Brandon Timok

CS422 Machine Learning

Junggab Son

Homework 2

# K-Nearest Neighbors (KNN) Implementation

## Explanation

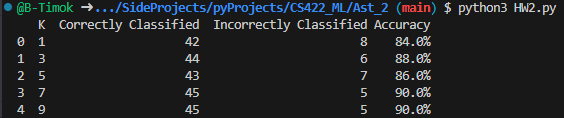
This program implements K-nearest neighbors (KNN) algorithm from scratch using Python.   
It takes two datasets, MNIST\_training.csv and MNIST\_test.csv, and follows the steps below:  
   
1. Load the training and test data using pandas.  
2. Calculate the Euclidean distance between test and training data using numpy.  
3. Finds the K-nearest neighbors and decide the majority class using numpy and Counter.  
4. Compares the prediction with the ground truth in the test data using numpy.  
5. Computes accuracy by counting correctly and incorrectly classified samples using numpy.  
6. Stores the results in a DataFrame and prints it using pandas.  
7. Saves the results in a Word document using the python-docx library.

## Results

|  |  |  |  |
| --- | --- | --- | --- |
| K | Correctly Classified | Incorrectly Classified | Accuracy |
| 1 | 42 | 8 | 84.0% |
| 3 | 44 | 6 | 88.0% |
| 5 | 43 | 7 | 86.0% |
| 7 | 45 | 5 | 90.0% |
| 9 | 45 | 5 | 90.0% |

## Screenshots

The program output is shown below:



## What I Learned

In this assignment, I learned how to implement the K-nearest neighbors (KNN) algorithm from scratch using Python.  
I learned how to calculate the Euclidean distance between test and training data using numpy.  
I learned how to find the K-nearest neighbors and decide the majority class using numpy and Counter.  
I learned how to compare the prediction with the ground truth in the test data using numpy.  
I also learned how to create a Word document using the python-docx library and add tables and images to it.