Programming Assignment 2

Put Title Here

Group No. 3

Sabo Vladimír

Hoxhallari Kejsi

Bleiberschnig Jakob

May 15, 2022

# Task 1 – Linear SVM

## Hyperparameter-Selection

The hyperparameters which can be adjusted for all three datasets are the following.

* Learning Rate (l-rate)
* Regularization Parameter (reg-param)
* Number of SGD-Epochs (epochs)

The adjustment of the learning rate and regularization parameter was done by repeatedly changing one parameter while the others stay fixed. This process was repeated until a local maximum of the accuracy was found. The number of epochs was selected by looking at the convergence plot. The number of epochs was selected such that the accuracy stayed stable for a couple of epochs.

With this approach the following hyperparameter were selected for the three datasets.

|  |  |  |
| --- | --- | --- |
| Toy-Dataset Tiny | Toy-Dataset Large | MNIST Dataset |
| * L-rate: 1 * reg-param: 0.001 * epochs: 10 | * L-rate: 1 * reg-param: 0.001 * epochs: 10 | * L-rate: 0.001 * reg-param: 0.001 * epochs: 10 |

## Convergence

The convergence was determined by the classification accuracy after each epoch. The following graphs show the convergence results. For the two toy-datasets test set of size 20% of the total number of samples is used to calculate the convergence. For the MNIST-dataset a validation set of size 14.3% (10000 samples) of the total number of samples was used for the calculation of the accuracies.

The training error would be 1-accuracy in this context.

|  |  |
| --- | --- |
| Toy-Dataset-Tiny | Toy-Dataset-Large |
| Chart, line chart  Description automatically generated |  |
| MNIST-Dataset |  |
| Chart, line chart  Description automatically generated |  |

## Accuracy

The results for the accuracy are shown in the following list.

* Toy-Dataset-Tiny: Accuracy (cross validation) = 0.99
* Toy-Dataset-Large: Accuracy (cross validation) = 0.99
* MNIST-Dataset: Accuracy (on test set) = 0.90

## Runtime

Here the time for fitting the model with the training data is considered as runtime. The times are shown in the following list. In case of cross validation for the toy datasets the average fitting time over all five data set combinations is taken.

* Toy-Dataset-Tiny: Runtime = 0.0087s
* Toy-Dataset-Large: Runtime = 6.75s
* MNIST-Dataset: Runtime (on training set) = 22.93s

## Description of the Implementation

The linear SVM is implemented in two classes. One for binary classification (LinearSVM) and one for multi-classification (LinearSVM\_MC). Each class has a “fit” and a “predict” function.

The former has the training data and the hyperparameters as input (in case of multi-classification the number of classes are passed as well) and returns the weights of the fitted SVM model after the last epoch and an array with the weights after each epoch. The array is later used to plot the convergence.

The predict function takes the dataset of interest (plus labels) and the weights as input and returns the predicted labels.

This program of the linear SVM implements the following optimization problem using SGD. The features get extended by a column of ones to incorporate the intersection parameter b into w.

Diagram

Description automatically generated with medium confidence

The SGD update scheme for the weights w which is used in this implementation can be described as follows.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Initialize as a random vector, with d as the feature dimension
2. For each epoch do
3. shuffle the dataset
4. For each sample in the dataset do
5. where is the learning rate
6. Return

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In case of the multiclass SVM the vector becomes a matrix with c as the number of classes.

The gradient of one sample with respect to is computed as

Diagram, schematic

Description automatically generated

or in the multiclass case as

Text, letter

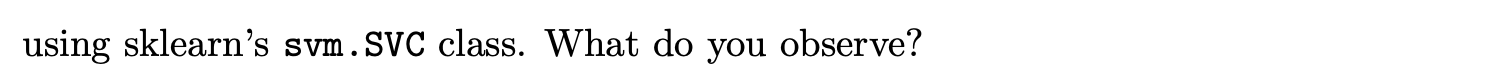
Description automatically generated

# Task 2

## Subheading

Text, letter

Description automatically generated



# Task 3

## Subheading

Text, letter

Description automatically generated

# Working time and Work-sharing

|  |  |
| --- | --- |
| Task | Person |
| Implementation of basic algorithm | Sabo, Hoxhallari, Bleiberschnig |
| Task1 Evaluation | Bleiberschnig |
| Task2 Implementation + Evaluation | Hoxhallari |
| Task3 Implementation + Evaluation | Sabo |
| Report | Sabo, Hoxhallari, Bleiberschnig |

Total working hours :

References