**Abstract**

One of the most basic techniques in Machine Learning is Linear Regression. In this project, we therefore implemented a closed-form algorithm as well as a gradient-descent algorithm to perfect our understanding of basic Machine Learning concepts.

**Introduction**

The trends behind popular content on the internet are often extremely hard to predict. In this project, we hence decided to build a machine learning model that aims at predicting the “popularity score” of comments on the social network “Reddit”. We used different features of a dataset of reddit comments (N = 12000) to build a model that uses Linear Regression to predict their “popularity scores”.

**Dataset**

**Results**

1) The runtime of the closed-form algorithm, while using the 3 principal features is of approximately 0.9978 ms, whereas the runtime of the gradient-descent approach with the parameter vector w0 initialized to 0 for all features and with a learning rate α of 0.0000065 is of approximately 174.45 ms.

2) Using the closed-form approach, by using 3 features, we obtain a Mean-Squared Error (MSE) of 1.08468307 for the training set, and 1.02032668 for the validation set. By using 63 features, we obtain an MSE of 1.06142257 for the training set, and 0.98696264 for the validation set. By using the 163 features, we obtain an MSE of 1.04871848 for the training set, and 0.99454456 for the validation set.

3)

**Discussion and Conclusion**

**Statement of Contribution**