

# Instanzbasiertes Lernen

Maschinelles Lernen auf die faule Art

## Das Eiscreme-Spiel: Gruppe 2

Prognose für  
Samstag, 18 Grad,  
stark bewölkt?

138!

Freitag, 30.3.2018	bewölkt	15	67
Dienstag, 10.4.2018	stark bewölkt	20	84
Samstag, 21.4.2018	schwach bewölkt	27	270
Mittwoch, 2.5.2018 ✓	stark bewölkt ✓	17 ✓	45
Sonntag, 13.5.2018 ✓	stark bewölkt ✓	17 ✓	138
Donnerstag, 24.5.2018	bewölkt	22	100
Montag, 4.6.2018	heiter bis wolkig	30	247
Freitag, 15.6.2018	mäßig bewölkt	23	177
Dienstag, 26.6.2018	bewölkt	22	159
Samstag, 7.7.2018	mäßig bewölkt	27	292
Mittwoch, 18.7.2018	schwach bewölkt	28	217

historische Verkäufe  
an einem  
sehr ähnlichen Tag

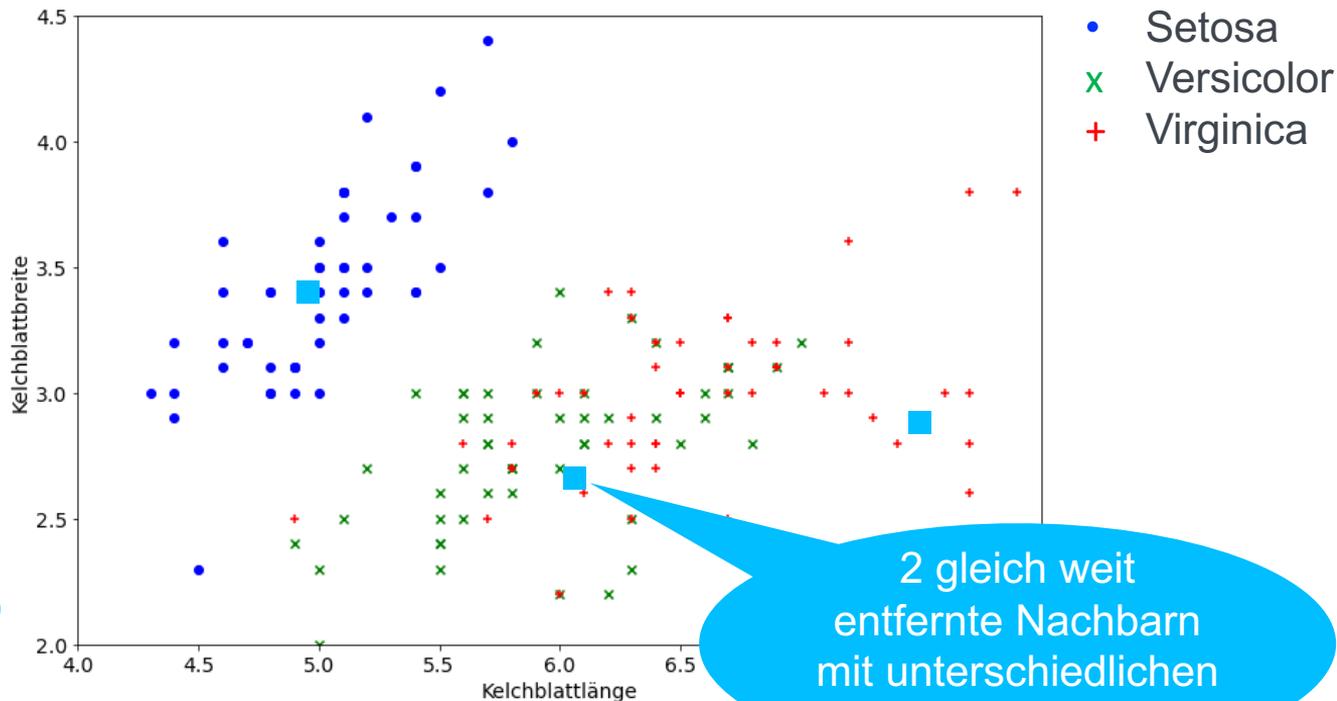
## Idee instanzbasiertes Lernen

- Instanzbasiert: Nutze für Vorhersage den Wert der **ähnlichsten** Instanz (des ähnlichsten Datenpunkts) der Trainingsdaten
- Hier: Regression
- Ähnlichkeit aus Tabelle “im Kopf“ bestimmt

## Noch ein Beispiel, grafisch dargestellt

- Nahe beieinander liegende Datenpunkte sind ähnlich
- Nahe beieinander liegende Datenpunkte gehören meistens zur selben Klasse

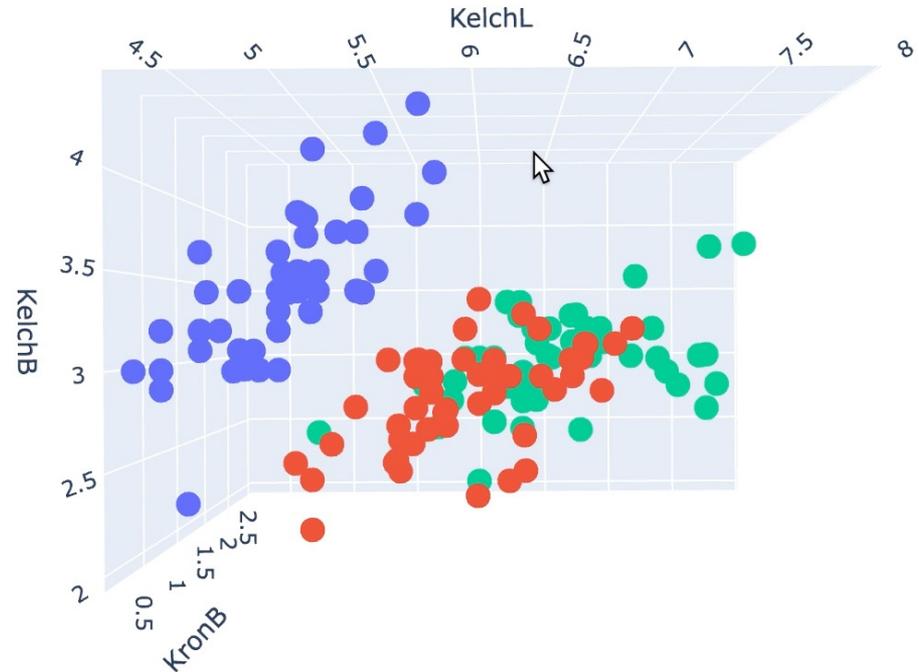
Idee: zur Klassifizierung nutzen!



2 gleich weit entfernte Nachbarn mit unterschiedlichen Klassen

## Ähnlichkeit im dreidimensionalen Raum

- Hier hilft möglicherweise die dritte Dimension Kronblattbreite
- Wenn alle drei Dimensionen berücksichtigt werden, ist die nächstgelegene Instanz hoffentlich von der korrekten Klasse



## Idee instanzbasiertes Lernen

- Instanzbasiert: Nutze für Vorhersage den Wert der **nächsten** Instanz (des nächsten Datenpunkts) der Trainingsdaten
- Hier: Klassifikation
- Nähe aus Scatterplot abgelesen
- In beiden Fällen: Distanz „berechnet“, nächsten Datenpunkt gesucht und die Vorhersage dort „abgeschaut“

Prognose für  
Samstag, 18 Grad,  
stark bewölkt?

## Das Eiscreme-Spiel: Gruppe 2

Freitag, 30.3.2018	0	bewölkt		15		67
Dienstag, 10.4.2018	0	stark bewölkt		20		84
Samstag, 21.4.2018	1	schwach bewölkt	3	27		270
Mittwoch, 2.5.2018 ✓	0	stark bewölkt ✓	8	17 ✓		45
Sonntag, 13.5.2018 ✓	1	stark bewölkt ✓	8	17 ✓		138
Donnerstag, 24.5.2018	0	bewölkt		22		100
Montag, 4.6.2018	$1-1=0$	heiter bis wolkig	$8-8=0$	30	$18-17=1$	247
Freitag, 15.6.2018	0	mäßig bewölkt	5	23		177
Dienstag, 26.6.2018	0	bewölkt		22		159
Samstag, 7.7.2018	1	mäßig bewölkt		27		292
Mittwoch, 18.7.2018	0	schwach bewölkt	3	28		217

## Idee instanzbasiertes Lernen

- In beiden Fällen: Distanz berechnet, nächsten Datenpunkt gesucht und die Vorhersage dort „abgeschaut“

### Lazy Learning – Faules Lernen

Die eigentliche Arbeit wird erst „erledigt“, wenn man die Daten für die Vorhersage bekommt (erst dann kann man den nächsten Nachbarn bestimmen).

Keine Abstraktion, keine Berechnung eines Modells anhand von Trainingsdaten.

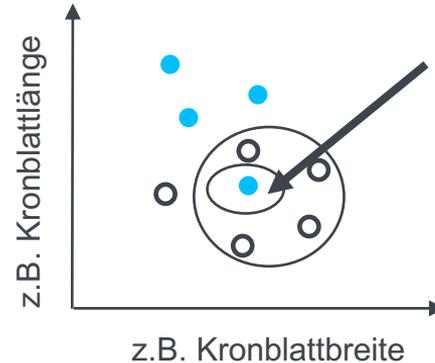
**Beim instanzbasierten Lernen werden Vorhersagen aufgrund von Ähnlichkeiten zu Instanzen in bereits vorhandenen Daten (den „Trainingsdaten“) getroffen.**

**Allerdings findet eigentlich kein „Training“ statt, es wird kein abstraktes Modell trainiert.**

# Nächste Nachbarn-Klassifikation und Regression

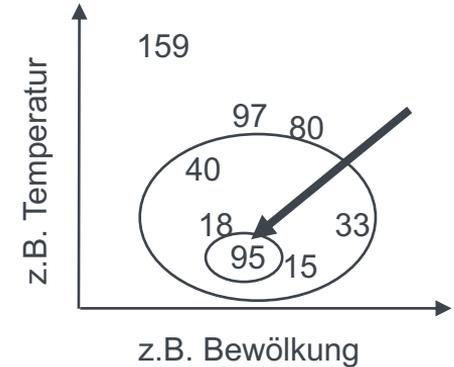
- Beispiele bisher: den einen ähnlichsten Nachbarn bestimmen
- Ähnlichkeit bestimmt durch Distanzfunktion  
⇒ nächster Nachbar

## Klassifikation



- Ein nächster Nachbar
- Oder k nächste Nachbarn und Mehrheitsentscheid

## Regression



- Ein nächster Nachbar
- Oder k nächste Nachbarn und Mittelwert

**Nächste-Nachbarn-Klassifikation (auch: k-Nächste-Nachbarn-Klassifikation) ist ein Beispiel für instanzbasiertes Lernen.**

**Für neue Datenpunkte werden in den Trainingsdaten die k nächsten Nachbarn gesucht. Als Vorhersage wird die bei diesen Nachbarn am häufigsten vorkommende Klasse verwendet.**

**Die Nähe wird mithilfe einer Distanzfunktion bestimmt.**

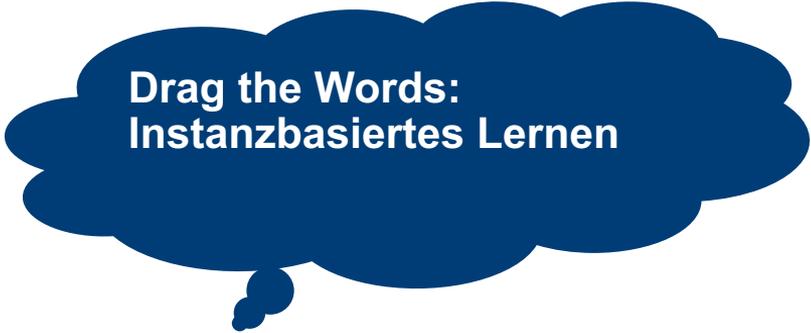
**Nächste-Nachbarn-Regression (auch: k-Nächste-Nachbarn-Regression) ist ein Beispiel für instanzbasiertes Lernen.**

**Für neue Datenpunkte werden in den Trainingsdaten die k nächsten Nachbarn gesucht. Als Vorhersage wird der Mittelwert der bei den Nachbarn vorkommenden Werte verwendet.**

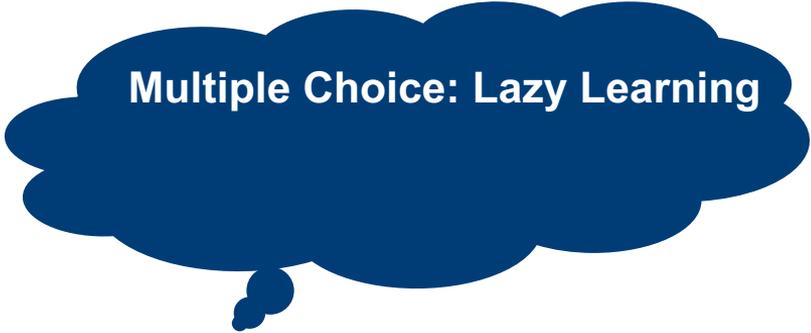
**Die Nähe wird mithilfe einer Distanzfunktion bestimmt.**



**Fill in the Blanks: Vorhersage  
von Verkäufen**



**Drag the Words:  
Instanzbasiertes Lernen**



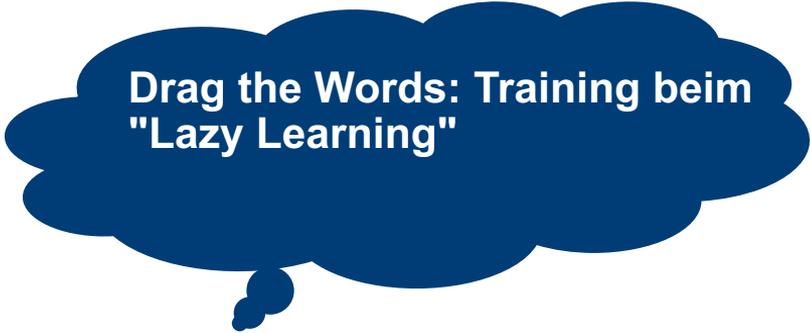
**Multiple Choice: Lazy Learning**



**Drag the Words:  
Zusammenfassung**



**Single Choice: Nächste-  
Nachbarn-Klassifikation**



**Drag the Words: Training beim  
"Lazy Learning"**

## Dr. Antje Schweitzer

Universität Stuttgart  
Institut für Maschinelle Sprachverarbeitung



### Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung  
Institut für Software Engineering



Reutlingen | Tübingen | Zollernalb



# Lizenzbestimmungen

“Instanzbasiertes Lernen“ von Antje Schweitzer, KI B<sup>3</sup> / Uni Stuttgart

Das Werk - mit Ausnahme der folgenden Elemente:

- Logos der Verbundpartner und des Förderprogramms
- im Quellenverzeichnis aufgeführte Medien

ist lizenziert unter:

 [CC BY 4.0 \(https://creativecommons.org/licenses/by/4.0/deed.de\)](https://creativecommons.org/licenses/by/4.0/deed.de)

(Namensnennung 4.0 International)

## Quellenverzeichnis

Titelfoto: [Manja Vitolic \(https://unsplash.com/@madhatterzone\)](https://unsplash.com/@madhatterzone), “lazy cat afternoon“, auf [Unsplash \(https://unsplash.com/photos/j6UmKxKDRPw\)](https://unsplash.com/photos/j6UmKxKDRPw), ist lizenziert unter [Unsplash-Lizenz \(https://unsplash.com/license\)](https://unsplash.com/license).

Bildausschnitt verändert.