Machine Learning

Assignment 1 - Linear Regression

Albert Cerfeda

Contents

27 April 2023
Università della Svizzera italiana
Faculty of Informatics
Switzerland

1   Tasks

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 8-10 pages and do not include any code in the report.

1.1   Task 1

Use the family of models $f(\mathbf{x}, \boldsymbol{\theta}) = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot \sin(x_2) + \theta_4 \cdot x_1 \cdot x_2$ to fit the data.

   a. Write in the report the formula of the model substituting parameters $\theta_0, \ldots, \theta_4$ with the estimates you've found:
$$f(\mathbf{x}, \boldsymbol{\theta}) = \_ + \_ \cdot x_1 + \_ \cdot x_2 + \_ \cdot \sin(x_2) + \_ \cdot x_1 \cdot x_2$$

   b. 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.

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

   b. 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 **0.022**, when evaluated on the test set. You will get extra points if you provide a model of your choice whose test performance is **better** (i.e., the MSE is lower) than ours. Of course, you must also tell us why your model is performing better.

2   Questions

2.1   Q1. Training versus Validation

Q1.1 What is the whole figure about?

A1.1

Q1.2 Explain the behaviours of the curves in each of the three highlighted sections in the figure, namely (a), (b), and (c).

A1.2

   Q1.2.a Can you identify any signs of overfitting or underfitting in the plot? If yes, explain which sections correspond to which concept.
   A1.2.a

   Q1.2.b How can you determine the optimal complexity of the model based on the given plot?
   A1.2.b

Q1.3 Is there any evidence of high approximation risk? Why? If yes, in which of the below subfigures?

A1.3

Q1.4 Do you think that increasing the model complexity can bring the training error to zero? And the structural risk?

A1.4

Q1.5 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)

A1.5

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:

Q2.1 $x_3 = x_1 + 0.2 \cdot x_2$.

A2.1

Q2.2 $x_3 = x_1 \cdot x_2 \cdot x_2$

A2.2

Q2.3 x_3 is a random variable independent from y.

A2.3

Q2.3 How would your answers change if you were using Lasso Regression?

A2.3

Q2.4 Explain the motivation behind Ridge and Lasso regression and their principal differences.

A2.4

2.3   Q3. Classification

Q3.1 Your boss asked you to solve the problem using a perceptron, and now he's upset because you are getting poor results. How would you justify the poor performance of your perceptron classifier to your boss?

A3.1

Q3.2 Would you expect better luck with a neural network with the activation function $h(x) = -x * e^(-2)$ for the hidden units?

A3.2

Q3.3 What are the main differences and similarities between the perceptron and the logistic regression neuron?

A3.3