

DATA SCIENCE 2

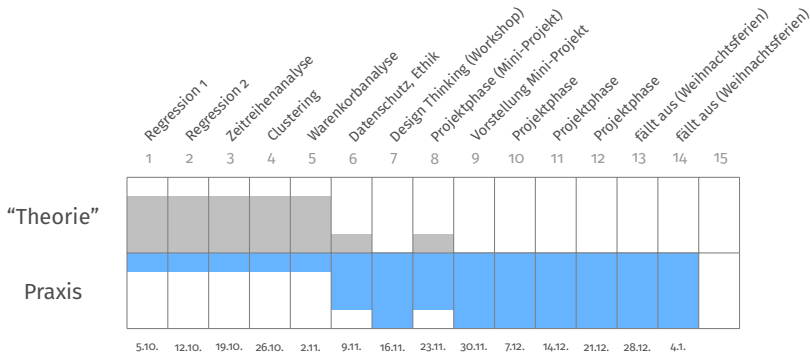
VORLESUNG 4 - INTRO

PROF. DR. CHRISTIAN BOCKERMANN

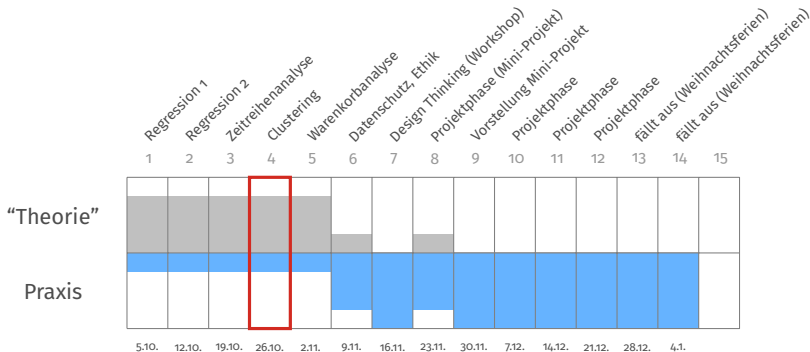
HOCHSCHULE BOCHUM

WINTERSEMESTER 2023 / 2024

Themen der Vorlesung



Themen der Vorlesung



Bisher: Überwachtes Lernen mit Zielvariable

X_{Hubraum}	X_{Leistung}	X_{Zylinder}	X_{Gewicht}	$X_{\text{Beschl.}}$	$Y_{l/100km}$
307.0	130	8	3504	12.0	15.7
350.0	165	8	3693	11.5	18.8
318.0	150	8	3436	11.0	15.7
304.0	140	8	3433	12.0	17.6

Bisher: Überwachtes Lernen mit Zielvariable

$X_{Hubraum}$	$X_{Leistung}$	$X_{Zylinder}$	$X_{Gewicht}$	$X_{Beschl.}$	$Y_{l/100km}$
307.0	130	8	3504	12.0	15.7
350.0	165	8	3693	11.5	18.8
318.0	150	8	3436	11.0	15.7
304.0	140	8	3433	12.0	17.6

Bisher: Überwachtes Lernen mit Zielvariable

X_{Hubraum}	X_{Leistung}	X_{Zylinder}	X_{Gewicht}	$X_{\text{Beschl.}}$	$Y_{l/100km}$
307.0	130	8	3504	12.0	15.7
350.0	165	8	3693	11.5	18.8
318.0	150	8	3436	11.0	15.7
304.0	140	8			17.6

Vorhersage

Bisher: Überwachtes Lernen mit Zielvariable

X_{Hubraum}	X_{Leistung}	X_{Zylinder}	X_{Gewicht}	$X_{\text{Beschl.}}$	$Y_{l/100km}$
307.0	130	8	3504	12.0	15.7
350.0	165	8	3693	11.5	18.8
318.0	150	8	3436	11.0	15.7
304.0	140	8			17.6

Vorhersage

Bisher: Überwachtes Lernen mit Zielvariable

X_{Hubraum}	X_{Leistung}	X_{Zylinder}	X_{Gewicht}	$X_{\text{Beschl.}}$	$Y_{l/100km}$
307.0	130	8	3504	12.0	15.7
350.0	165	8	3693	11.5	8.1
318.0	150	8	3436	11.0	11.7
304.0	140	8	3433	12.0	7.1

Unüberwachtes Lernen: Es gibt keine Vorhersage-Variable!

Clustering sucht Aufteilung von Daten in ähnliche Gruppen

- Datenmenge \mathbf{X} von Beispielen (keine Klassen gegeben!)
- Parameter k zu findender Gruppen
- Abstandsmaß $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$
- Qualitätsfunktion q

Ziel:

- Abstand *innerhalb* der Gruppen soll minimiert, Abstand *zwischen* den Gruppen soll maximiert werden