

Part I: Pen and paper

- 1.) FILL ME
- 2.) FILL ME
- 3.) FILL ME
- 4.) FILL ME

Part II: Programming

- 5.) To compare the performance of a Linear Regression Model and two MLP Regressors with 2 hidden layers of 10 neurons each, except one has no activation function while the other has ReLU activation functions, we averaged out the performance of each model over 10 separate runs on different 80-20 train-test splits. The resulting boxplots for each model are:

MAE Comparison of Linear Regressor, MLP without activation and MLP with ReLU activation

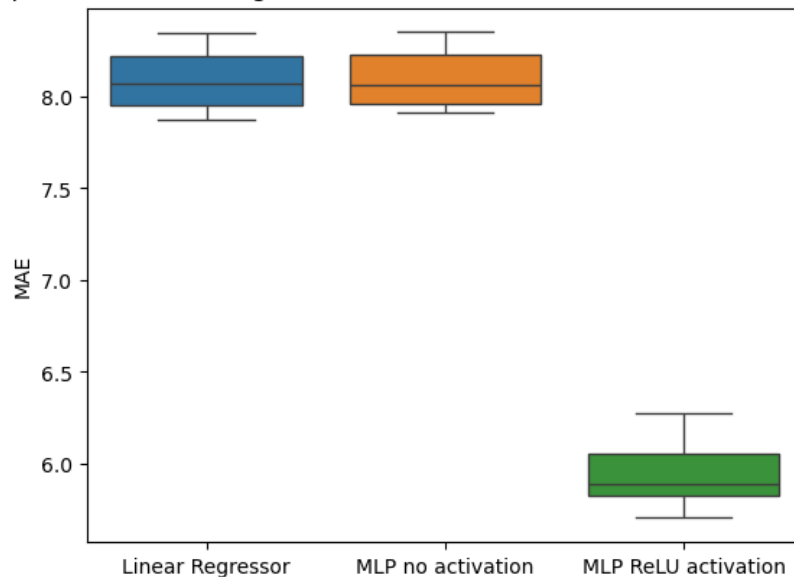


Figure 1: Boxplot graphs for the Mean Absolute Error of different regression models, accross 80-20 train-test splits.

- 6.) As expected the Linear and MLP without activation regressors have high similar absolute errors, since they can only be used in linearly separable datasets. By adding an activation function, such as ReLU, on the MLP model we are able to learn non-linear patterns within the dataset, which obviously results in a lower absolute error.

The reason that no activation function, which in reality is the default sign function, leads to bad results is because it creates a linear decision boundary, which is not suitable for non-linear datasets, that can't be separated by simple lines, but rather complex non-linear boundaries.

- 7.) MLP regressors have varied hyperparameters that can be tuned to improve their performance. To find the best combination of hyperparameters, within the choices presented in the assignment, we used a grid search approach, with a single split, applied on all combinations. The results were as follows:

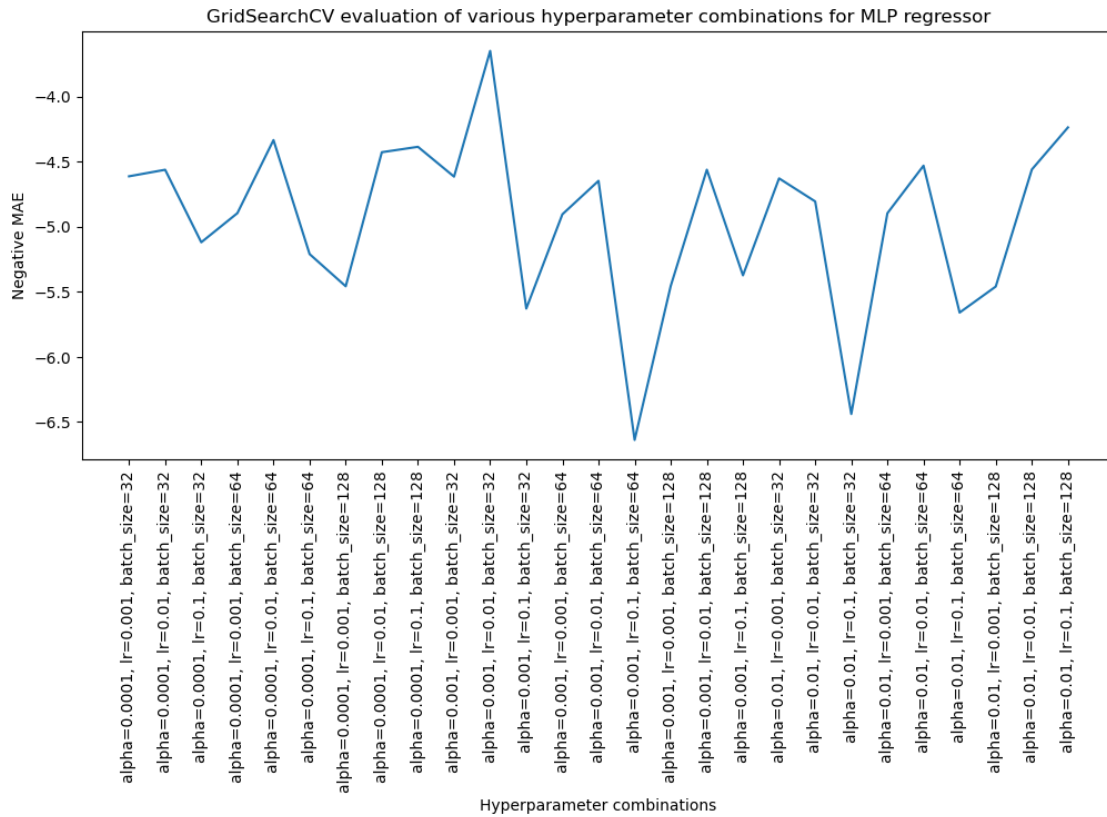


Figure 2: Performance of different hyperparameters combinations on the MLP Regressor

As it can be seen on the plot, the best L2 penalty - Learning Rate - Batch size combination is the one that leads to the lowest Mean Absolute Error, which is the following combination:

- L2 penalty: 0.001
- Learning Rate: 0.01
- Batch Size: 32

MISSING: DISCUSSION OF TRADE-OFFS BETWEEN COMBINATIONS