



# Machine Learning (SS 24)

## Assignment 04: Linear regression

Yi Wang

[yi.wang@ki.uni-stuttgart.de](mailto:yi.wang@ki.uni-stuttgart.de)

Akram Sadat Hosseini

[Akram.Hosseini@ipvs.uni-stuttgart.de](mailto:Akram.Hosseini@ipvs.uni-stuttgart.de)

Tim Schneider

[timphillip.schneider@ipvs.uni-stuttgart.de](mailto:timphillip.schneider@ipvs.uni-stuttgart.de)

Farane Jalali/Farahani

[Akram.Hosseini@ipvs.uni-stuttgart.de](mailto:Akram.Hosseini@ipvs.uni-stuttgart.de)

Osama Mohammed

[osama.mohammed@ipvs.uni-stuttgart.de](mailto:osama.mohammed@ipvs.uni-stuttgart.de)

Daniel Frank

[daniel.frank@ipvs.uni-stuttgart.de](mailto:daniel.frank@ipvs.uni-stuttgart.de)

This assignment sheet consists of 3 pages with the following 2 tasks:

- Task 1: Simple Linear Regression (50 Points) <sup>2</sup>
- Task 2: Multiple linear regression (50 Points) <sup>3</sup>

Submit your solution in ILIAS as a single PDF file.<sup>1</sup> Make sure to list full names of all participants, matriculation number, study program and B.Sc. or M.Sc. on the first page. Optionally, you can *additionally* upload source files (e.g. PPTX files). If you have any questions, feel free to ask them in the exercise forum in ILIAS.

**Submission is open until Monday, 13.05.2024, 12:00 noon.**

---

<sup>1</sup>Your drawing software probably allows to export as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like [pdfarranger](#)) to combine the PDFs into a single file.



## Task 1: Simple Linear Regression (50 Points)

Assume you are a graduate who has just started a new job at a company. Your manager has provided historical salary data to help you understand how your earnings will increase over time. Now, you want to use your knowledge of linear regression to predict your future salary and plan your career development.

Your manager shows you the following salary data points:

- $(x_1, y_1) = (1, 50)$  - After 1 year, employees earn 50,000 €,
- $(x_2, y_2) = (2, 55)$  - After 2 year, employees earn 55,000 €,
- $(x_3, y_3) = (3, 65)$  - After 3 year, employees earn 65,000 €.

### 1. Task (16 Points)

Calculate the best fit of the linear regression function  $y = \beta_0 + \beta_1 x + \epsilon$  according to the given data (hint: use the least square approximation).

2. **Task (10 Points)** Predict your future salary after 5 and 10 years.

3. **Task (12 Points)** Compute the RSS loss of the dataset.

4. **Task (3 Points)** Express the RSS loss with  $l_1$  and  $l_2$  regularization.

5. **Task (9 Points)** Explain the concept of regularization and the effect of  $l_1$  and  $l_2$  regularization on the model parameters (just explain, don't calculate).



## Task 2: Multiple linear regression (50 Points)

Now, you have been told that the salary progress is affected by the number of years each employee has worked and their level of education. Your manager provides you with a dataset that includes these new variables:

- $x_{i,1}$ : Years of experience,
- $x_{i,2}$ : Education level (0 for Bachelor's degree, 1 for Master's and 2 for PhD)

**Table 1** Salary

Years of experience	Education level	Salary
1	1	52
4	1	63
6	0	62
8	2	91
10	1	75

1. **Task (20 Points)** The scalar representation of the multiple linear regression is  $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$ . Calculate the best fit of the parameters  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  according to the given data.
2. **Task (10 Points)** Assume you are an employee with a Master's degree. Predict your future salary after 5 years.
3. **Task (10 Points)** Compute the RSS loss of the dataset.
4. **Task (10 Points)** Express the RSS loss with  $l_2$  regularization and explain the potential benefits and drawbacks of choosing a smaller vs. larger  $\lambda$  (don't calculate).