Data Science 1

Vorlesung 7 - Intro

Prof. Dr. Christian Bockermann

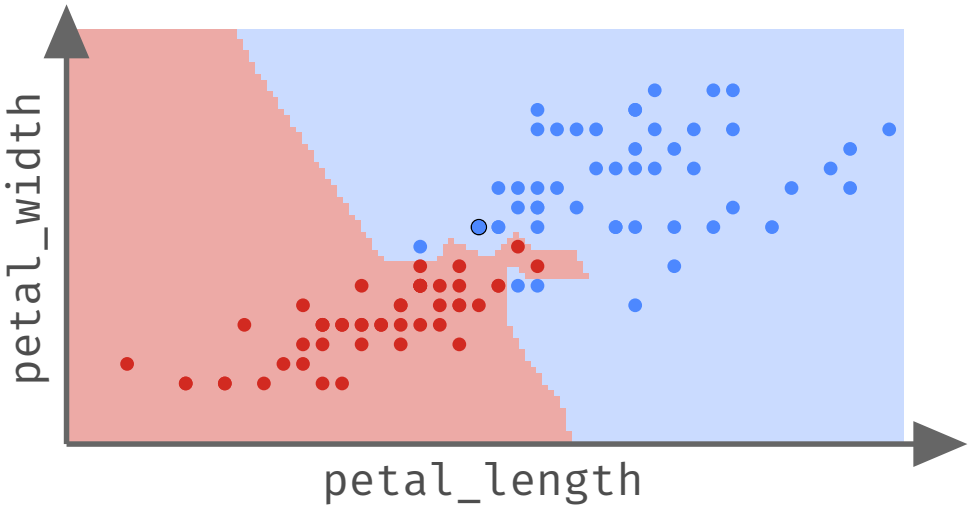
Hochschule Bochum

Sommersemester 2023

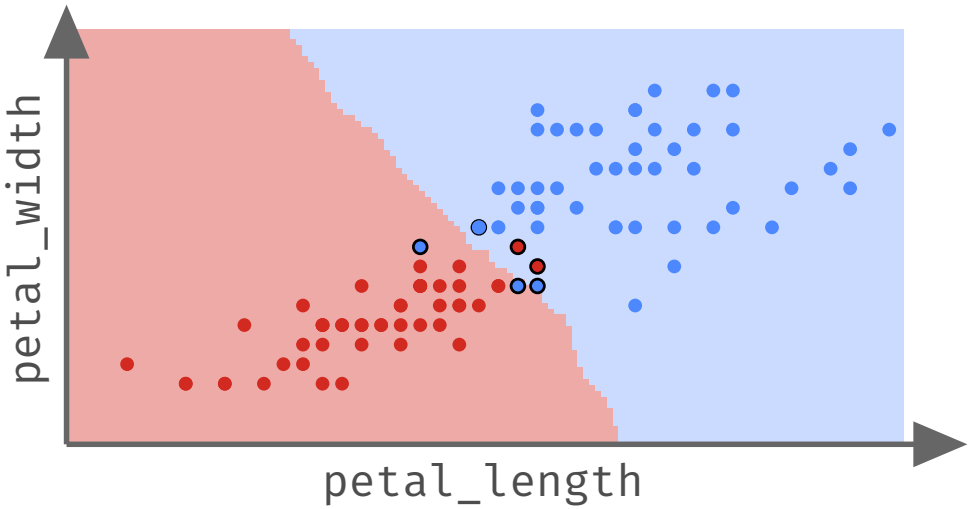
Was geschah zuletzt?

- Instanzbasiertes Lernen über Ähnlichkeit
- Distanz-Funktion auf Beispielen (eukl. Distanz)
- Normalisierung von Daten (Min/Max-, z-Normalisierung)
- k-NN als Vorhersagemodell

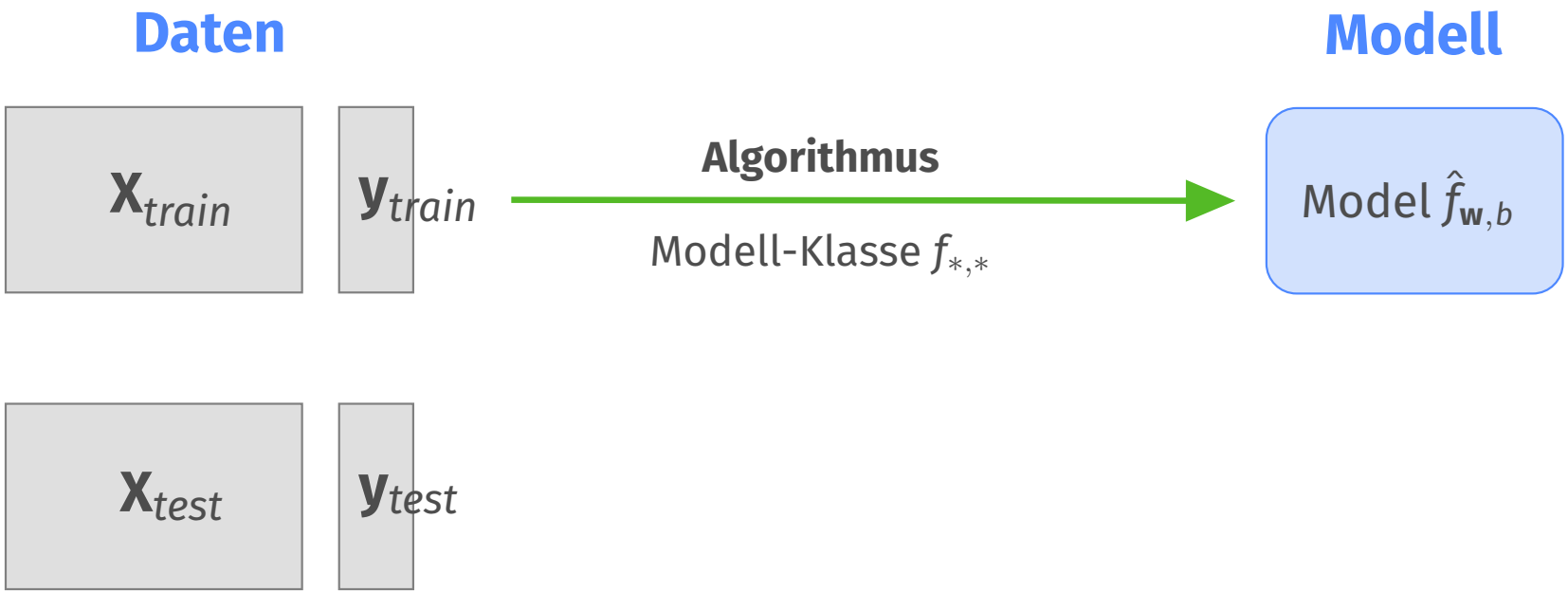
k=1



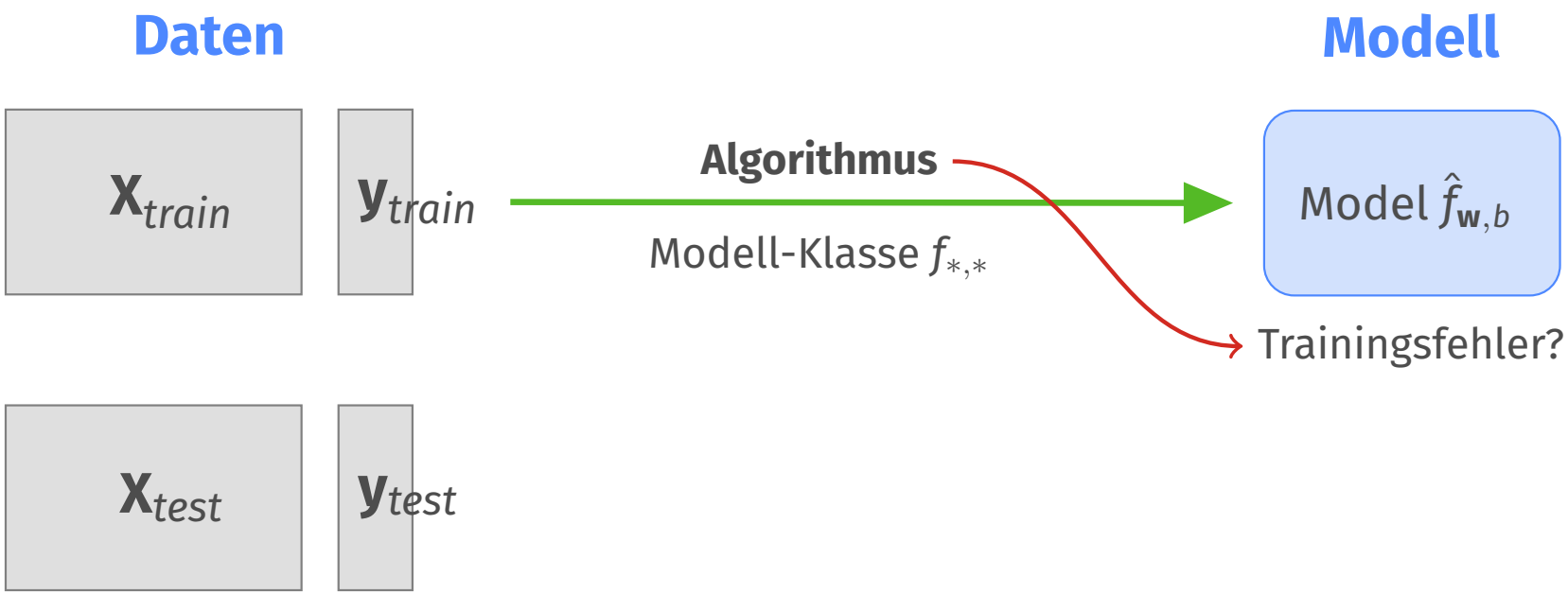
k=16



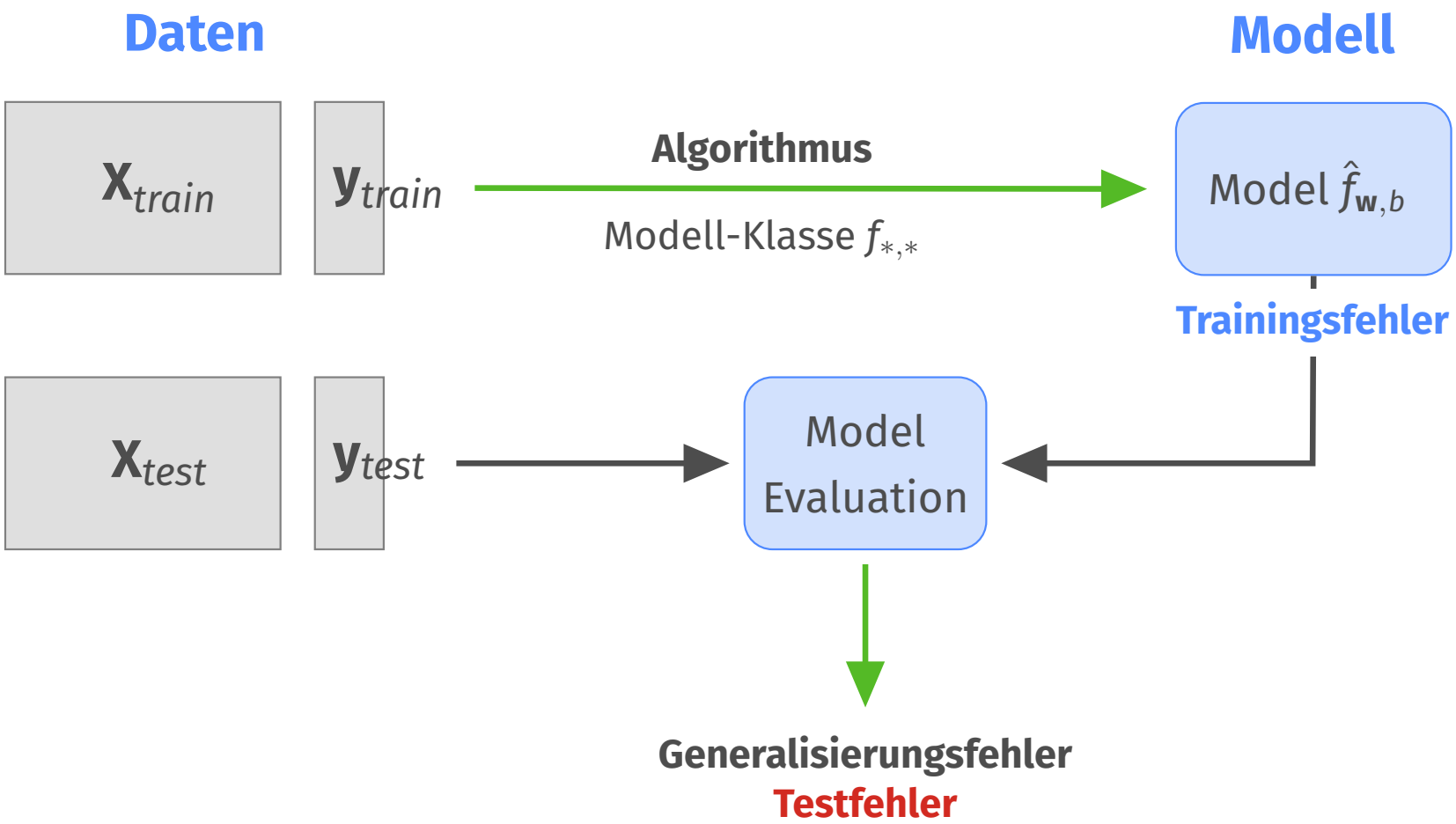
Bewertung der Modell-Güte – Generalisierungsfehler



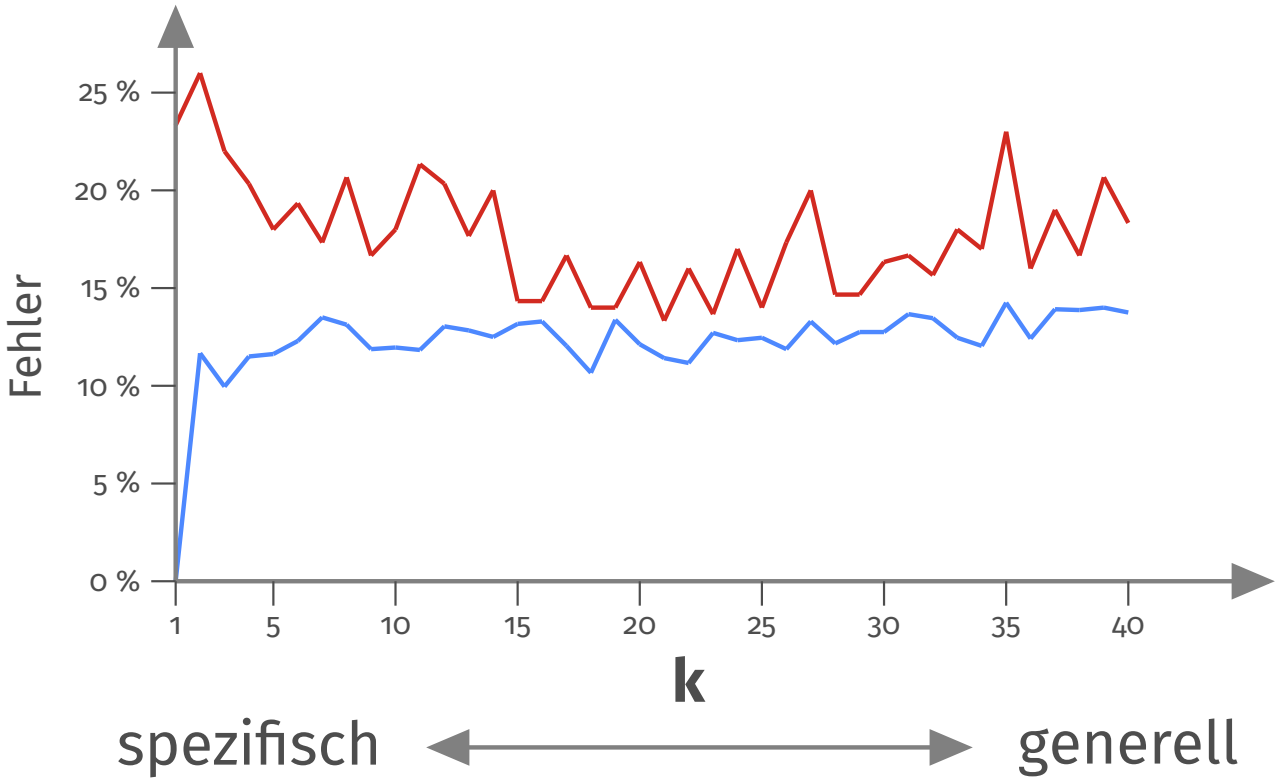
Bewertung der Modell-Güte – Generalisierungsfehler



Bewertung der Modell-Güte – Generalisierungsfehler



Trainings- und Test-Fehler auf generiertem Datensatz (k-NN)



Overfitting

“Das Modell passt nur zu den Trainingsdaten.”

Overfitting

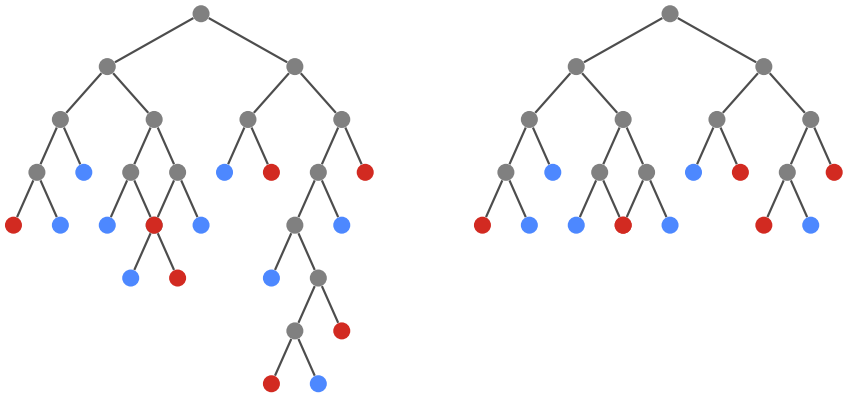
"Das Modell passt nur zu den Trainingsdaten."

	Trainingsfehler **klein**	Trainingsfehler **groß**
Testfehler **klein**	Das sieht gut aus!	
Testfehler **groß**	**Overfitting!**	Das Modell lernt nicht!?

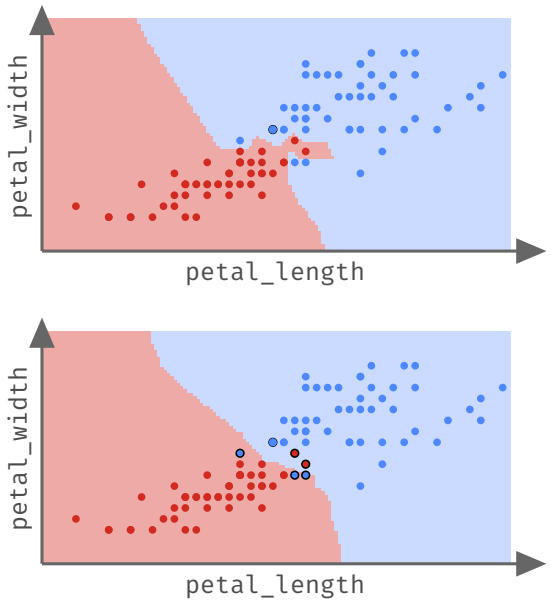
Overfitting - zu spezifisches Modell

- Modell zu sehr an die Trainingsdaten angepasst
- Vorhersage auf unbekannten Daten schlechter
- Modellkomplexität begrenzen (generelleres Modell)

Tiefe bei Bäumen beschränken



k bei k-NN erhöhen



Wo sind wir heute (Vorlesung 7) ?