

Exercise Sheet: Linear Regression

1 Exercises

1.1

Determine for $D = \{(1, 2), (2, 2), (2, 3), (3, 3), (4, 3)\} \subset \mathbb{R}_x \times \mathbb{R}_y$ both regression lines, as well as the regression coefficients r_{xy} and r_{yx} , by hand (calculator or CAS is allowed, but, e.g., no Python package!). Sketch the data points and the regression lines.

1.2

In this exercise, ϕ denotes the objective function from Theorem 2.1 and ψ the objective function such that

$$\psi(f) := \sum_{i=1}^n |f(x_i) - y_i|$$

(i) Find a dataset D such that

$$\arg \min\{\phi(f) \mid f : \mathbb{R} \rightarrow \mathbb{R} \text{ affine linear}\} \neq \arg \min\{\psi(f) \mid f : \mathbb{R} \rightarrow \mathbb{R} \text{ affine linear}\}$$

holds.

(ii) Find a dataset where not all x -values are the same and yet there is more than one function f^* with

$$\psi(f^*) = \min\{\psi(f) \mid f : \mathbb{R} \rightarrow \mathbb{R} \text{ affine linear}\}.$$

1.3

Carry out the method of linear regression for the following dataset. Below, the exam preparation time is the x -variable and the exam result is the y -variable. Draw all data points and also the regression line.

Student	Exam preparation in h	Exam result in %
1	21.0	82
2	18.0	69
3	15.0	29
4	8.0	41
5	8.0	44
6	1.5	8
7	0.0	10

1.4

Let $D = \{(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \dots, n\}$ be a dataset, where neither all x_i nor all y_i are equal. Show that the first and second regression lines coincide if and only if $r_{xy} = \pm 1$.