École Polytechnique Fédérale de Lausanne

EPFL

Introduction to Machine Learning CS-233(a)

Milestone 1 Report

Ann-Kristin Bergmann, 362802

Caspar Martin Schön, 362091

Adrian Zvizdenco, 362739

November 11, 2022

1 Logistic Regression

The accuracy of classification using Logistic Regression, depends in general on 2 hyper-parameters (**learning rate** and **maximum iteration number**). We decided to fix the maximum iteration parameter to 1000 iterations and apply cross-validation on the learning rate. The results of different search parameters in the range 10^{-2} to 10^{-5}, are plotted below, with the Macro F1-score on the y-axis. The best hyper-parameter found is 0.0001. Another optimization we applied in the algorithm, is to stop the gradient descend algorithm, whenever the accuracy does not further improve (commented as it fails one pre-coded test). That allowed the model to predict the test labels with up to $\sim 78\%$ accuracy and a 0.76 F1-score, but we noticed that the cross-entropy loss of the final trained model is around ~ 1260. Having several iterations of the logistic regression re-run, we observed that a higher accuracy in the test validation, would correspond to a higher cross-entropy
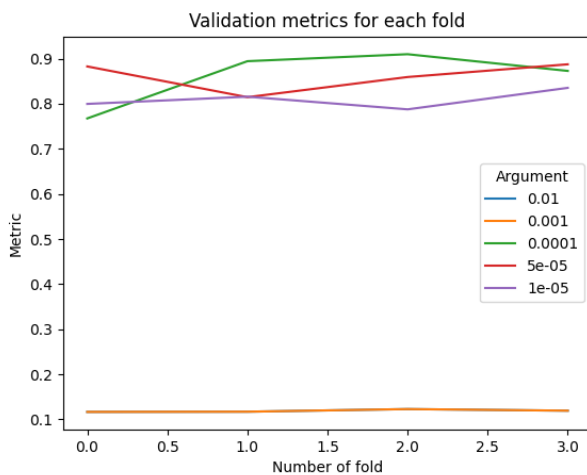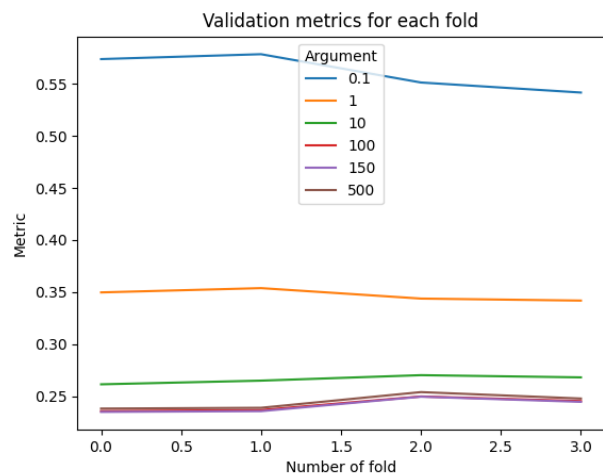

Figure 1: Cross-validation for learning rate

Figure 2: Cross-validation for ridge λ

```
Best hyperparameter found is -> 0.0001
Metric for the best hyperparameter -> 0.8610786156492214

Started Logistic Regression training with learning rate 0.0001 and max iterations 1000

Training accuracy at iteration 0 is 0.6512166859791425
Final accuracy after training is 0.9165701042873696

Runtime of Logistic Regression training: 14.186721086502075 sec

Final classification accuracy is 77.831715210356
Final macro F1 score is 0.764447313911666
```

```
Best hyperparameter found is -> 150
Metric for the best hyperparameter -> 0.239901975251897

FIT & PREDICT phase ...
Started Ridge/Linear Regression training with lambda 150

MSE value for ridge/linear regression training 0.16365054275831062
Final loss is 0.36744674733568694
```

2 Ridge/Linear Regression

The Ridge and Linear Regression are combined within the same class, where the hyper-parameter is the **regularization influence** λ. A value of $\lambda = 0$, would correspond to the Linear Regression. In order to find the best hyper-parameter for the regression problem, cross-validation was applied to a range of arguments from 0.1 to 500. The best hyper-parameter found was 150, having the least *MSE* loss. The results of cross-validation, can be visualized in the plot above, having the *MSE* metric on x-axis and the number of the fold as y-axis. Finally, the *MSE* loss of the trained Ridge Regression was lowered to 0.369. For the Linear Regression, the final *MSE* loss of the predictions is 4073.2175, showing a numerical instability of the dataset.