HA2 – MNIST Dataset

Mathias Løkkebø Øvreseth

Computer engineering

Norwegian University of Science and Technology

Ålesund, Norway

mathilov@stud.ntnu.no

Michal Åsebø Berg

Computer engineering

Norwegian University of Science and Technology

Ålesund, Norway

michalb@stud.ntnu.no

Contents

INTRODUCTION

The MNIST dataset contains a training set of 60,000, and a test set of 10,000 examples of handwritten digits ranging from zero to nine (Yann LeCun, Courant Institute, NYU, Corinna Cortes, Google Labs, New York 2022) . In this report we will compare how three local models compare to a global model. The local models are trained on a subset of all the available data. The models used Is: MLPClassifier, linear Perceptron and KNeighborsClassifier from sklearn

METHODS

Data description

This work used the publicly available MNIST dataset and was imported using python from TensorFlow's repository of datasets. The dataset was then split into three local sections and grouped together by digits ranging from 0-2, 3-5 and 6-9. The models were evaluated using confusion matrix, precision score and recall score, and the results were used to make comparisons between the quality of the models.

Prediction model description

Each subset of the data was trained using a different model from sklearn. The digits ranging from 0-2 was trained using MLPClassifier, digits ranging from 3-5 using Perceptron, and 6-9 using KneighborsClassifier. Furthermore, the global model was trained using the MLPClassifier. A hard voting classifier was used to train on the local models and get a combined output based on the majority predicted value from the local models.

RESULTS

Learner 1 – digits 0-2:

Learner 1 used a MLP model for predicting and delivered precision score of 99.367% and a recall score of 99.364%. It produced the following confusion matrix.

```
Confusion matrix for learner 1 MLP:
[[ 978     1     1]
 [   1  1128     6]
 [  10     1  1021]]
Precision score = 99.367 %
Recall score = 99.364 %
```

Figure 1: Confusion matrix, precision score and recall score for learner 1

Learner 2 – digits 3-5:

Learner 2 used a linear perceptron model and delivered a precision score of 95.948% and a recall score of 95.908%. It produced the following confusion matrix.

```
Confusion matrix for learner 2 linear perceptron:
[[981    4  25]
 [  6  971   5]
 [ 64   14 814]]
Precision score = 95.948 %
Recall score = 95.908 %
```

Figure 2: Confusion matrix, precision score and recall score for learner 2

Learner 3 – digits 6-9:

Learner 3 used a KNeighbor classifier model and delivered a precision score of 98.668% and a recall score of 98.665%. It produced the following confusion matrix.

```
Confusion matrix for learner 3 Kneighbor:
[[ 957     1     0     0]
 [   0  1013     1    14]
 [   8     7   954     5]
 [   1    12     4   992]]
Precision score = 98.668 %
Recall score = 98.665 %
```

Figure 3: Confusion matrix, precision score and recall score for learner 3

Global model – digits 0-9:

The global model used the highest scoring model of all the three learners, the MLP model, and delivered precision score of 95.027% and a recall score of 95.010%. It produced the following confusion matrix.

```
Confusion matrix for learner 4 MLP Global model:
[[ 964    0    1    4    1    2    3    0    3    2]
 [   1 1108    2    7    0    1    1    2   13    0]
 [   5    3  978   14    7    1    4   13    6    1]
 [   4    0    7  961    2   15    1    8   12    0]
 [   2    0    3    0  941    0   10    3    2   21]
 [   4    1    1   24    1  825   11    2   11   12]
 [   6    2    1    1    9    8  924    0    7    0]
 [   1    7   13    9    5    1    0  969    1   22]
 [   5    3    7   13    4   10    7    5  910   10]
 [   4    4    0    8   44    3    2    8   15  921]]
Precision score = 95.027 %
Recall score = 95.010 %
```

Figure 4: Confusion matrix, precision score and recall score of the global learner

The Voting classifier using the three local models produced the following confusion matrix and had an accuracy of 95.86%
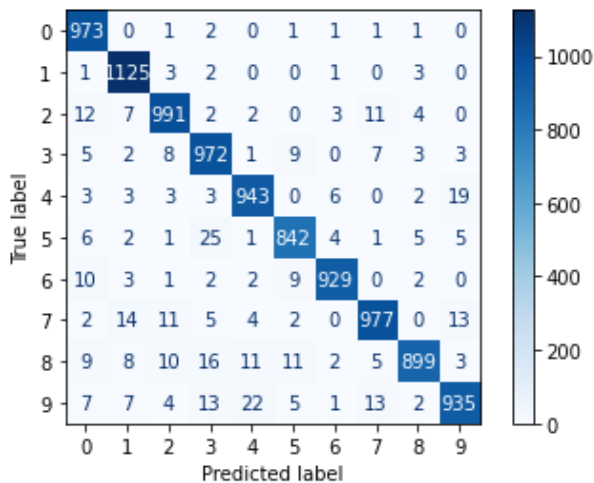
DISCUSSION

Using the voting classifier, the performance of the model improved. The voting classifier is an optimal way to boost the accuracy of a model. The linear perceptron model provided the worst accuracy from the subsets while the MLPclassifier had the best.When using the MLP on the entire dataset, it was observed that the precision score fell from 99,367% to 95.027%. This indicates that the digits range had an impact on both the precision score and the recall score performance of the three models used.

I. BIBLIOGRAFI

Yann LeCun, Courant Institute, NYU, Corinna Cortes,
Google Labs, New York. 2022. *Yann LeCun.*
Funnet Oktober 22, 2022.
http://yann.lecun.com/exdb/mnist/.