

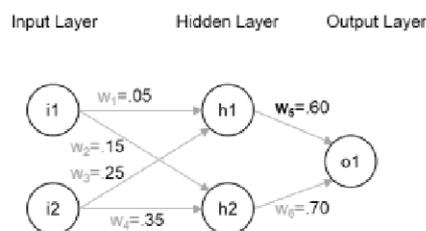
Gedächtnisprotokoll DataScience-Toolbox

General: 4 Parts with 22P, 25P, 18P and 25P = 90P in total

1. Machine Learning 25P

- This was basically exactly this from the tutorial questions with different numbers and only 2 of the hints:

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.



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
$\frac{\partial E_{total}}{\partial out_{o1}}$		
$\frac{\partial out_{o1}}{\partial nct_{o1}}$		
$\frac{\partial nct_{o1}}{\partial w_5}$		

Now, using gradient descent, conduct the update step. Assume a learning rate of 0.5. What is the new value of w_5 ?

Hints:

$z = \sum_i w_i x_i + b$	Derivate of Sigmoid function: $f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$ $\frac{d}{dx} f(x) = \frac{e^x \cdot (1 + e^x) - e^x \cdot e^x}{(1 + e^x)^2}$ $\frac{d}{dx} f(x) = \frac{e^x}{(1 + e^x)^2} = f(x)(1 - f(x))$
$f(x) = \frac{1}{1 + e^{-x}}$	
$E_{total} = \sum \frac{1}{2} (target - output)^2$	

2. Web Scraping + Experiment 22P

- 4P: How would you get all the names like “James” from that website with JSoup for example, describe your steps. (You had a simple HTML-DOM on the left side, where some divs had the id “name” and their content were the names.)
- 2P: Write pseudo code in any language on how you would get the names and print them out in the console.

- 16P: You want to find out if beginner programmers profit more from generative ai or advanced programmers. Budget was 500€. Then it was basically this from the tutorial:

- b) What are your main research questions?
- c) How does your treatment design look like?
- d) Which experimental task would you choose?
- e) How do you incentivize participants?
- f) What is the experiment's main outcome variable?
- g) How large should you sample be and how do you recruit participants?
- h) Which ethical considerations would you make?

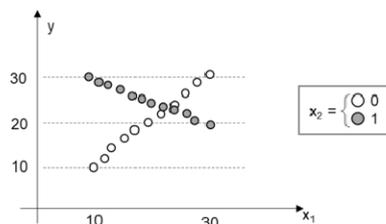
3. Multiple Choice (True or False) 18P

- 6*3 Questions
- 0-3 right answers
- 1P per right answer
- Questions are very close to the slides and the guest lecture "Data Mesh" was also relevant.
- for example: "If the p-value of the null hypothesis is greater than 5%, the null hypothesis is not rejected."

4. Linear Regression 25P

- 12P Basically exactly this exercise for the tutorial question:

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					
$x_1 * x_2$					

