

Exercise 1:

A medical research group wants to train a supervised regression model to predict how long a patient stays in the hospital after admission (measured in days). The three researchers Holger, Lisa and Son have gathered data on each patient's age (in years), systolic blood pressure (in mmHg, a continuous scale), and weight (in kg). Formally, this gives rise to the target space $\mathcal{Y} = \mathbb{R}$ (where we technically allow for negative values as a prerequisite of the linear model) and the feature space $\mathcal{X} = (\mathbb{R}_0^+)^3$, with $\mathbf{x}^{(i)} = (x_{age}, x_{blood\ pressure}, x_{weight})^{(i)} \in \mathcal{X}$ for $i = 1, 2, \dots, n$ observations. At first, they believe all of these three features should be modeled to have a linear effect on the target variable. To construct their supervised regression model, they remember how most of such models can be described:

Learning = Hypothesis Space + Risk + Optimization.

With the first of the three components of learning in his mind, Holger comes up with the following hypothesis space:

$$\mathcal{H} = \{f : f(\mathbf{x}) = \theta_0 + \theta_1 \cdot x_{age} + \theta_2 \cdot x_{blood\ pressure} + \theta_3 \cdot x_{weight} \mid \boldsymbol{\theta} \in \mathbb{N}^3\} \quad (1)$$

1) Is the hypothesis space formulated above correct? If not, find the correct one.

Lisa recalls the second component of learning: risk. She remembers two slightly different functions for calculating the empirical risk, but is not sure what difference they will make in optimization:

$$\mathcal{R}_{\text{emp}}(f) = \sum_{i=1}^n L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right) \quad (2)$$

$$\bar{\mathcal{R}}_{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^n L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right) \quad (3)$$

2) What difference may the two functions make for optimization?

As a next step, the research group debates the influence the loss function (L1 or L2 loss) they choose has on the three components of learning.

3) Which of the three components of learning are impacted by the choice of the loss function in the case of the linear regression model?

Son suggests to use L2 loss for their linear regression model, and notes this would require no numerical optimization to find the parametrization with lowest empirical risk.

4) Is this true or false? Explain your answer.

After training the linear regression model, Holger proposes to add more flexibility by training a polynomial regression model with degree $d = 1$ for age and systolic blood pressure, and $d = 2$ for weight. Once again, he formulates a hypothesis space:

$$\mathcal{H} = \{f : f(\mathbf{x}) = \theta_0 + \theta_1 \cdot x_{age} + \theta_2 \cdot x_{blood\ pressure} + \theta_3 \cdot (x_{weight} + x_{weight}^2) \mid \boldsymbol{\theta} \in \mathbb{R}^4\} \quad (4)$$

5) Is the hypothesis space formulated above correct? If not, find the correct one.

The research group decides to train the polynomial regression model using L2 loss. Again, Son is worried about the optimization component of learning. He is not entirely sure how to find the parametrization with lowest empirical risk.

6) Explain how the model with the lowest empirical risk can be found in the case of the polynomial regression model with L2 loss.