

# **Machine Learning: Assignment Sheet #7**

Due on April 5, 2022 at 10:00

**Group HB**

Henri Sota, Enis Mustafaj

## Problem 7.1

In this task, we consider the training data

$$\mathcal{T} = \{((1, 7)^\top, 25), ((2, 5)^\top, 21), ((2, 6)^\top, 14), ((3, 3)^\top, 32), ((7, 1)^\top, 14), ((3, 1)^\top, 14), ((4, 2)^\top, 25), ((5, 4)^\top, 18), ((5, 6)^\top, 12)\}$$

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

The predictor for kNN regression with  $k = 6$  is

$$f(x) = \frac{1}{6} \cdot \sum_{x_i \in N_6(x)} y_i$$

where  $N_6(x)$  are the 6 nearest neighbours of point  $x$ . We compute the predicted values for each input data:

$$\begin{aligned} f(x_1) &= \frac{25+14+21+12+32+18}{6} = 20.33 & f(x_2) &= \frac{21+14+25+12+32+18}{6} = 20.33 & f(x_3) &= \frac{21+14+25+12+32+18}{6} = 20.33 \\ f(x_4) &= \frac{21+14+25+14+32+18}{6} = 20.67 & f(x_5) &= \frac{18+14+25+12+32+14}{6} = 19.17 & f(x_6) &= \frac{21+14+25+18+32+14}{6} = 20.67 \\ f(x_7) &= \frac{21+14+25+14+32+18}{6} = 20.67 & f(x_8) &= \frac{18+21+25+12+32+14}{6} = 20.33 & f(x_9) &= \frac{21+12+25+18+32+14}{6} = 20.33 \end{aligned}$$

Now we compute the training error by the formula:

$$TE(f, \mathcal{T}) = \frac{1}{N} \cdot \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{9} \cdot \sum_{i=1}^9 (y_i - f(x_i))^2$$

By substituting the values from above, we get  $TE(f, \mathcal{T}) = 39.51$

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

Firstly, we need to find the coefficients of the model by the formula:

$$\beta = (\mathcal{X}^T \cdot X)^{-1} \cdot X^T \cdot y$$

where

$$\mathcal{X} = \begin{pmatrix} 1 & 1 & 7 \\ 1 & 2 & 5 \\ 1 & 2 & 6 \\ 1 & 3 & 3 \\ 1 & 7 & 1 \\ 1 & 3 & 1 \\ 1 & 4 & 2 \\ 1 & 5 & 4 \\ 1 & 5 & 6 \end{pmatrix} \quad \mathcal{X}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 & 7 & 3 & 4 & 5 & 5 \\ 7 & 5 & 6 & 3 & 1 & 1 & 2 & 4 & 6 \end{pmatrix} \quad y = \begin{pmatrix} 25 \\ 21 \\ 14 \\ 32 \\ 14 \\ 14 \\ 25 \\ 18 \\ 12 \end{pmatrix}$$

By solving the linear equation:

$$\mathcal{X}^T \cdot \mathcal{X} \cdot \beta = \mathcal{X}^T \cdot y$$

We get the vector of coefficients  $\beta = \begin{pmatrix} \frac{55170}{1831} \\ -\frac{7465}{3662} \\ -\frac{1619}{1831} \end{pmatrix}$  and the predictor  $f(x) = x^T \cdot \beta$

Now, we calculate the the predicted values for each of our input data

$$\begin{aligned} f(x_1) &= x_1^T \cdot \beta = 21.9 & f(x_2) &= x_2^T \cdot \beta = 21.6 & f(x_3) &= x_3^T \cdot \beta = 19.7 \\ f(x_4) &= x_4^T \cdot \beta = 21.4 & f(x_5) &= x_5^T \cdot \beta = 15 & f(x_6) &= x_6^T \cdot \beta = 23.1 \\ f(x_7) &= x_7^T \cdot \beta = 20.2 & f(x_8) &= x_8^T \cdot \beta = 16.4 & f(x_9) &= x_9^T \cdot \beta = 14.6 \end{aligned}$$

Now we compute the training error by the formula:

$$TE(f, \mathcal{T}) = \frac{1}{N} \cdot \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{9} \cdot \sum_{i=1}^9 (y_i - f(x_i))^2$$

By substituting the values from above in the formula, we get  $TE(f, \mathcal{T}) = 31.7$

## Problem 7.2

In this task, we consider the data set

$$\mathcal{T} = \{(-2, 4), (2, 4), (1, 1), (-1, 1), (0, 0), (3, 9)\}$$

- 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.) Following Algorithm 10 from the Lecture Notes for estimating the generalization error using  $\mathcal{K}$ -fold cross validation:

1. Split  $\mathcal{T}$  into equally sized sets  $\mathcal{T}_1, \dots, \mathcal{T}_{\mathcal{K}}$
2. For  $i \in \{1, \dots, \mathcal{K}\}$ :
  - a)  $\mathcal{T}_{\text{train}} \leftarrow \mathcal{T} \setminus \mathcal{T}_i$
  - b)  $f_{\mathcal{T}_{\text{train}}} \leftarrow \text{train}(f, \mathcal{T}_{\text{train}})$
  - c)  $\varepsilon_i \leftarrow \frac{1}{|\mathcal{T}_i|} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_i} L(\mathbf{y}, f_{\mathcal{T}_{\text{train}}}(\mathbf{x}))$
3.  $\varepsilon \leftarrow \frac{1}{\mathcal{K}} \sum_{i=1}^{\mathcal{K}} \varepsilon_i$
4. Return  $\varepsilon$

The following splits are created deterministically while using  $\mathcal{K}$ -fold cross validation with  $\mathcal{K} = 3$ :

| Split | Training Set $\mathcal{T}_{\text{train}}$ | Test Set $\mathcal{T}_i$ |
|-------|---|--------------------------|
| 1     | $\{(1, 1), (-1, 1), (0, 0), (3, 9)\}$     | $\{(-2, 4), (2, 4)\}$    |
| 2     | $\{(-2, 4), (2, 4), (0, 0), (3, 9)\}$     | $\{(1, 1), (-1, 1)\}$    |
| 3     | $\{(-2, 4), (2, 4), (1, 1), (-1, 1)\}$    | $\{(0, 0), (3, 9)\}$     |

- **Split 1:**

| Data Point | True | Predicted |
|------------|------|-----------|
| (-2)       | (4)  | (0.5)     |
| (2)        | (4)  | (5)       |

Mean squared error for the split is  $\varepsilon_1 = 6.625$ .

- **Split 2:**

| Data Point | True | Predicted |
|------------|------|-----------|
| (1)        | (1)  | (2)       |
| (-1)       | (1)  | (2)       |

Mean squared error for the split is  $\varepsilon_2 = 1$ .

- **Split 3:**

| Data Point | True | Predicted |
|------------|------|-----------|
| (0)        | (0)  | (1)       |
| (3)        | (9)  | (2.5)     |

Mean squared error for the split is  $\varepsilon_3 = 21.625$ .

The estimation of the (expected) generalization error is  $\varepsilon = 9.75$ .

- 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.)

Following Algorithm 9 from the Lecture Notes for estimating the generalization error using Leave One Out cross validation:

- For  $i \in \{1, \dots, N\}$ :
  - $\mathcal{T}_{\text{train}} \leftarrow \mathcal{T} \setminus \{(\mathbf{x}_i, \mathbf{y}_i)\}$
  - $f_{\mathcal{T}_{\text{train}}} \leftarrow \text{train}(f, \mathcal{T}_{\text{train}})$
  - $\varepsilon_i \leftarrow L(\mathbf{y}_i, f_{\mathcal{T}_{\text{train}}}(\mathbf{x}_i))$
- $\varepsilon \leftarrow \frac{1}{K} \sum_{i=1}^N \varepsilon_i$
- Return  $\varepsilon$

The following splits are created while using Leave One Out cross validation:

| Split | Training Set $\mathcal{T}_{\text{train}}$      | Test Set $\mathcal{T}_i$ |
|-------|--|--------------------------|
| 1     | $\{(2, 4), (1, 1), (-1, 1), (0, 0), (3, 9)\}$  | $\{(-2, 4)\}$            |
| 2     | $\{(-2, 4), (1, 1), (-1, 1), (0, 0), (3, 9)\}$ | $\{(2, 4)\}$             |
| 3     | $\{(-2, 4), (2, 4), (-1, 1), (0, 0), (3, 9)\}$ | $\{(1, 1)\}$             |
| 4     | $\{(-2, 4), (2, 4), (1, 1), (0, 0), (3, 9)\}$  | $\{(-1, 1)\}$            |
| 5     | $\{(-2, 4), (2, 4), (1, 1), (-1, 1), (3, 9)\}$ | $\{(0, 0)\}$             |
| 6     | $\{(-2, 4), (2, 4), (1, 1), (-1, 1), (0, 0)\}$ | $\{(3, 9)\}$             |

- **Split 1:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 1 + 2x$$

The true value of the data point  $(-2)$  is 4. Predicted value for the data point  $(-2)$  is  $-3$ .

Loss using mean squared error for the split is  $\varepsilon_1 = 49$ .

- **Split 2:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 2.78378378 + 1.08108108x$$

The true value of the data point  $(2)$  is 4. Predicted value for the data point  $(2)$  is 4.94594595.

Loss using mean squared error for the split is  $\varepsilon_2 = 0.89481373$ .

- **Split 3:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 3.1627907 + 1.09302326x$$

The true value of the data point  $(1)$  is 1. Predicted value for the data point  $(1)$  is 4.25581395.

Loss using mean squared error for the split is  $\varepsilon_3 = 10.6003245$ .

- **Split 4:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 2.86486486 + 0.91891892x$$

The true value of the data point  $(-1)$  is 1. Predicted value for the data point  $(-1)$  is 1.94594595.

Loss using mean squared error for the split is  $\varepsilon_4 = 0.89481373$ .

- **Split 5:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 3.25581395 + 0.90697674x$$

The true value of the data point  $(0)$  is 0. Predicted value for the data point  $(0)$  is 3.25581395.

Loss using mean squared error for the split is  $\varepsilon_5 = 10.6003245$ .

- **Split 6:** Linear model fit from the training set data has the prediction function:

$$\hat{y} = 0 + 2x$$

The true value of the data point  $(3)$  is 9. Predicted value for the data point  $(3)$  is 2.

Loss using mean squared error for the split is  $\varepsilon_6 = 49$ .

The estimation of the (expected) generalization error is  $\varepsilon = 20.165046077460385$ .

## Problem 7.3

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.

The training error of kNN regression with neighbourhood size  $k = 1$  is 0. In case of  $k = 0$ , the formula of kNN regression becomes:

$$f(x) = \frac{1}{1} \cdot \sum_{x_i \in N_1(x)} y_i = \sum_{x_i \in N_1(x)} y_i$$

where  $N_1(x)$  is the nearest neighbour of the point  $x$ . In this case the nearest neighbour of the point  $x$  is the point itself, so the result of the predictor will be the output of the input data. So, the formula of kNN regression will be:

$$f(x) = y$$

By substituting this on the training error formula we will have:

$$TE(f, \mathcal{T}) = \frac{1}{N} \cdot \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{9} \cdot \sum_{i=1}^9 (y_i - y_i)^2 = 0$$

- 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.

The training error of linear regression with an output dimension of  $K = 1$  for  $N = 2$  distinct training samples is 0, because there exists a line which connects 2 different points. Let's take a set of 2 points  $\{(x_1, y_1), (x_2, y_2)\}$  and try to find the vector of coefficients by the least squares formula:

$$\beta = (\mathcal{X}^T \cdot \mathcal{X})^{-1} \cdot \mathcal{X}^T \cdot y$$

where

$$\mathcal{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix} \quad \mathcal{X}^T = \begin{pmatrix} 1 & 1 \\ x_1 & x_2 \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

By substituting in the formula we get:

$$\beta = \begin{pmatrix} 2 & x_1 + x_2 \\ x_1 + x_2 & x_1^2 + x_2^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} y_1 + y_2 \\ y_1 \cdot x_1 + y_2 \cdot x_2 \end{pmatrix}$$

Since we have only 2 data points  $\mathcal{X}^T \cdot \mathcal{X}$  will always produce a  $2 \times 2$  matrix and its inverse can easily be found

$$\beta = \frac{1}{(x_1 - x_2)^2} \cdot \begin{pmatrix} x_1^2 + x_2^2 & -x_1 - x_2 \\ -x_1 - x_2 & 2 \end{pmatrix} \cdot \begin{pmatrix} y_1 + y_2 \\ y_1 \cdot x_1 + y_2 \cdot x_2 \end{pmatrix}$$

$$\beta = \frac{1}{(x_1 - x_2)^2} \cdot \begin{pmatrix} (y_2 \cdot x_1 - y_1 \cdot x_2) \cdot (x_1 - x_2) \\ (y_1 - y_2) \cdot (x_1 - x_2) \end{pmatrix}$$

$$\beta = \begin{pmatrix} \frac{y_2 \cdot x_1 - y_1 \cdot x_2}{x_1 - x_2} \\ \frac{y_1 - y_2}{x_1 - x_2} \end{pmatrix}$$

By inserting each input data point in the formula:

$$f(x) = \frac{y_2 \cdot x_1 - y_1 \cdot x_2}{x_1 - x_2} + \frac{y_1 - y_2}{x_1 - x_2} \cdot x$$

we get the same output as the initial data point. So, the training error formula becomes:

$$TE(f, \mathcal{T}) = \frac{1}{N} \cdot \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{2} \cdot \sum_{i=1}^2 (y_i - y_i)^2 = 0$$

## Programming Problem 7.1

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.