

DATA SCIENCE 2

VORLESUNG 2 - INTRO

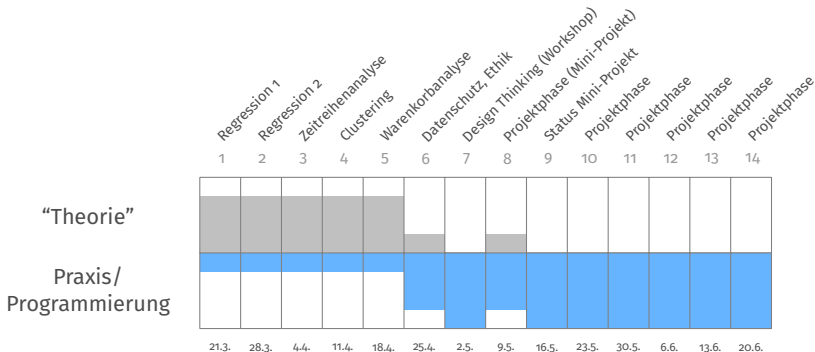
PROF. DR. CHRISTIAN BOCKERMANN

HOCHSCHULE BOCHUM

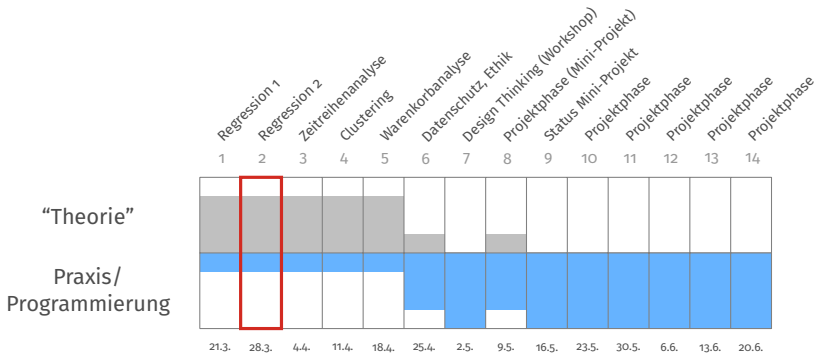
SOMMERSEMESTER 2023

1 Wiederholung - lineare Regression

Themen der Vorlesung



Themen der Vorlesung



Wiederholung - lineare Regression

Regression liefert reellwertige Vorhersagen

- Für Regression gilt $\mathcal{Y} = \mathbb{R}$
- Menge $\mathbf{X} \times \mathbf{y}$, d.h. jedem Beispiel x_i ist ein $y_i \in \mathbb{R}$ zugeordnet
- Qualitätsfunktion $q : (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \rightarrow \mathcal{Y}) \rightarrow \mathbb{R}$

Regression liefert reellwertige Vorhersagen

- Für Regression gilt $\mathcal{Y} = \mathbb{R}$
- Menge $\mathbf{X} \times \mathbf{y}$, d.h. jedem Beispiel x_i ist ein $y_i \in \mathbb{R}$ zugeordnet
- Qualitätsfunktion $q : (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \rightarrow \mathcal{Y}) \rightarrow \mathbb{R}$

Ziel:

- Finde Modell

$$f : \mathcal{X} \rightarrow \mathcal{Y},$$

das die Qualitätsfunktion optimiert.

Auch hier wieder: Lernen als **Optimierungsproblem!**

Lineare Regression

Einfache Modell-Klasse linearer Funktionen

$$\begin{aligned}\hat{f}(\mathbf{x}) &= b + \sum_{i=1}^p w_i x_i \\ &= \sum_{i=0}^p w_i x_i \quad \text{mit } w_0 = b, x_0 = 1 \\ &= \mathbf{w}^T \mathbf{x} = \hat{f}_{\mathbf{w}}(\mathbf{x})\end{aligned}$$

Lineare Regression

Einfache Modell-Klasse linearer Funktionen

$$\begin{aligned}\hat{f}(\mathbf{x}) &= b + \sum_{i=1}^p w_i x_i \\ &= \sum_{i=0}^p w_i x_i \quad \text{mit } w_0 = b, x_0 = 1 \\ &= \mathbf{w}^T \mathbf{x} = \hat{f}_{\mathbf{w}}(\mathbf{x})\end{aligned}$$

Parameter $\mathbf{w} = (w_0, w_1, \dots, w_p)$ bestimmen das Modell

Lineare Regression

Einfache Modell-Klasse linearer Funktionen

$$\begin{aligned}\hat{f}(\mathbf{x}) &= b + \sum_{i=1}^p w_i x_i \\ &= \sum_{i=0}^p w_i x_i \quad \text{mit } w_0 = b, x_0 = 1 \\ &= \mathbf{w}^T \mathbf{x} = \hat{f}_{\mathbf{w}}(\mathbf{x})\end{aligned}$$

Parameter $\mathbf{w} = (w_0, w_1, \dots, w_p)$ bestimmen das Modell

Idee: Bestimme \mathbf{w} so, dass der Trainingsfehler minimal wird!