**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**

These days, we often hear about “influencers” who are about to create “viral” content on the internet. It is however very difficult to pin-point what makes a publication on a social network website popular. In this project, we hence decided to investigate the “popularity score” of comments on the social network “Reddit”. We used different features of a dataset of reddit comments (N = 12000) to perform a linear regression in the hope of developing a model that would allow us to predict the “popularity score” of reddit comments.

**Dataset**

**Results**

1)

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**