1 Individual / small group task: Training & Prediction error

You are given the data set

$$\mathcal{T} = \{(0,1),(1,1),(2,1),(3,3),(4,4)\}$$

that you split into the training data set

$$\mathcal{T}_{train} = \{(0,1),(2,1),(4,4)\}$$

and the validation data set

$$\mathcal{T}_{val} = \mathcal{T} \setminus \mathcal{T}_{train}\,.$$

1. Compute the linear model on the training data.
2. Evaluate the training error.
3. Evaluate the (expected) generalization error according to the validation set approach.

2 Implementing training & prediction error estimation

In an earlier live lecture, we loaded a molecular energy data set and created representations for the inputs. Use the given Jupyter notebook for this data set.

1. Split the given 500 data samples in the first 400 samples as training set and the last 100 samples as a validation set.
2. Build the kNN regression predictor for $k = 1, 3$ and compute the training error and the (expected) generalization error using the validation set approach.
3. Build the linear model by least squares and compute the training error and the (expected) generalization error using the validation set approach.