CS 4375

ASSIGNMENT \_\_\_\_\_1\_\_\_\_\_\_

Names of students in your group:  
- Trent Hardy (TTH190003)

Number of free late days used: \_\_\_\_\_\_2\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

Data set: Tsanas,Athanasios and Little,Max. (2009). Parkinsons Telemonitoring. UCI Machine Learning Repository. <https://doi.org/10.24432/C5ZS3N>.

Public Link to Dataset: https://cdn.jsdelivr.net/gh/UnknownSilence/Linear-Regression-using-Gradient-Descent@main/parkinsons\_updrs.csv

### **Almost all metrics, logs, plots, and answers on this report can be found in the part1.ipynb and part2.ipynb files. I include them there for convenience, so I recommend looking at those first.**

# Part 1

## MSE Log

A screenshot of a computer

Description automatically generated

## Plot

A screen shot of a chart

Description automatically generated

## Are you satisfied that the package has found the best solution. How can you check. Explain.

In general, a lower MSE indicates a better fit of the model to the data. The R^2 score is a measure of the proportion of the variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1.

Based on the results:

- Best Learning Rate: 0.0001

- Best Number of Iterations: 50

- Test Dataset MSE: 59.97

- R-squared (R2) Score: 0.0604

In this case, the R^2 score is relatively low.

The model's performance on the test dataset, as measured by the Mean Squared Error (MSE), is satisfactory. The MSE of 59.97 suggests that the model is properly fitting the data.

# Part 2

## MSE Log

A screenshot of a computer

Description automatically generated

## Plot

A screen shot of a chart

Description automatically generated

## Are you satisfied that the package has found the best solution. How can you check. Explain.

Weight Coefficients: [ 0.3041745 -0.00082082 0.01059351 0.00449969 0.01814347 0.04549573

0.14280999 -0.01711548 0.12248709]

Best Learning Rate: 0.0001

Best Number of Iterations: 100

Test Dataset MSE: 59.85061219551369

R-squared (R2) Score: 0.062332162529287305  
  
In the process of hyperparameter tuning with the Stochastic Gradient Descent (SGD) Regressor, optimal settings for the learning rate and number of iterations were identified. Despite fine-tuning, the model exhibited modest performance on the test dataset, as evidenced by a Mean Squared Error (MSE) and R^2 score that suggest substantial room for improvement. The weight coefficients for the features were also determined, providing some insights into feature importance. While the model does offer a degree of predictive capability, its relatively high MSE and low R^2 score indicate that further optimization is needed. Additional strategies such as feature engineering, algorithm selection, or broader hyperparameter tuning could enhance the model's predictive accuracy and robustness.