNN from Scratch	`(0.60)`
Scikit-learn KNN	`(0.60)`

Table 1: Accuracy Comparison of Both Models

4 Confusion Matrix

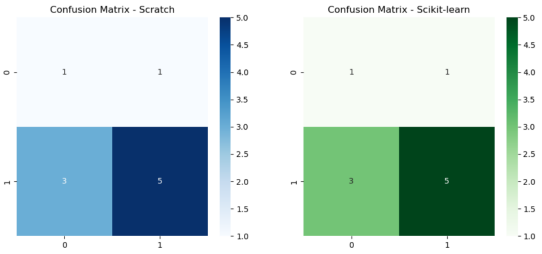

Figure 1: Confusion Matrix

5 KNN accuracy comparison

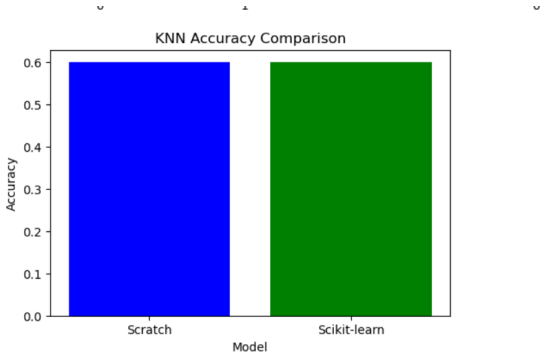

Figure 2: Accuracy

6 Conclusion

Both implementations show similar accuracy, but Scikit-learn's version is optimized for speed and scalability. The custom implementation provides deeper insight into how KNN works.