**Machine Learning for Materials Scientists**
*Summer Semester 2024*

**Prof. Bernhard Eidel**
**Narkhede, Rahul Vishnu**
*rahul-vishnu.narkhede@student.tu-freiberg.de*

TECHNISCHE UNIVERSITÄT
BERGAKADEMIE FREIBERG
The University of Resources. Since 1765.

# Exercise 1 - Linear Regression

*May 30, 2024*

1. In the given jupyter notebook `Linear regression.ipynb` update the functions for `update_weights_sgd` `update_weights_mbgd` and `update_weights_mbgd_regularized`. A batch size of 8 is given as a default, but the function should able to return the updated feature values for any batch size. Use these functions and perform gradient descent on the given dataset. The learning rate and the $\lambda$ value might also have to be changed to obtain a low loss value.

2. The BostonHousing dataset has been provided to you. For this dataset, calculate the parameters of Linear regression model following the same steps as in the case of the Exercise dataset. The following tasks need to be completed:

   (a) Conduct basic data analysis to determine four features that have the highest correlation with the target feature **medv**

   (b) Split the dataset into train and test datasets. Pick a reasonable ratio (80% train and 20% test). Only use the train dataset to calculate the model parameters.

   (c) Create functions for model setup and put them all together to train your model. Implement the three variants of gradient descent: batch gradient descent, stochastic gradient descent and mini-batch gradient descent.

   (d) Calculate the loss of your model on the test as well as the train datasets. Try this with varying values of the regularization variable and the three variants of gradient descent.

Update the code and the analysis of the dataset in the jupyter notebook which has been provided for the exercise. Additionally, provide a pdf document(based on LaTeX) explaining the results with:

1. Plots for change in loss value with number of iterations/steps of gradient descent.
2. Plots for change in loss value on test and train datasets for different values of regularization parameter $\lambda$.
3. Provide the parameters (learning rate, number of epochs, regularization parameter) and the test loss of your final best performing model.

***Deadline: 12-June-2024, 23:59h***