École Polytechnique Fédérale de Lausanne

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 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 as this value showed improvements for some learning rates 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. An optimization we applied in the algorithm, is to stop the gradient descend algorithm, whenever the accuracy gets stuck to a certain value (in the Logistic *fit()* function it's commented in order to pass all tests). That allowed the model to predict the test labels with up to $\sim 83\%$ accuracy and a 0.78 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 classification, would correspond to a higher cross-entropy loss in training. It can be concluded that by stopping earlier the gradient descend algorithm, we avoid over-fitting the model and therefore, improve it's test validation metric.

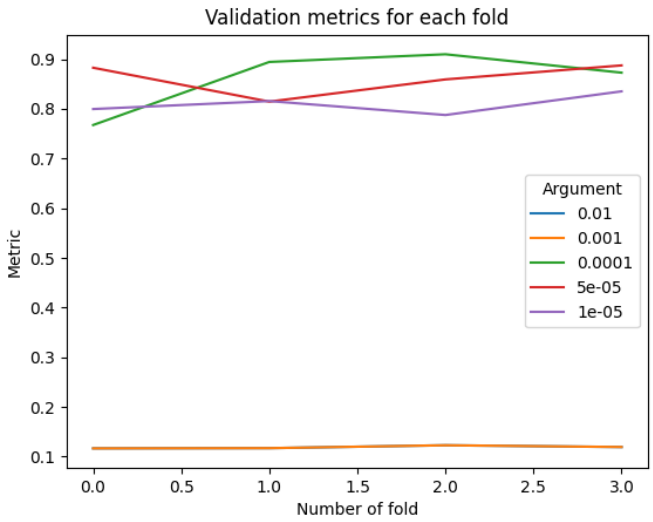

Figure 1: Cross-validation for learning rate

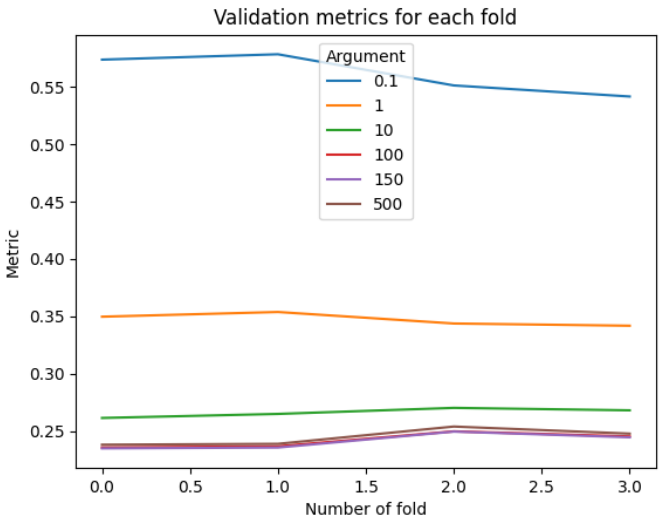

Figure 2: Cross-validation for ridge λ

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 data set.

Some of the console outputs we obtained are screenshot and saved in the **output** folder in the project