Machine Learning
Università della Svizzera italiana

Assignment 1

Surname Name

April 6, 2022

The assignment is split into two parts: you are asked to solve a regression problem, and answer some questions. You can use all the books, material, and help you need. Bear in mind that the questions you are asked are similar to those you may find in the final exam, and are related to very important and fundamental machine learning concepts. As such, sooner or later you will need to learn them to pass the course. We will give you some feedback afterwards.

!! Note that this file is just meant as a template for the report, in which we reported **part of** the assignment text for convenience. You must always refer to the text in the README.md file as the assignment requirements.

1 Regression problem

This section should contain a detailed description of how you solved the assignment, including all required statistical analyses of the models' performance and a comparison between the linear regression and the model of your choice. Limit the assignment to 2500 words (formulas, tables, figures, etc., do not count as words) and do not include any code in the report.

1.1 Task 1

Use the family of models $f(\mathbf{x}, \boldsymbol{\theta}) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot \cos(x_2) + \theta_4 \cdot x_1^2$ to fit the data. Write in the report the formula of the model substituting parameters $\theta_0, \ldots, \theta_4$ with the estimates you've found:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \_ + \_ \cdot x_1 + \_ \cdot x_2 + \_ \cdot x_1 \cdot x_2 + \_ \cdot \sin(x_1)$$

Evaluate the test performance of your model using the mean squared error as performance measure.

1.2 Task 2

Consider any family of non-linear models of your choice to address the above regression problem. Evaluate the test performance of your model using the mean squared error as performance measure. Compare your model with the linear regression of Task 1. Which one is **statistically** better?

1.3 Task 3 (Bonus)

In the **Github repository of the course**, you will find a trained Scikit-learn model that we built using the same dataset you are given. This baseline model is able to achieve a MSE of **0.0197**, when evaluated on the test set. You will get extra points if the test performance of your model is better (i.e., the MSE is lower) than ours. Of course, you also have to tell us why you think that your model is better.

2 Questions

2.1 Q1. Training versus Validation

1.Q. Explain the behaviors of the curves in each of the three highlighted sections in the figure, namely (a), (b), and (c);

1.A.

2.Q. Do you think the figure gives useful information to reduce the approximation risk? And to reduce the estimation risk? Explain why.

2.A.

3.Q. Would you think that by further increasing the model complexity you would be able to bring the structural risk to zero? How would your answer change if your data was not affected by noise?

3.A.

4.Q. If the X axis represented the training iterations instead, would you think that the training procedure that generated the figure used early stopping? Explain why. (**NB:** ignore the subfigures and the dashed vertical lines)

4.A.

2.2 Q2. Linear Regression

Comment and compare how the (a.) training error, (b.) test error and (c.) coefficients would change in the following cases:

1.Q. $x_3 = x_1 + 3.0 \cdot x_2$.

1.A.

2.Q. $x_3 = x_1 \cdot x_2 \cdot x_2$

2.A.

3.Q. Can we make any educated guess on what would be the value of θ_3 for each of the preceeding cases if we used Lasso Regression?

3.A.

4.Q. Explain the motivation behind Ridge and Lasso regression and their principal differences.

4.A.

2.3 Q3. Non-Linear Regression

1.Q. Do you think a model of the family $f(x, \theta) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2$ is a good choice for such task? Why?

1.A.

2.Q. Do you think using a feed forward neural network with activation function $h(x) = 1 - 2 \cdot x \cdot e^{-2}$ would improve the results?

2.A.

3.Q. Do you think it would be possible to achieve good performance with a linear model? How?

3.A.

4.Q. What is the purpose of the hidden activation function in a feed forward neural network?

4.A.