# Machine Learning

Spring Semester 2023
Prof. Dr. Peter Zaspel

## Assignment Sheet 7.

Submit on **Tuesday, March 28, 2023, 10:00**.

**Excercise 1.** (Training error)

In this task, we consider the training data

$$\mathcal{T} = \left\{ \left((1,1)^\top, 15\right), \left((1,3)^\top, 24\right), \left((3,2)^\top, 32\right), \left((4,4)^\top, 16\right), \left((5,5)^\top, 15\right), \left((5,7)^\top, 21\right), \left((6,3)^\top, 12\right), \left((7,6)^\top, 6\right), \left((8,1)^\top, 18\right) \right\}$$

a) Build a predictor using kNN regression for $k = 6$ and evaluate the training error of the predictor.

b) Build a predictor using linear regression by least squares and evaluate the training error of the linear model.

(4 Points)

**Excercise 2.** (Cross validation)

In this task, we consider the data set

$$\mathcal{T} = \left\{ (1,1), (3,2), (4,5), (-4,-3), (-3,-4), (-1,-1) \right\}.$$

a) Evaluate the (expected) generalization error of the kNN regressor with $k = 2$ by $\mathcal{K}$-fold cross validation with $\mathcal{K} = 3$. (Do a deterministic, i.e. not randomized, splitting of the given data following the ordering of the samples.

b) Evaluate the (expected) generalization error of the linear model by leave-one-out cross validation. (It is fine to use the help of a computer to fit the individual linear models.)

(4 Points)

**Excercise 3.** (Observations on the training error)

Provide quantitative reasoning to the questions below.

a) What is the training error of kNN regression with neighbourhood size $k = 1$? Give the result and an explanation of how you get to the result.

b) What is the training error of linear regression with an output dimension of $K = 1$ for $N = 2$ distinct training samples? Give the result and an explanation of how you get to the result.

(4 Points)

**Programming Exercise 1.** (Implementation of gerneralization error evlauation)

In this programming exercise, you will implement the validation set approach and $\mathcal{K}$-fold cross validation, while recalling that leave-one-out cross validation is $\mathcal{K}$-fold cross validation for $\mathcal{K} = N$, if $N$ is the number of training samples.

Implement the study carried out in Example 6.3 from the lecture notes. Note that due to randomization, you might get other results than the ones shown in the example.

Reference solutions will only be provided in Python+Matplotlib. The submission format for Python is a Jupyter notebook. The submission format for C/C++ is standard source files. Choose an appropriate format for the Gnuplot-related submission.

(4 Points)