CS 622 – Homework 4

Shen Chan Huang

April 12 2023

1 Python Version and Libraries Used

- Python Version: 3.8.13

- Pandas

- Numpy

- scikit-learn

2 Dataset

The dataset we are working on is the 'MNIST_HW4.csv'. It contains 1000 rows of data with a column as label and pixel values in other columns.

3 Description

To run the script, go to cmd and use "python -u Huang_HW4_CS622.py"

In this assignment, I implemented 5 fold cross-validation with the SVM model using 'linear', 'rbf', and 'poly'. The problem is to model using Support Vector Machine for classifying digits with different kernels.

First, I loaded the data set. Then I initialized standardization and minmax normalization. I then declared the models I am going to use and defined the k-fold cross-validation. Then I iterated through the splits.

4 Experimental Results

Fold	accuracy
1	0.865
2	0.855
3	0.89
4	0.88
5	0.864
Mean	0.871

(a) Linear kernel without normalization.

Fold	accuracy
1	0.805
2	0.83
3	0.885
4	0.835
5	0.824
Mean	0.836

(b) Poly kernel without normalization.

Fold	accuracy
1	0.895
2	0.86
3	0.925
4	0.89
5	0.874
Mean	0.889

(c) RBF kernel without normalization

Table 1: Accuracy scores of different kernels without normalization.

Fold	accuracy
1	0.855
2	0.825
3	0.885
4	0.835
5	0.829
Mean	0.846

(a) Linear kernel with standardization.

Fold	accuracy
1	0.5
2	0.505
3	0.66
4	0.48
5	0.568
Mean	0.543

(b) Poly kernel with standardization.

Fold	accuracy
1	0.835
2	0.825
3	0.865
4	0.805
5	0.789
Mean	0.824

(c) RBF kernel with standardization.

Table 2: Accuracy scores of different kernels with standardization.

Fold	accuracy
1	0.87
2	0.855
3	0.895
4	0.88
5	0.859
Mean	0.872

(a) Linear kernel with min max normalization.

Fold	accuracy
1	0.805
2	0.825
3	0.885
4	0.835
5	0.824
Mean	0.835

(b) Poly kernel with min max normalization.

Fold	accuracy
1	0.895
2	0.86
3	0.92
4	0.85
5	0.844
Mean	0.874

(c) RBF kernel with min max normalization.

Table 3: Accuracy scores of different kernels with min max normalization.