# Machine Learning

## Exercise Sheet 2

Winter Term 2023/2024 Available: 09.11.2023
Prof. Dr. Niels Landwehr Hand in until: 16.11.2023 11:59am
Dr. Ujjwal Exercise sessions: 20.11.2023/22.11.2023

**Task 1 – Normal Equations for Linear Regression** [15 points]

In this exercise, we will study the simple linear regression example given in the lecture (Slides 39-42) in more detail.
Consider the one-dimensional linear regression model $f : \mathbb{R} \to \mathbb{R}$ given by

$$\begin{aligned} f_{\boldsymbol{\theta}}(\mathbf{x}) &= \theta_1 x_1 + \theta_2 x_2 \qquad (1) \\ &= \theta_1 x_1 + \theta_2 \qquad (2) \end{aligned}$$

where we are assuming a constant attribute $x_2 = 1$ in the instance representation. That is, instances are of the form $\mathbf{x} = (x_1, 1)^T$, where $x_1$ is the one-dimensional input to the model.
We want to train the model on a training data set $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), (\mathbf{x}_3, y_3)\}$ with $N = 3$ instances. The training data can be given in matrix form as

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix} \qquad (3)$$

where the rows of $\mathbf{X}$ contain the instances $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and $\mathbf{y}$ contains the targets $y_1, y_2, y_3$.
Specifically, we want to estimate the least squares model parameters

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \qquad (4) \\ &= \frac{1}{N} \sum_{n=1}^{N} (f_{\boldsymbol{\theta}}(\mathbf{x}_n) - y_n)^2 \qquad (5) \end{aligned}$$

a) Compute the least squares model parameters by solving the normal equations.

b) Verify that the gradient of the loss function $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ is zero at the least-squares solution $\boldsymbol{\theta}^*$.

If you like, you can use a programming environment such as Numpy to solve the matrix equations, but please write down the corresponding equations and give intermediate results to facilitate comparison of solutions.

**Task 2 – Gradient Descent for Linear Regression** [15 points]

In this task, we continue to work with the simple linear regression model $f_{\boldsymbol{\theta}}$ and data $\mathbf{X}, \mathbf{y}$ given in Task 1. Instead of computing least-squares parameters based on the normal equations, we study how to perform gradient descent for this model.

1. Manually carry out two iterations of gradient descent for this model. As starting point for gradient descent, initialize the parameter vector to $\boldsymbol{\theta}_0 = (0, 0)^T$. Then compute new parameter vectors $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ according to the update rule given in the gradient descent algorithm. Use a learning rate of $\eta = 0.01$.

2. Compute the initial loss $L(\boldsymbol{\theta}_0)$ on the training data and the losses $L(\boldsymbol{\theta}_1)$, $L(\boldsymbol{\theta}_2)$ of the two updated models.

Again, you can use a programming environment such as Numpy to solve the matrix equations if you like, but please give intermediate results to facilitate comparison of solutions.

**Task 3 – Positive Semidefinite Matrices** [10 points]

Prove the following remark made in the lecture: For any matrix $\mathbf{A} \in \mathbb{R}^{N \times M}$, the matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$ is positive semidefinite. Reminder: a squared matrix $\mathbf{H} \in \mathbb{R}^{M \times M}$ is positive semidefinite, if for all $\boldsymbol{\theta} \in \mathbb{R}^M$ it holds that $\boldsymbol{\theta}^\mathsf{T}\mathbf{H}\boldsymbol{\theta} \geq 0$.