

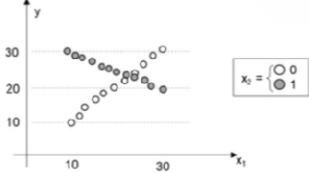
DST Exam recap

Question 1. (25 points)

First 2/3 of this question was very similar to the Exercise 6 from the Tutorial 4.

Exercise 6: Linear Regression II

The following figure shows a schematic scatter plot of data points, the y-values of which are to be explained by different linear regression models based on x_1 and x_2 .

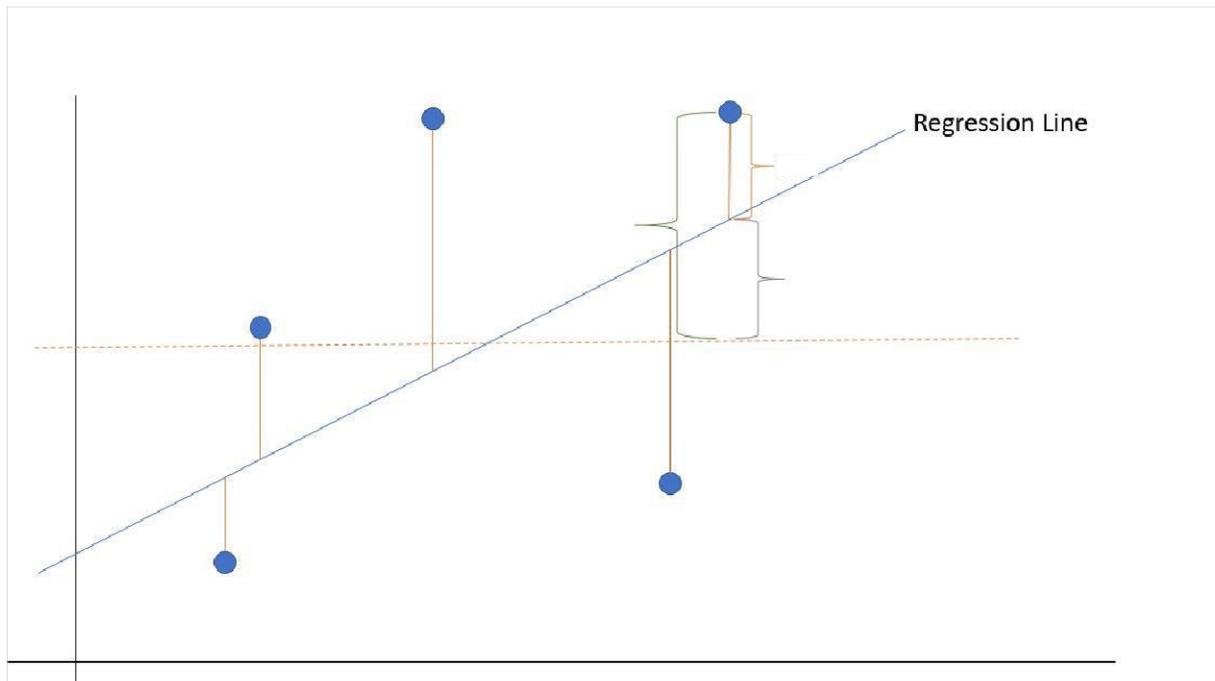


a) In the following table, please indicate the coefficients for the respective regression model estimates (grey fields do not exist in the respective model).

	Model I	Model II	Model III	Model IV	Model V
Intercept					
x_1					
x_2					
$x_1 * x_2$					

Seite 2/3

In the last 1/3 a graph was given to us. It kinda looked like the following:



We were then asked indirectly about SST, SSR and SSE. Meaning, where is which one of these on the graph, as well as their formula.

Problem 2. (18 points)

Multiple Choice. 6 Topics with 3 questions each = 18 Questions in total.

All topics, from Web-scraping to Data Visualization were covered. You can tick (ankreuzen) one, two, three or no answers at all. If you followed and understood lecture material, questions were all very manageable.

Problem 3. (22 or 25 points)

You want to conduct an experiment, in which you want to find out in how much can use of A.I. increase/decrease a skill gap between experienced coders vs. beginners. Your subject pool are university students. You can use the economics department lab of your institution. Your budget is 500€.

(hint: use within-subject design)

3.1: What is the experiment's main outcome variable?

3.2: How does your treatment design look like? & Which experimental task would you choose?

3.3: Which 4 ethical considerations would you make?

3.4: How would you incentivize participants?

Problem 4. (25 or 22 points)

This problem was almost 1 to 1 to the Exercise 1 of the Tutorial 5.

Data Science Toolbox
Tutorial 5
Machine Learning

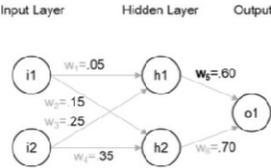


Prof. Dr. Timm Teubner
Lukas Hermanns

Exercise 1: Neural Networks

Consider the following neural network with two input nodes, two nodes in a single hidden layer, and one output node. The network's current weights are indicated on the edges.

Input Layer Hidden Layer Output Layer



For **input values of i1=1, i2=2**, compute the nodes' output values $\Phi(z)$ as well as the intermediate z values. Please use the sigmoid activation function (see below). Assume all bias units (b) to be zero.

	z (=net)	$\Phi(z)$ (=out)
h1		
h2		
o1		

Assume that the **target value was 0**. What is the total error?

Now, conduct one update step for w_5 (using backward propagation). To do so, make use of the piecewise multiplicativity between total and partial derivatives and first determine the individual derivatives and values.

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5}$$

Partial Derivative	Formula	Value
$\frac{\partial E_{total}}{\partial out_{o1}}$		
$\frac{\partial net_{o1}}{\partial w_5}$		

isis.tu-berlin.de

However, with a 1-2 exceptions.

For example, we were asked NOT to use the sigmoid activation function.

The second part of this problem was about **Data Visualization**. We were given a graph and were asked to point out two things that were odd/wrong in terms of visualization of data.

On the left part we had a bar chart with absolute numbers of serial killers by country (top 15 countries), with USA having 10x as many as the second country in the list which happened to be Russia, very closely followed by the UK.

And in the middle there was a world map with dots of different sizes placed on each country that was in the top-15. These dots corresponded with the ranking of a country, and not with an actual amount of serial killers, e.g. a dot above America was almost the same size as dots above Russia or the UK.

That sums up all that i could remember. Hope it helps :)