

Übung 7

Übung zu Informationstechnik II und Automatisierungstechnik – Felix Pistorius

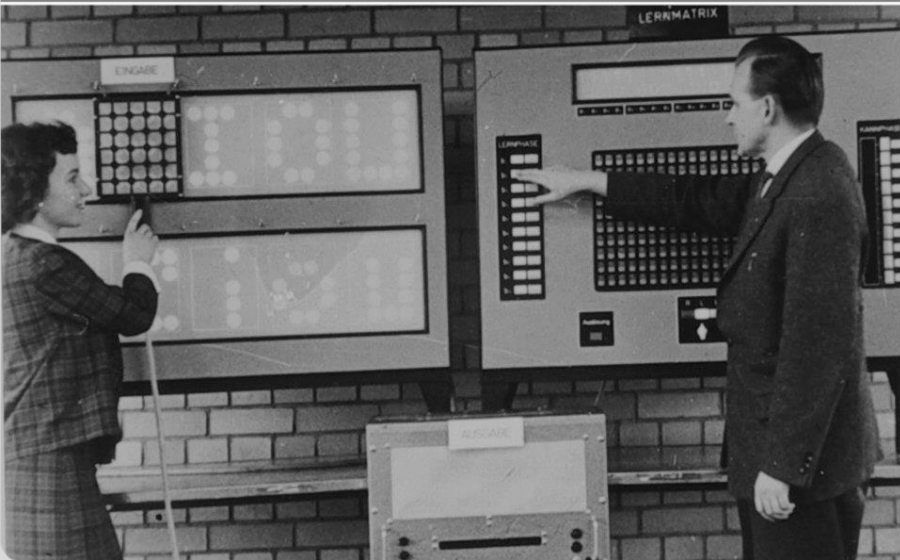
Prof. Dr.-Ing. Eric Sax

Institutsleitung

Prof. Dr.-Ing. J. Becker

Prof. Dr.-Ing. E. Sax

Prof. Dr. rer. nat. W. Stork



Allgemeines

- Mindestens eine Antwort pro Frage korrekt
- Deckblatt des MC-Teils muss ausgefüllt werden
 - Name und Matrikelnummer
 - Siehe Folgefolie
 - Dafür gesondert Zeit
- Keine negativen Punkte
- Ankreuzprozedere siehe Folgefolien

Name und Matrikelnummer

Prüfungsbogen: 1, Seite: 1/5

EvaExam Informationstechnik II und Automatisierungstechnik

Karlsruher Institut für Technologie Prof. Dr.-Ing. Eric Sax

Institut für Technik der Informationsverarbeitung ITIV

Bitte so markieren: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.

Korrektur: ☐ ☒ ☐ ☐ Bitte beachten Sie die Folie 1.

Bitte ausfüllen (Die Angabe des Namens ist freiwillig.):

Vorname: _____

Nachname: _____

Prüfungsteilnehmer-ID für den Prüfungsbogen Nr.: 1:

1	☐	☐	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐
0	☐	☐	☐	☐	☐	☐	☐

Für die eindeutige Zuordnung der Prüfung übertragen Sie bitte Ihre Prüfungsteilnehmer-ID gewissenhaft in die dafür vorgesehenen Felder. Alle Seiten sind vollständig individualisiert und nicht mit anderen Prüfungen tauschbar.

Ausfüllen &
Ankreuzen

Bitte ausfüllen (Die Angabe des Namens ist freiwillig.):

Vorname: _____

Nachname: _____

Prüfungsteilnehmer-ID für den Prüfungsbogen Nr.: 1:

	☐	☐	☐	☐	☐	☐	☐
1	☐	☐	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐
0	☐	☐	☐	☐	☐	☐	☐

Für die eindeutige Zuordnung der Prüfung übertragen Sie bitte Ihre Prüfungsteilnehmer-ID gewissenhaft in die dafür vorgesehenen Felder. Alle Seiten sind vollständig individualisiert und nicht mit anderen Prüfungen tauschbar.

Bitte ausfüllen (Die Angabe des Namens ist freiwillig.):

Vorname: Erika

Nachname: Mustermann

Prüfungsteilnehmer-ID für den Prüfungsbogen Nr.: 1:

	0	8	3	6	6	1	2
1	☐	☐	☐	☐	☐	☒	☐
2	☐	☐	☐	☐	☐	☐	☒
3	☐	☐	☒	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐	☐
6	☐	☐	☒	☒	☐	☐	☐
7	☐	☐	☐	☐	☐	☐	☐
8	☒	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐
0	☒	☐	☐	☐	☐	☐	☐

Für die eindeutige Zuordnung der Prüfung übertragen Sie bitte Ihre Prüfungsteilnehmer-ID gewissenhaft in die dafür vorgesehenen Felder. Alle Seiten sind vollständig individualisiert und nicht mit anderen Prüfungen tauschbar.

Antworten ankreuzen

EvaExam Informationstechnik II und Automatisierungstechnik Prüfungsbogen: 1, Seite: 1/5
Karlsruher Institut für Technologie Prof. Dr.-Ing. Eric Sax
Institut für Technik der Informationsverarbeitung ITIV

Bitte so markieren: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
Korrektur: ☐ ☒ ☐ ☐ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

Bitte ausfüllen (Die Angabe des Namens ist freiwillig.):

Vorname: _____

Nachname: _____

Prüfungsteilnehmer-ID für den Prüfungsbogen Nr.: 1:

1	☐	☐	☐	☐	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	☐	☐	☐	☐	☐	☐	☐	☐	☐

Für die eindeutige Zuordnung der Prüfung übertragen Sie bitte Ihre Prüfungsteilnehmer-ID gewissenhaft in die dafür vorgesehenen Felder. Alle Seiten sind vollständig individualisiert und nicht mit anderen Prüfungen tauschbar.

Allgemeine Hinweise

Die Punktevergabe erfolgt separat pro Teilaufgabe.

Manuelle Korrektur:

Tragen Sie hier nur Aufgaben ein, bei denen Sie ein geschwärztes Kreuz gewertet haben möchten (deutlich in der Aufgabe kennzeichnen).

Aufgabennummer

.....

.....

.....

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☐ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☐ Kann man ankreuzen
- ☐ Muss leer bleiben

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☐ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☐ Kann man ankreuzen
- ☐ Muss leer bleiben

Antwort
ankreuzen

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☒ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

Antworten korrigieren

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☒ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

Antwort
korrigieren

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☐ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

Nochmals
korrigieren

Manuelle Korrektur:

Tragen Sie hier nur Aufgaben ein, bei denen Sie ein geschwärztes Kreuz gewertet haben möchten (deutlich in der Aufgabe kennzeichnen).

Aufgabennummer

A1

&

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☒ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

Antworten korrigieren

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☒ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

Manuelle Korrektur:

Tragen Sie hier nur Aufgaben ein, bei denen Sie ein geschwärztes Kreuz gewertet haben möchten (deutlich in der Aufgabe kennzeichnen).

Aufgabennummer

A1

Antwort
korrigieren

Auf erster
Seite
vermerken

&

Aufgabe 1: Beispielaufgabe

Welche Aussagen treffen auf ein Kästchen zu?

- ☐ Hat vier Ecken
- ☐ Hat fünf Seiten
- ☒ Kann man ankreuzen
- ☐ Muss leer bleiben

EvaExam Informationstechnik II und Automatisierungstechnik Prüfungsbogen: 1, Seite: 1/5

Karlsruher Institut für Technologie Prof. Dr.-Ing. Eric Sax

Institut für Technik der Informationsverarbeitung Elektronik Paper

Bitte so markieren: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
Korrektur: ☐ ☒ ☐ ☐ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

Bitte ausfüllen (Die Angabe des Namens ist freiwillig.): Prüfungsteilnehmer-ID für den Prüfungsbogen Nr.: 1:

Vorname: 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Nachname: 2 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

3 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

4 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

5 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

6 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

7 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

8 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

9 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

0 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Für die eindeutige Zuordnung der Prüfung übertragen Sie bitte Ihre Prüfungsteilnehmer-ID gewissenhaft in die dafür vorgesehenen Felder. Alle Seiten sind vollständig individualisiert und nicht mit anderen Prüfungen tauschbar.

Allgemeine Hinweise

Die Punktevergabe erfolgt separat pro Teilaufgabe.

Manuelle Korrektur:

Tragen Sie hier nur Aufgaben ein, bei denen Sie ein geschwärztes Kreuz gewertet haben möchten (deutlich in der Aufgabe kennzeichnen).

Aufgabennummer

.....

.....

.....

.....

SoSe 2020

Übung zu Informationstechnik II und Automatisierungstechnik

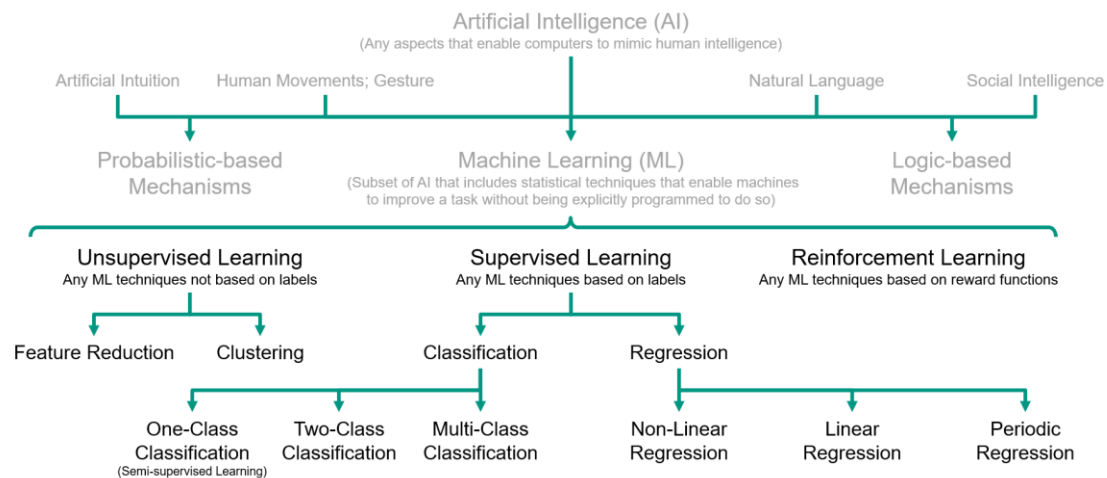
Institut für Technik der Informationsverarbeitung (ITIV)

WIEDERHOLUNG ÜBUNG 6



Supervised Learning

Modelbildung – Übersicht über Verfahren



Regression und Klassifikation

Gemeinsamkeiten und Unterschiede



Überwachtes Lernen

- Bei Trainingsdaten ist das Vorhersageattribut bekannt
- Zielgröße neuer Datensätze werden auf Basis des gelernten Modells vorhergesagt

Regressionsprobleme

- Idee
 - Bestimmung eines unbekannten **numerischen** Attributwertes (ordinal oder kategorisch durch Schwellwertsetzung)
 - Unter Benutzung beliebiger Attributwerte
- Beispiele:
 - Vorhersage von Kosten, Aufwand, etc.
 - Vorhersage von Kundenverhalten (Kündigungszeitpunkt)
 - Vorhersage zu Verkaufszahlen
 - uvm

Klassifikationsprobleme

- Idee
 - Bestimmung eines unbekannten **kategorischen** Attributwertes (ordinal mit Einschränkungen)
 - Unter Benutzung beliebiger Attributwerte
- Beispiele:
 - Klassifikation von Spam
 - Vorhersage von Kundenverhalten (Kündigung)
 - Vorhersage von Kreditwürdigkeit
 - uvm

Wiederholung Übung 6

Maschinelle Lernmethoden

Regression

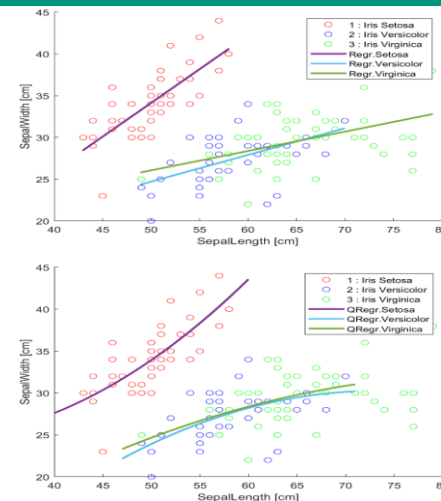
Bestimmung eines unbekannten numerischen Attributwertes

Regressionsgerade

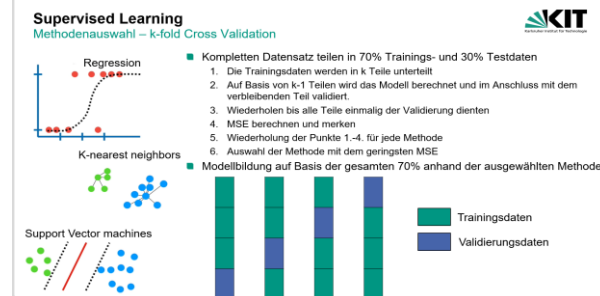
$$y = \frac{S_{xy}}{S_{xx}}x + \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x}$$

Regressionskurve

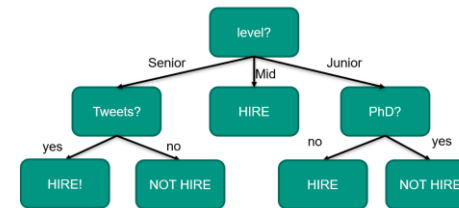
$$y = ax^2 + bx + c$$



K-fold Cross Validation Methodenauswahl



Klassifikation – Decision Tree



Confusion Matrix

		Actual value	
		Yes	No
Predicted value	Positive	True positive (TP) „Vorhersage richtig, Objekt vorhanden“ 	False positive (FP) „Vorhersage falsch, Objekt nicht vorhanden“ 
	Negative	False negative (FN) „Vorhersage falsch, Objekt vorhanden“ 	True negative (TN) „Vorhersage richtig, Objekt nicht vorhanden“ 

Gütekriterium:

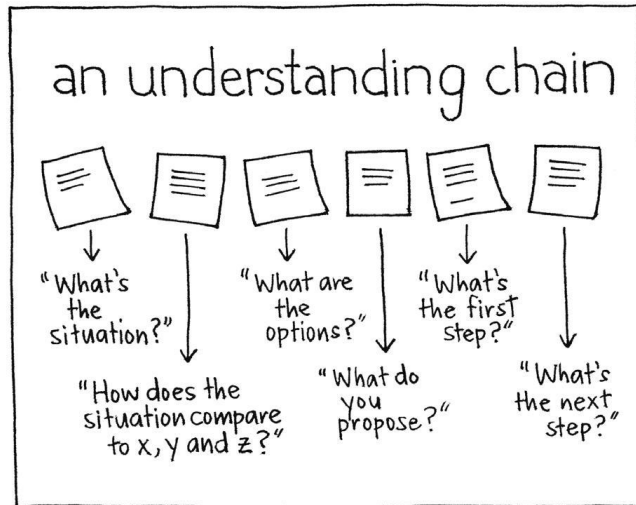
Klassifikationsrate: $P(\text{richtig klassifiziert}) = \frac{t_p + t_n}{n_{ges}}$

(accuracy) $P(\text{falsch klassifiziert}) = \frac{f_p + f_n}{n_{ges}}$

INHALT ÜBUNG 7



Data Understanding



Data Preparation



Modeling



Ziele der heutigen Übung



■ Nach der heutigen Übung können Sie....

- ...Ansätze der unüberwachten maschineller Lernverfahren auf ihre Anwendbarkeit und Wirksamkeit einschätzen

1

- ... Unterschiede zwischen unüberwachten und überwachten Lernverfahren benennen

2

- ... die Familien der Clusterverfahren mit Vor- und Nachteilen benennen

3

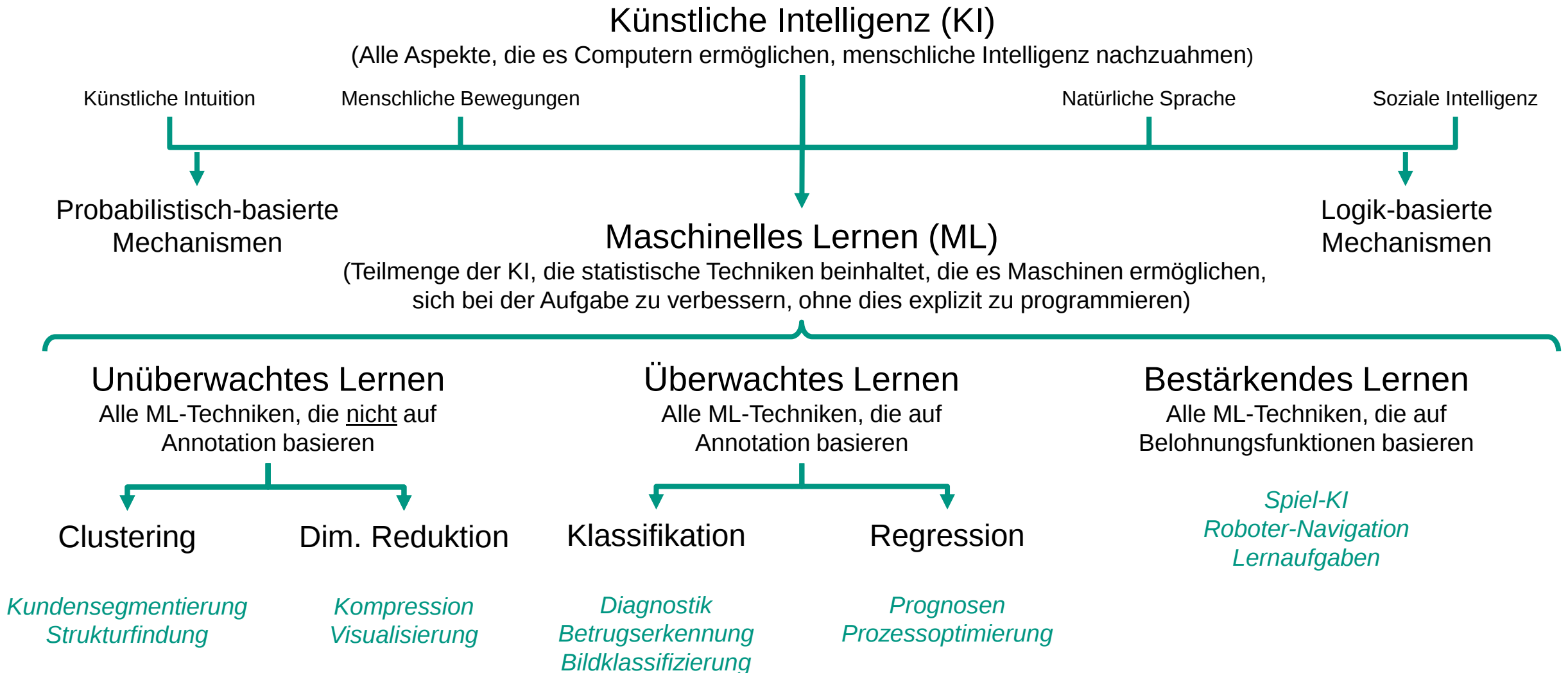
- ... Modelle unter Anwendung unüberwachter maschineller Lernverfahren aufstellen

UNSUPERVISED LEARNING CLUSTERING



Künstliche Intelligenz und Maschinelles Lernen

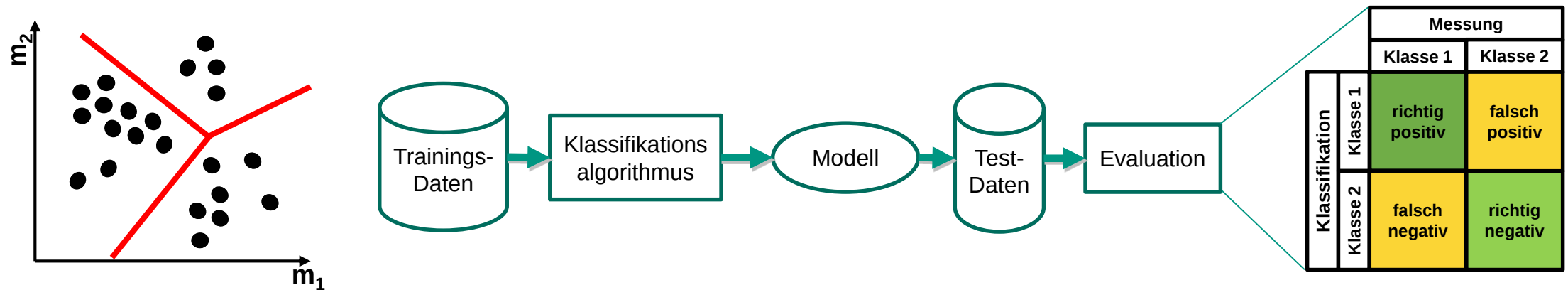
Übersicht der Verfahren und Anwendungsbeispiele



Überwachtes Lernverfahren

Klassifikation

- Überwachte Verfahren (Klassifikation) → Ground Truth ist beim Training und Test bekannt
 - Separierung von vorannotierten Daten

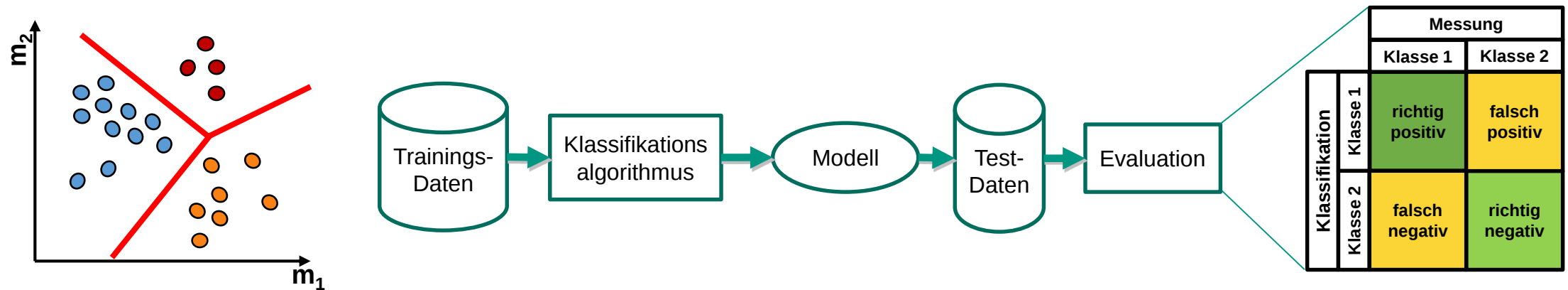


- Vorteil:
 - Zielgerichtet durch Vorannotation
 - Evaluation durch Testdaten
 - Vergleich der Modelle möglich (Optimierung)
 - Vergleich erleichtert Merkmalsselektion
- Nachteil:
 - Nur anwendbar auf vorannotierte Daten
 - Anwendung beschränkt auf die vorgegebenen Klassen

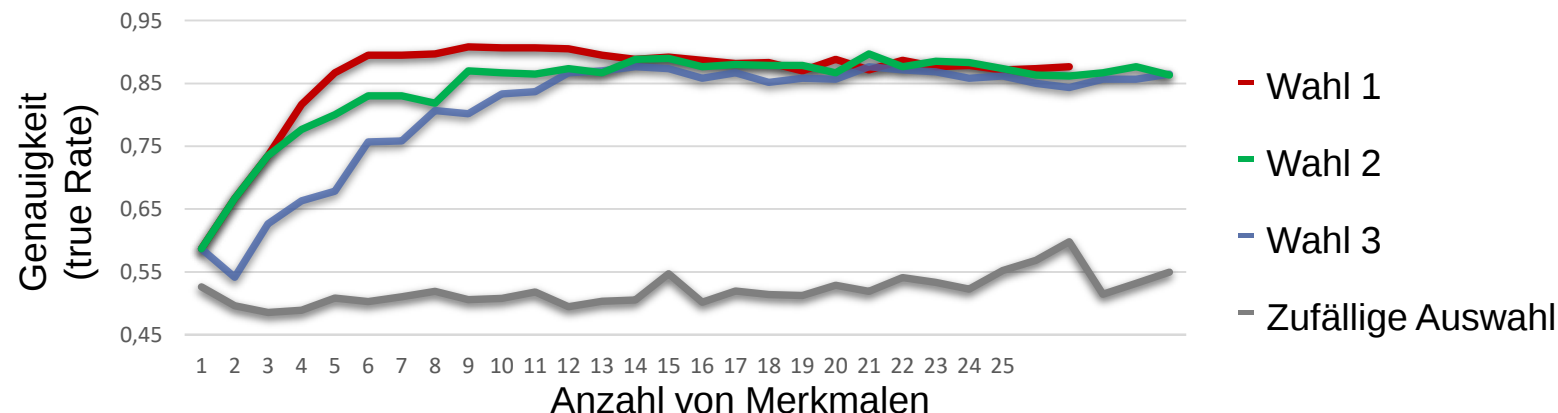
Überwachtes Lernverfahren

Klassifikation

- Überwachte Verfahren (Klassifikation) → Ground Truth ist beim Training und Test bekannt
 - Separierung von vorannotierten Daten



- Beispiel der Merkmalsselektion anhand vom Madelon Datensatz [1] und Decision Tree Klassifikation

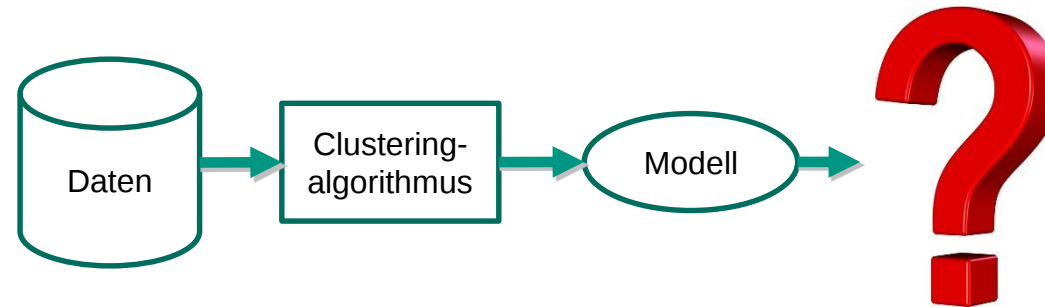
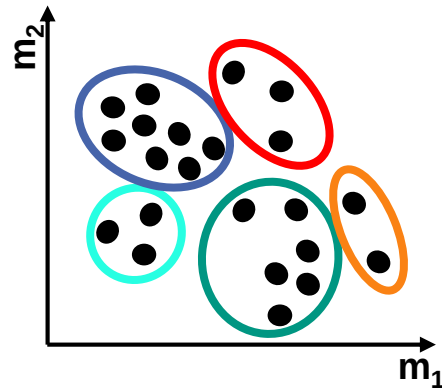


[1] Datensatz von NIPS (Neural Information Processing Systems) im Jahr 2003 für die „feature selection challenge“ erstellt

Unüberwachtes Lernverfahren

Clustering

- Unüberwachte Verfahren (Clustering) → Ground Truth ist unbekannt
 - Gruppierung von gleichartigen Datenpunkte / Situationen

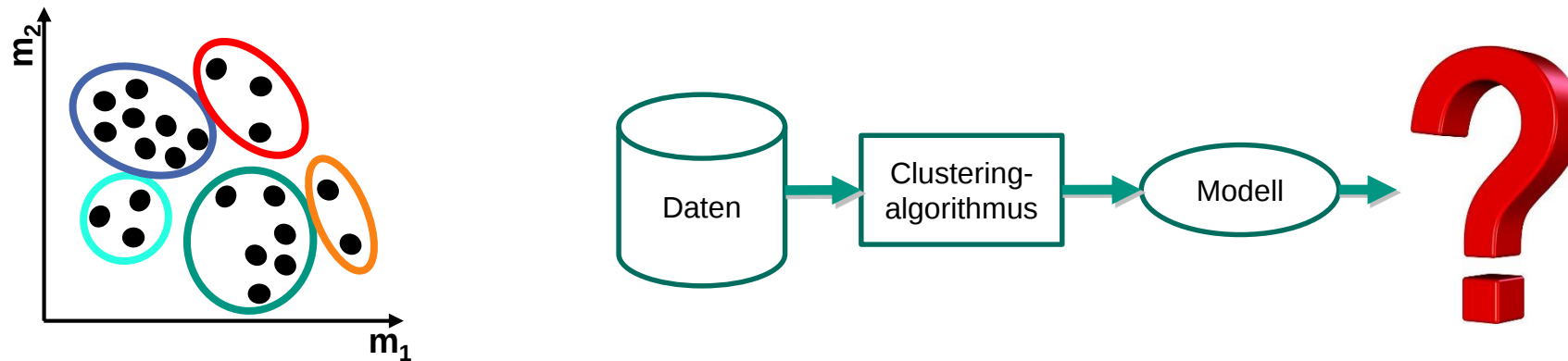


- Vorteil:
 - Benötigt keine Vorannotationen
 - Entdeckung von bisher unbekannten Differenzierbarkeiten
- Nachteil:
 - Nicht zielgerichtet
 - stark von Datenselektion, Data Preparation und Clustering-Verfahren abhängig
 - Bewertung der Ergebnisse schwer

Unüberwachtes Lernverfahren

Clustering

- Unüberwachte Verfahren (Clustering) → Ground Truth ist unbekannt
 - Gruppierung von gleichartigen Datenpunkte / Situationen

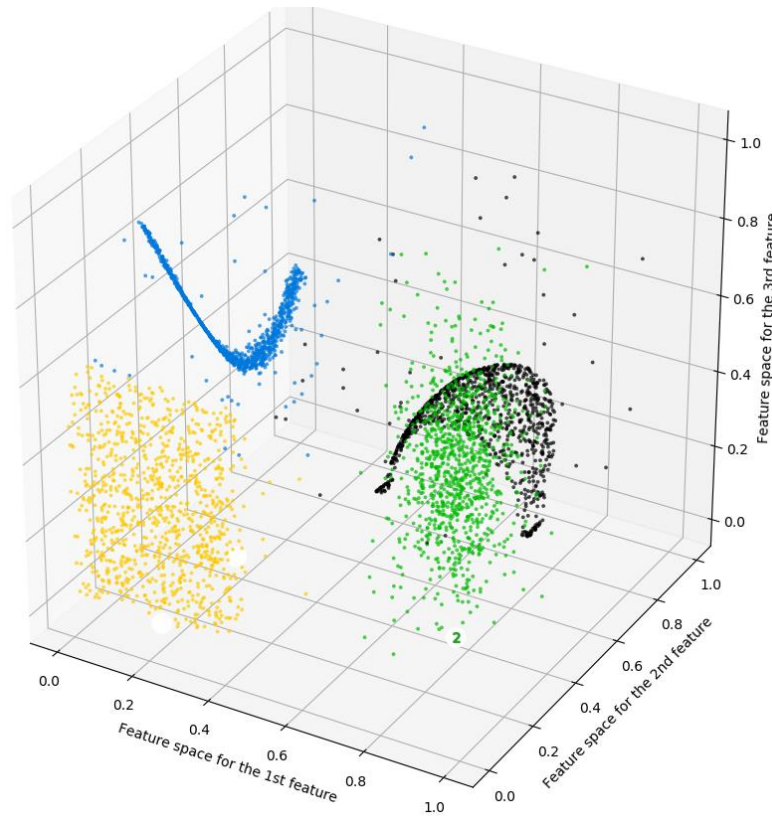


- Gütemaß von Clustering-Ergebnissen nur durch Berechnung der „Separierung“ möglich
→ eine „gute“ Separierung ist oft unabhängig von der gewünschten Anwendung!
- Beispiele für Berechnungen der Clustering-Güte:
 - Sum of Squared Errors (SSE)
 - Distanz jedes Objekts zum Cluster-Zentrum des Objekts
 - $SSE = \sum_{i=1}^k \sum_{x \in C_i} (dist(x, c_i))^2$
 - Silhouetten Koeffizient (S_C)
 - $S_C = \frac{1}{n} \sum_{o=1}^n \frac{b(o) - a(o)}{\max\{a(o), b(o)\}}$
 - $a(o)$: Abstand eines Objekts o zum Repräsentanten seines Clusters
 - $b(o)$: Abstand zum Repräsentanten des nächsten Clusters

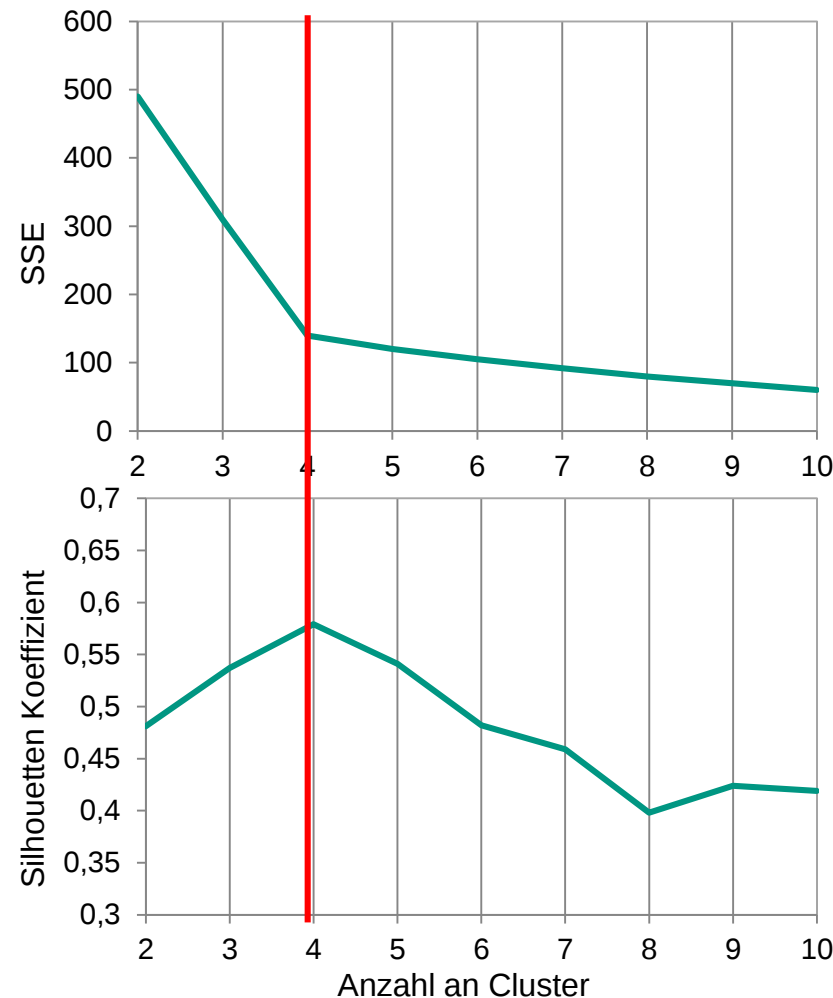
Unüberwachtes Lernverfahren

Bestimmung der Clustering-Güte bzw. Anzahl der gesuchten Cluster

Synthetischer-Beispiel-Datensatz



SSE und Silhouette Koeffizient (S_C)
für verschiedene # an Cluster



Faustformel:

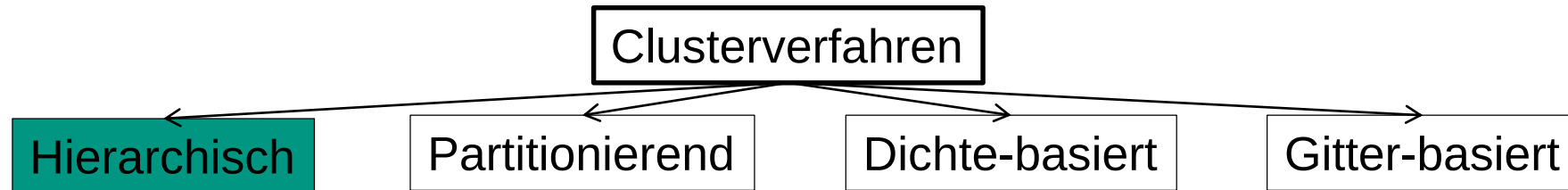
Beste Anzahl an Cluster ist
beim „Ellenbogen“ (ein Knick
in der Linie)

Faustformel:

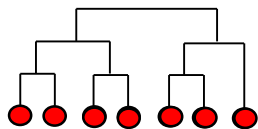
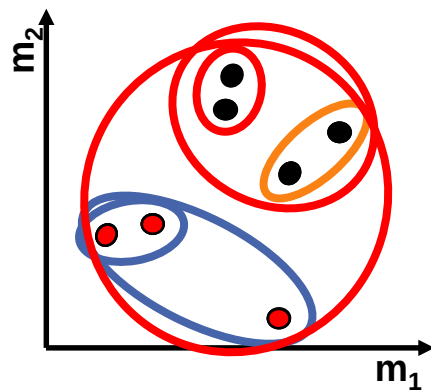
$S_C > 0,7$ „starke“
Struktur

$S_C > 0,5$ „brauchbare“
Struktur

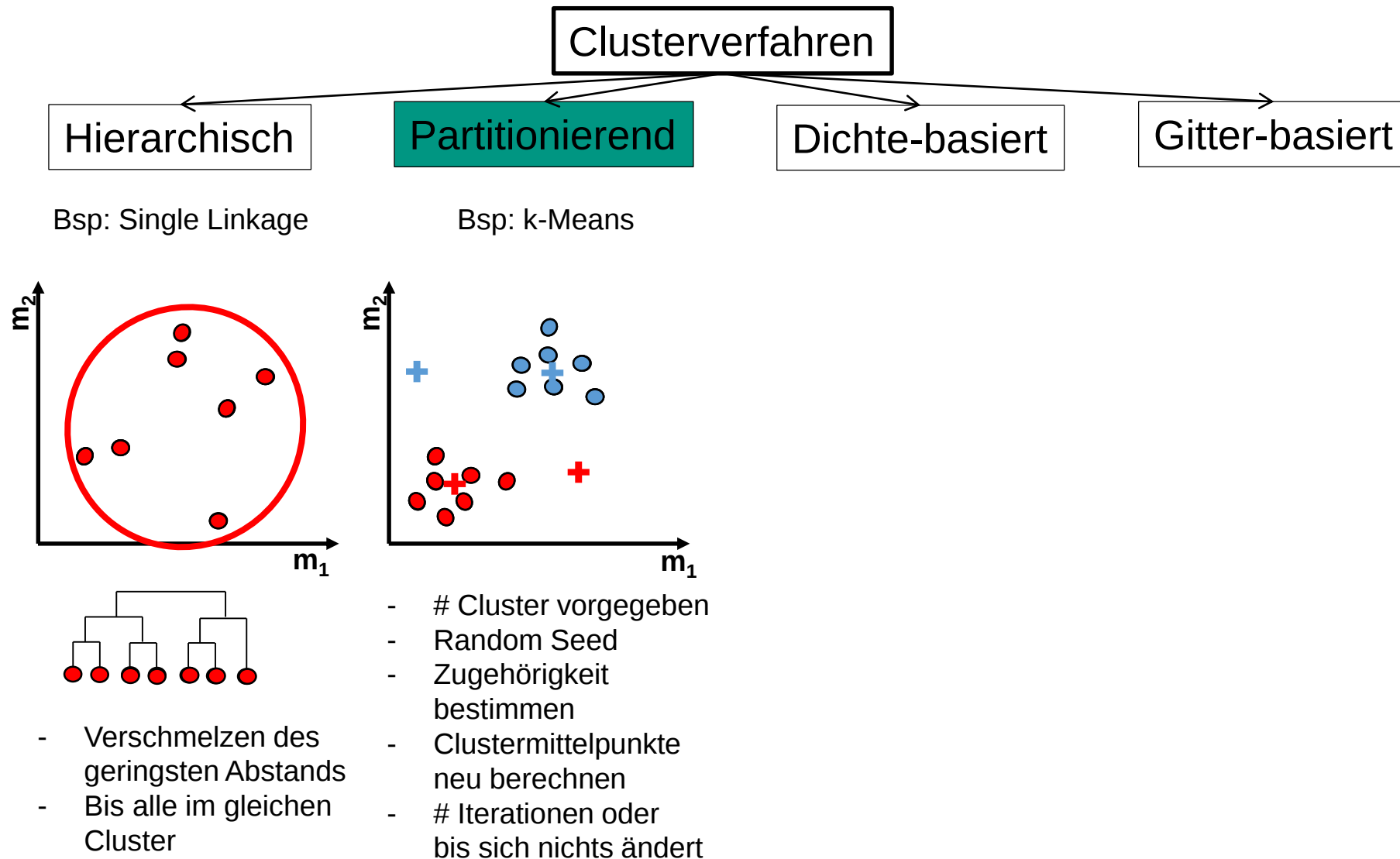
$S_C < 0,2$ „unbrauchbare“
Struktur

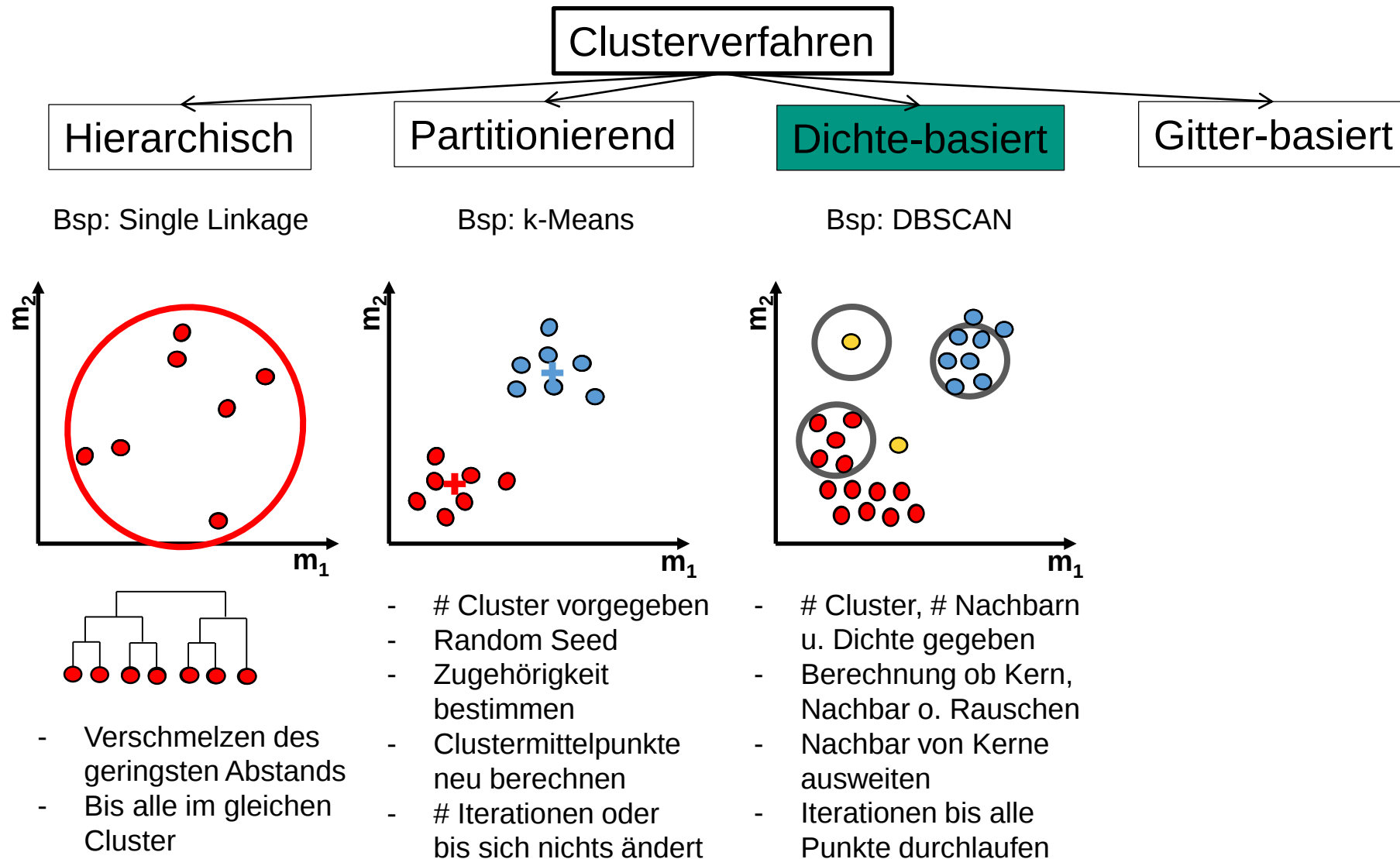


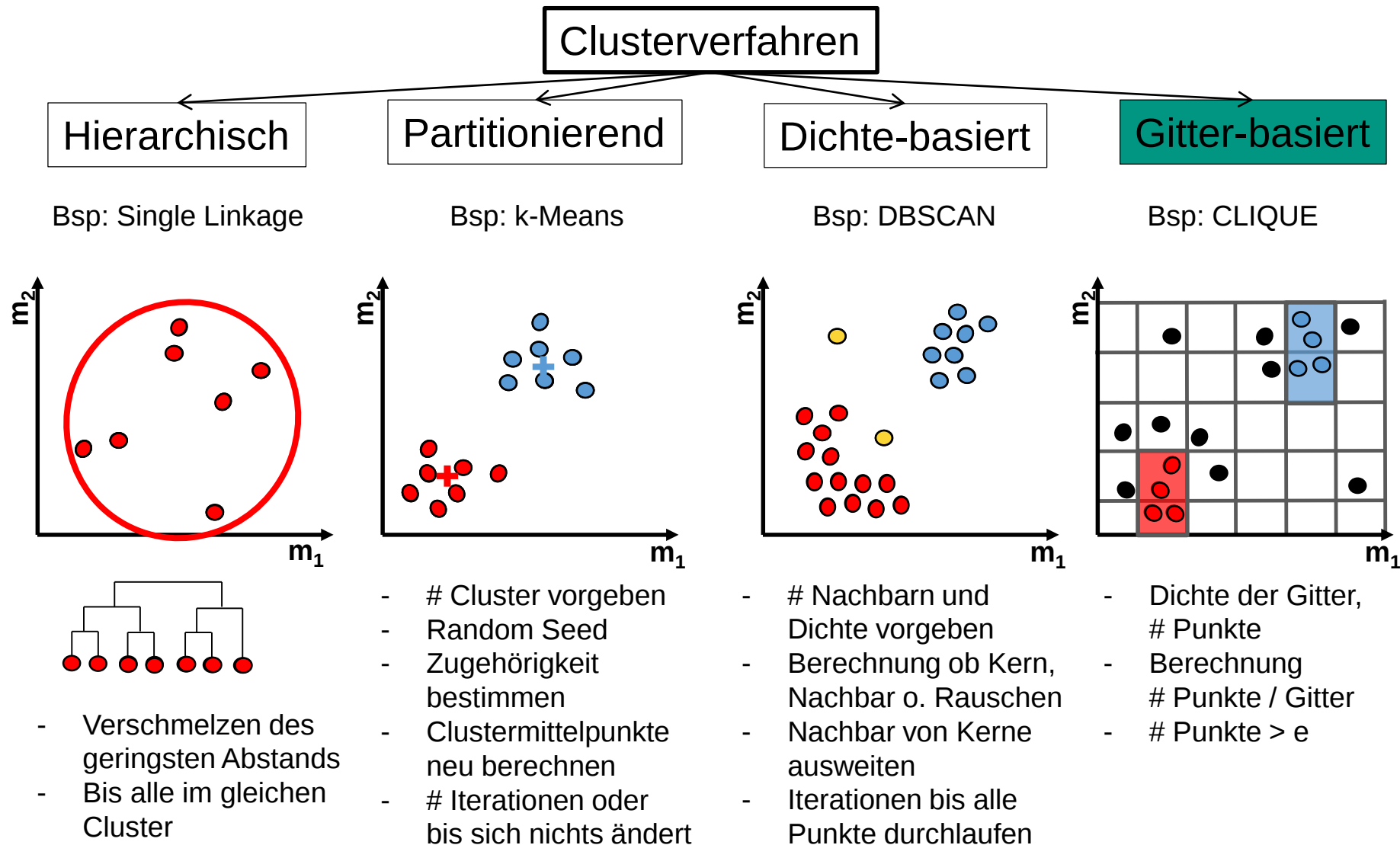
Bsp: Single Linkage



- Verschmelzen des geringsten Abstands
- Bis alle im gleichen Cluster







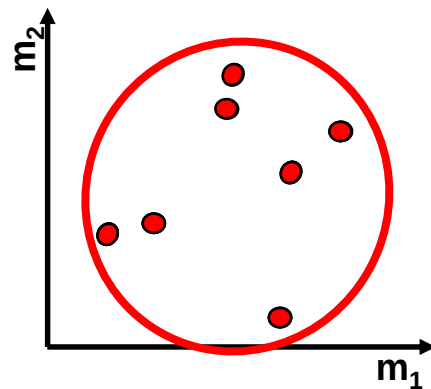
Verschiedene Clusterverfahren

Zwischenübung

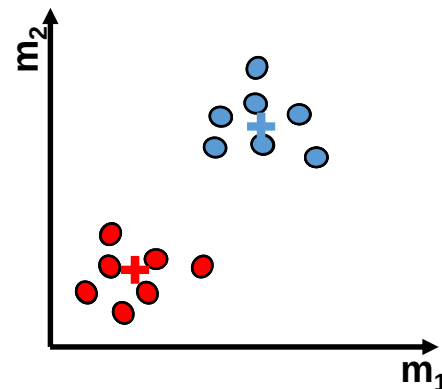
- Geben Sie für jeden der vier Familien der Clusterverfahren je ein Vor- und ein Nachteil an.



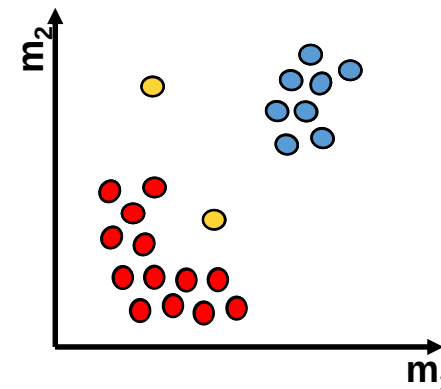
Hierarchisch



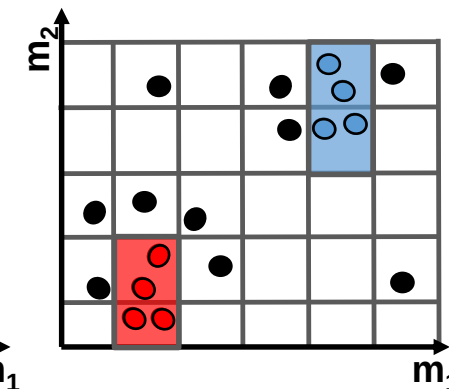
Partitionierend



Dichte-basiert



Gitter-basiert



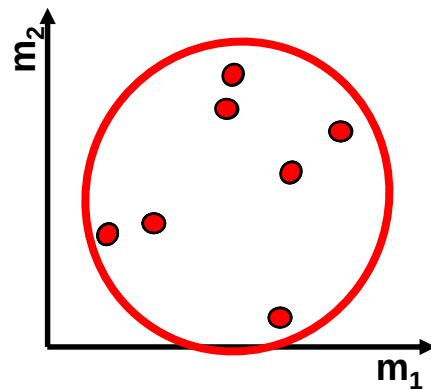
Verschiedene Clusterverfahren

Zwischenübung - Lsg

- Geben Sie für jeden der vier Familien der Clusterverfahren je ein Vor- und ein Nachteil an.



Hierarchisch

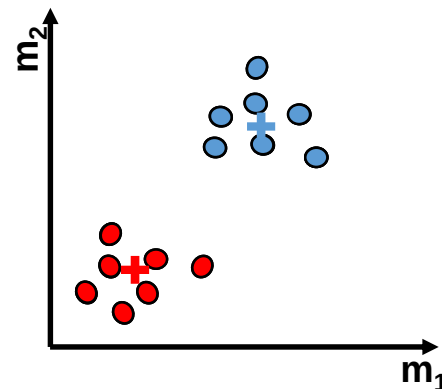


- Cluster Hierarchie
- Keine Parameterwahl



- Keine Iteration
- Keine Optimierung

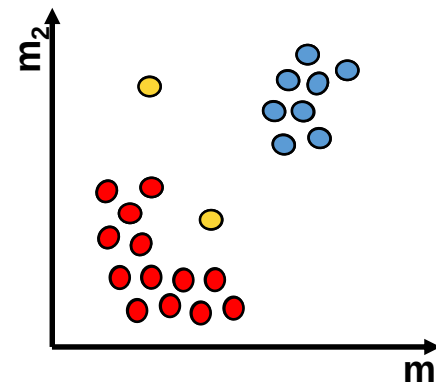
Partitionierend



- Schnell
- Daten unabhängig

- Anzahl Cluster muss gewählt werden

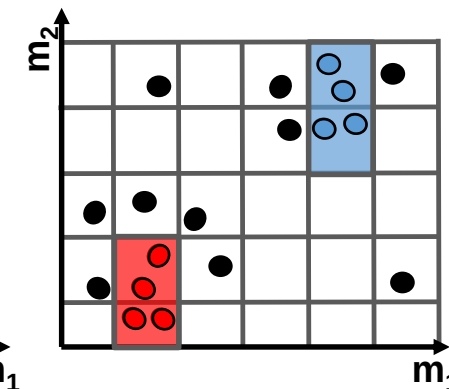
Dichte-basiert



- Unregelmäßige Formen
- Rauschunterdrückung

- Sensibel auf Parameterwahl

Gitter-basiert



- Für hochdimensionale Räume geeignet

- Eher grob
- Wenig Optimierungspotenzial

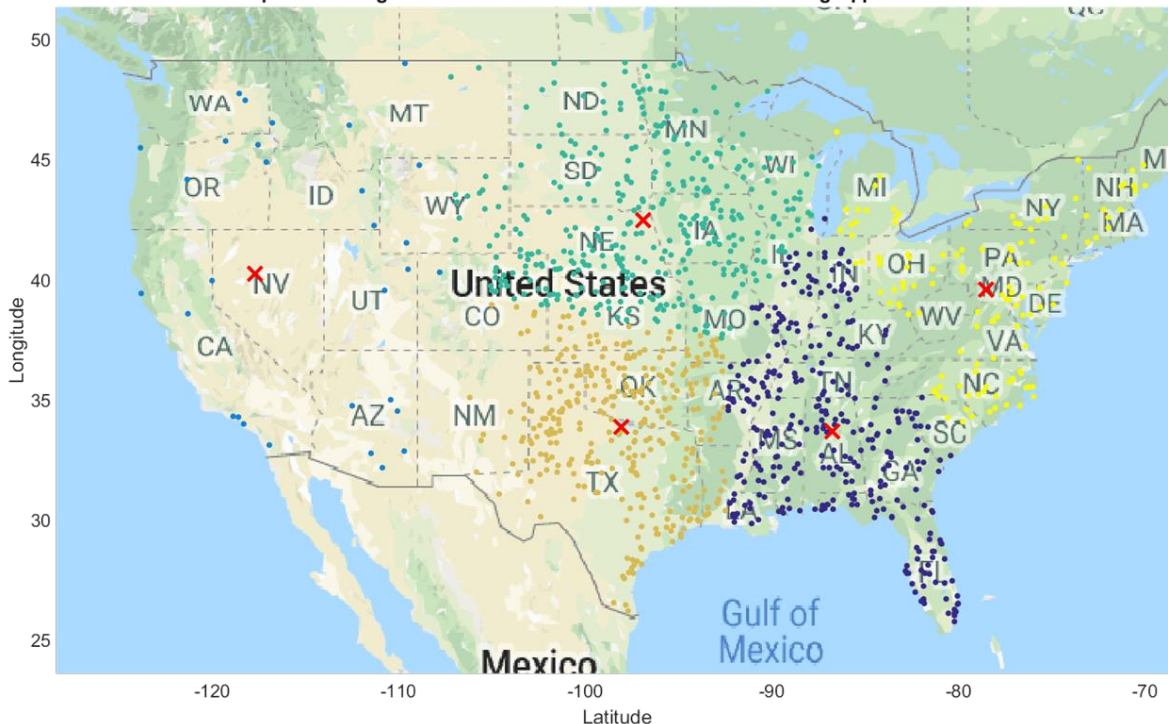
Clustering - Algorithmen

K-means

- Verfahren der Vektorquantisierung
- Bildung Mengen von ähnlichen Objekten in k Gruppen
möglichst geringe Varianz, Gruppen ähnlicher Größen

Datensatz in k Partitionen teilen, so dass Summe der quadrierten Abweichungen von Cluster-Schwerpunkten minimal wird

Exemplarisch einige Stürme in den USA zwischen 1950 und 2015 gruppiert mit k-means



$$J = \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

x_j : Datenpunkt
 S_i : Cluster
 μ_i : Schwerpunkt

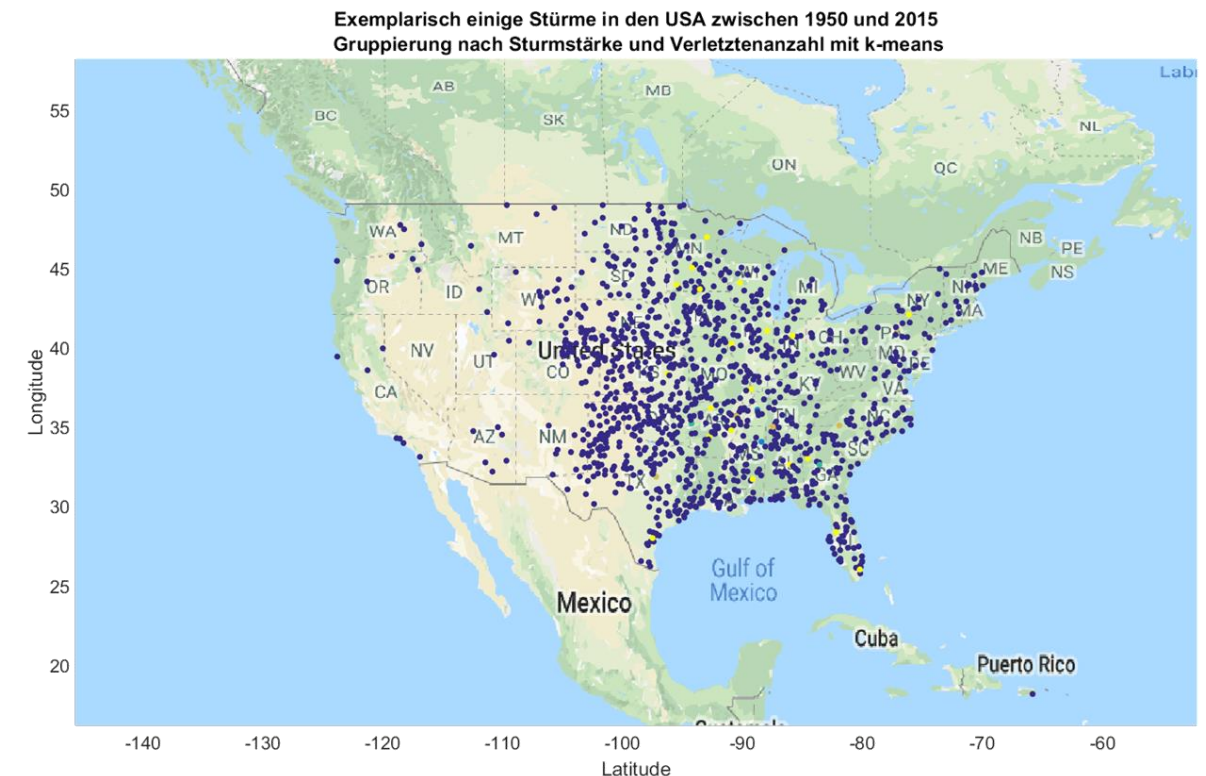
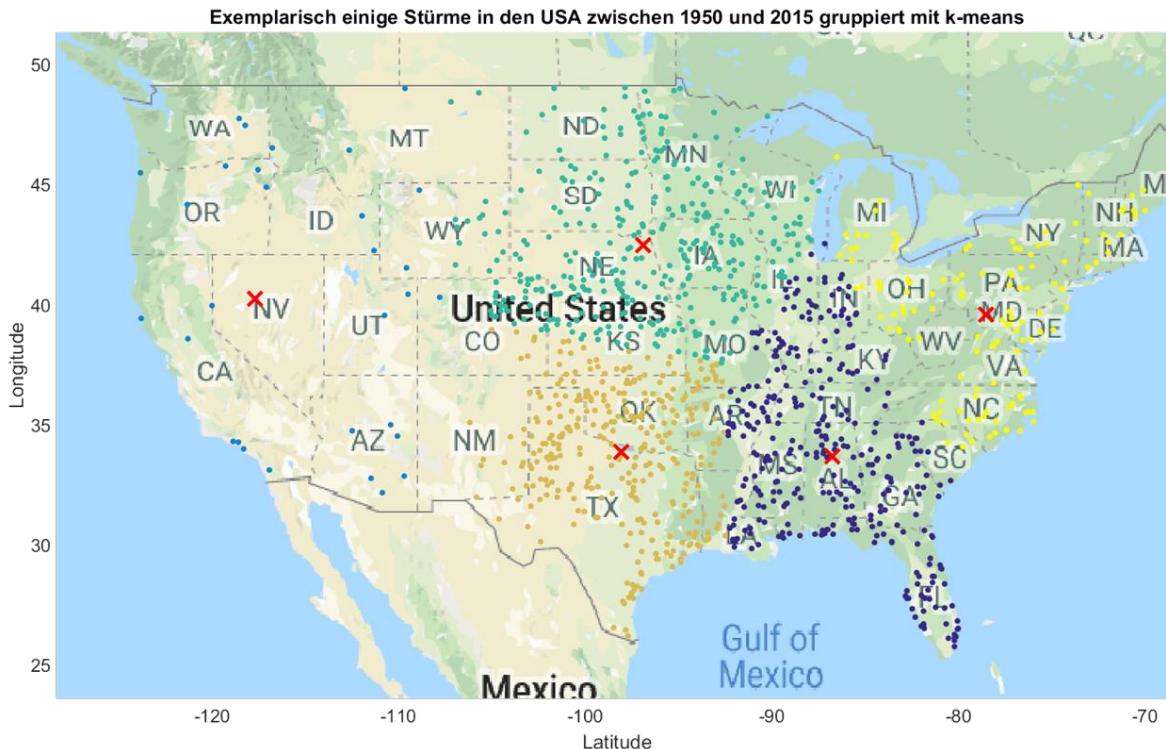
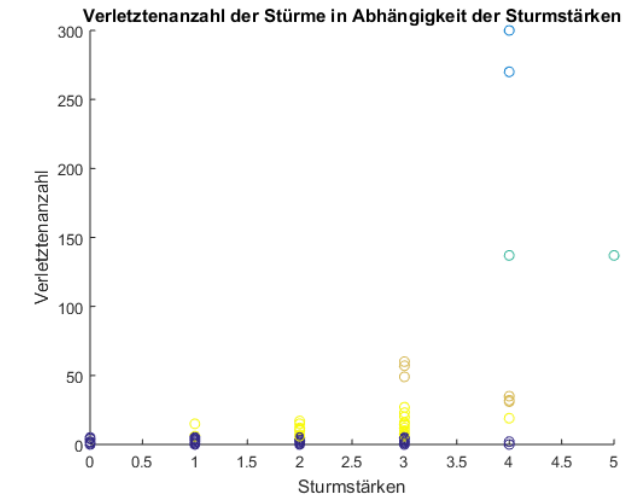
Lloyd Algorithmus:

1. k unterschiedliche Zentren $c_1, c_2, \dots, c_k$
2. Solange sich die Zielfunktion verbessert:
Partitioniere P in Cluster $S_1, S_2, \dots, S_k$ dass S_i die Punkte aus P enthält, deren nächstgelegenes Zentrum c_i ist
Für jedes $1 \leq i \leq k$ sei $c_i \leftarrow \mu(C_i)$

Clustering

K-means

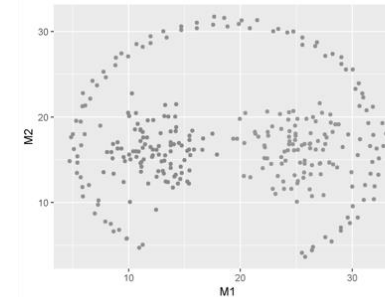
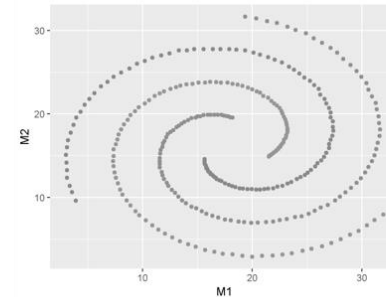
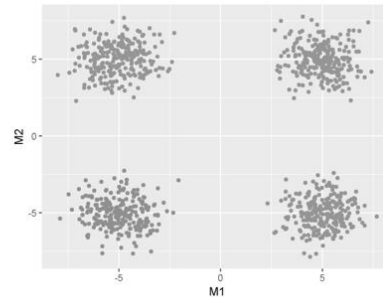
- Links:
Clusterbildung nach GPS-Daten
- Rechts:
Clusterbildung nach Sturmstärken und Verletztenanzahl



Clustering - Algorithmen

Zwischenübung - Lsg

- Benennen Sie jeweils den Clusteralgorithmus, den Sie verwenden würden um den jeweiligen Datensatz bestmöglich zu clustern. Benennen Sie außerdem die nötigen Hyperparameter, die bestimmt werden müssen, damit der jeweilige Clusteralgorithmus ausgeführt werden kann.

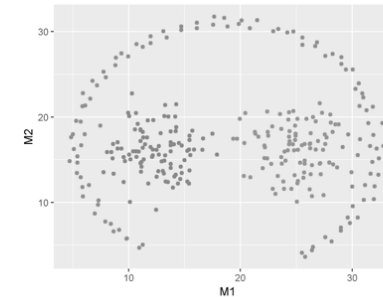
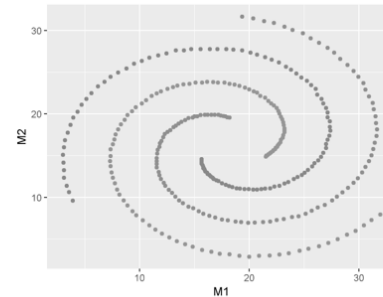
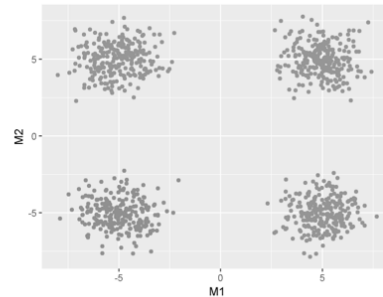


Algorithmus-Vorschlag:			
Welche Hyperparameter müssen für diesen Alg. bestimmt werden:			

Clustering - Algorithmen

Zwischenübung - Lsg

- Benennen Sie jeweils den Clusteralgorithmus, den Sie verwenden würden um den jeweiligen Datensatz bestmöglich zu clustern. Benennen Sie außerdem die nötigen Hyperparameter, die bestimmt werden müssen, damit der jeweilige Clusteralgorithmus ausgeführt werden kann.



Algorithmus-Vorschlag:	K-means	DBSCAN	DBSCAN
Welche Hyperparameter müssen für diesen Alg. bestimmt werden:	• K – Anzahl der Klassen (hier 4)	• MinPts • Epsilon	• MinPts • Epsilon

Clustering

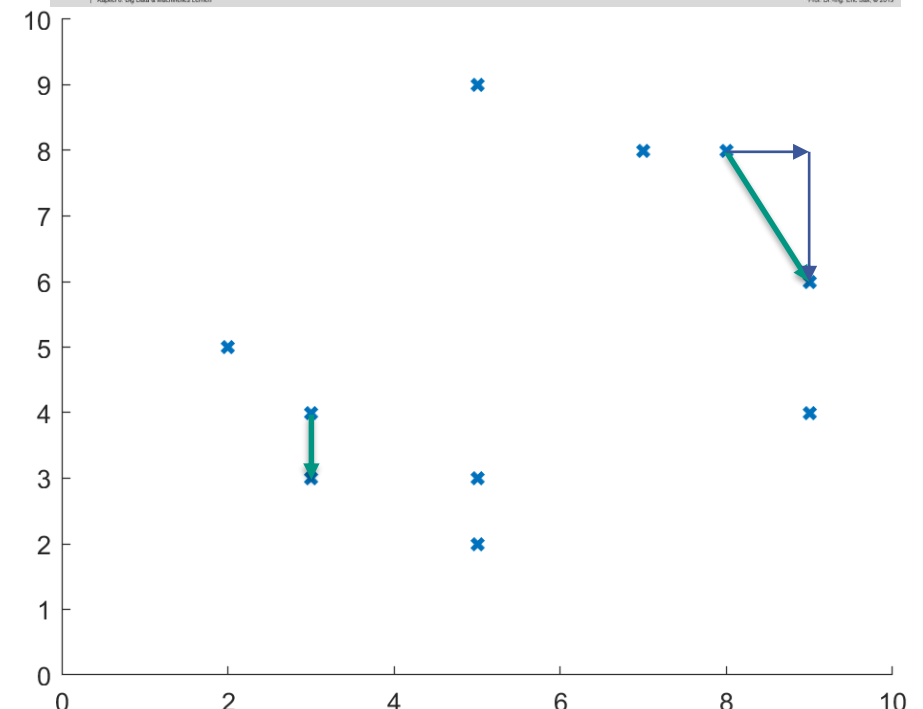
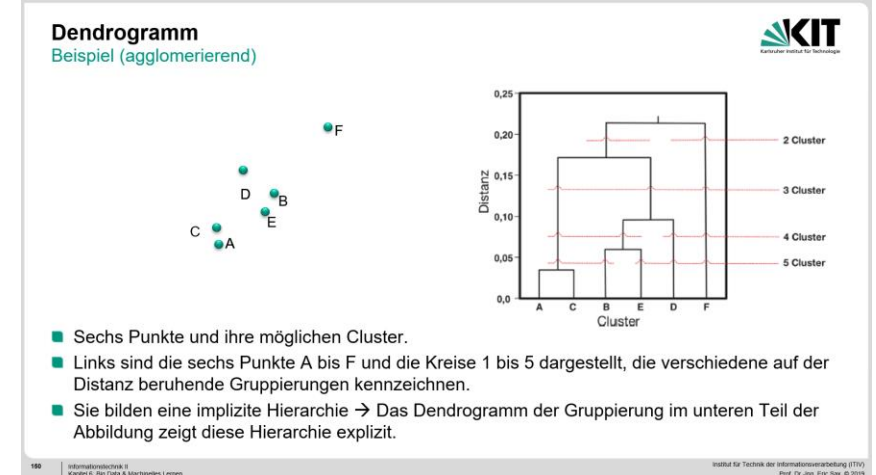
Dendrogramme

- Visualisierung hierarchischer Baumstruktur
 - Jede Wurzel repräsentiert ein einzelnes Cluster, welches die komplette Datenmenge umfasst
 - Jedes Blatt repräsentiert ein Cluster, in denen sich lediglich ein Objekt befindet
 - Jeder Knoten repräsentiert die Vereinigung aller Kindknoten/Blätter
 - Die Verbindungen stellen Ähnlichkeitsmaß dar

Grafische Erstellung über Entfernung
Gesucht Abstand zwischen Datenpunkten

$$|\overline{x_a}| = \begin{vmatrix} x_{a1} \\ x_{a2} \end{vmatrix} = \sqrt{x_{a1}^2 + x_{a2}^2} \downarrow$$

Vektor zwischen zwei Punkten:
Linearkombination $\overline{x_a} = \overline{x_1} + \overline{x_2}$



Clustering

Dendrogramm

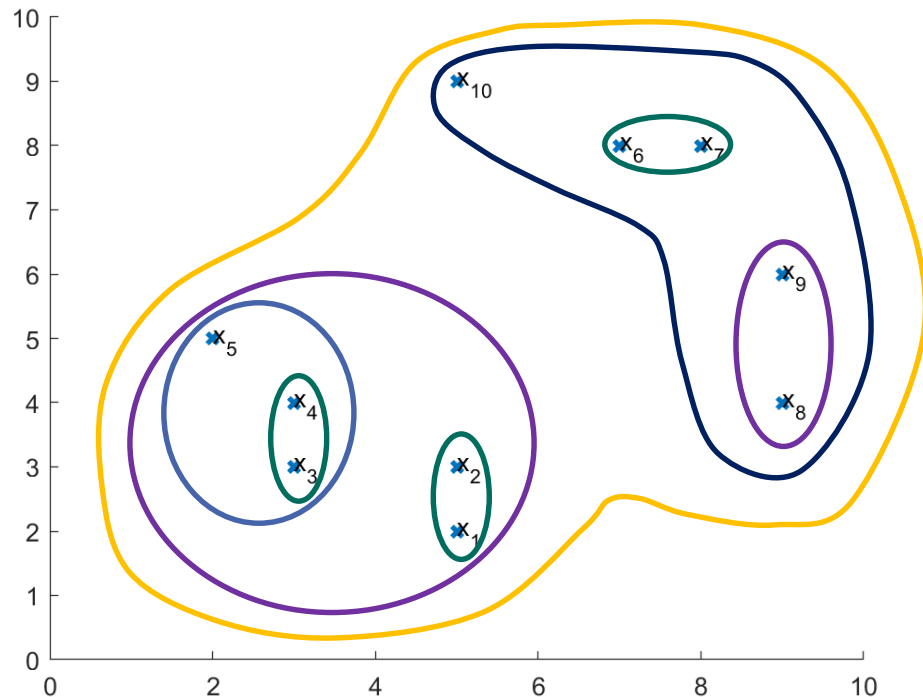
Ebene

Abstand

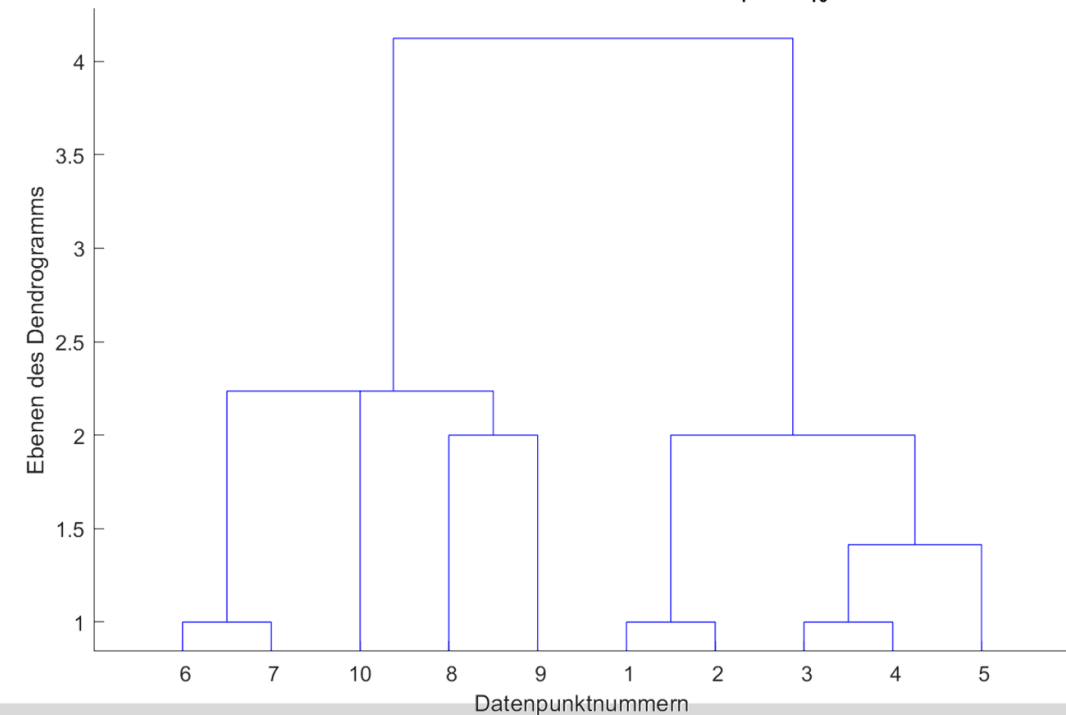
Punkte

Vektor zwischen zwei Punkten: $\overline{x_a} = \overline{x_1} + \overline{x_2}$

Abstand: $|\overline{x_a}| = \left| \begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} \right| = \sqrt{x_{a1}^2 + x_{a2}^2}$



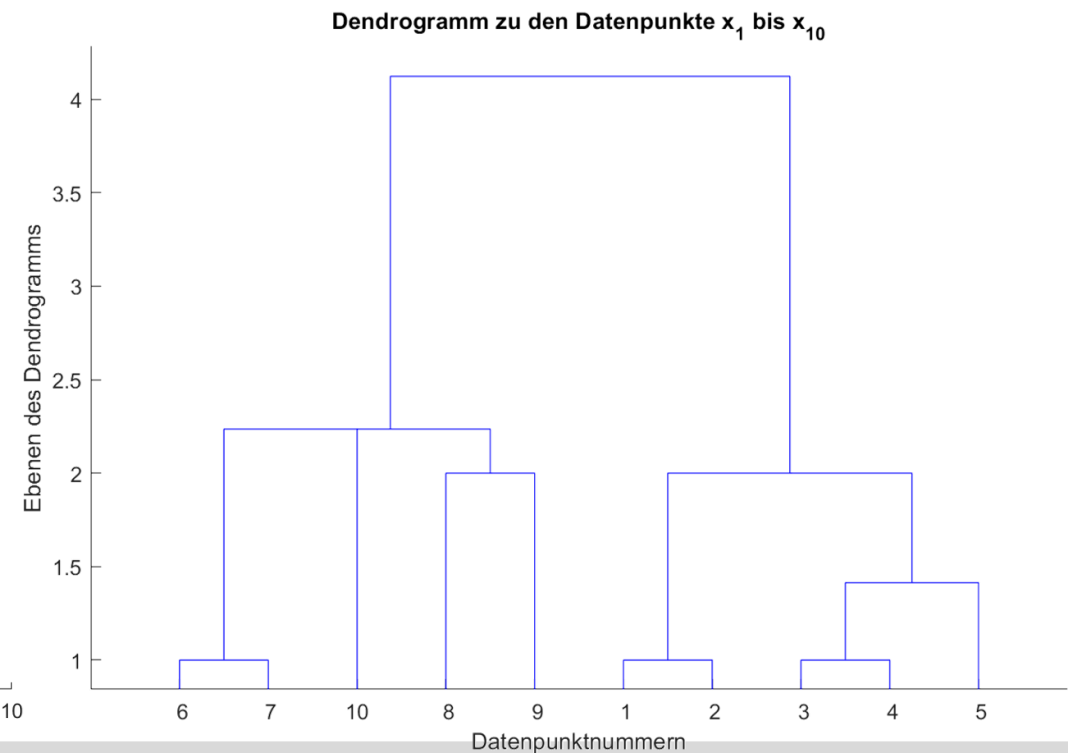
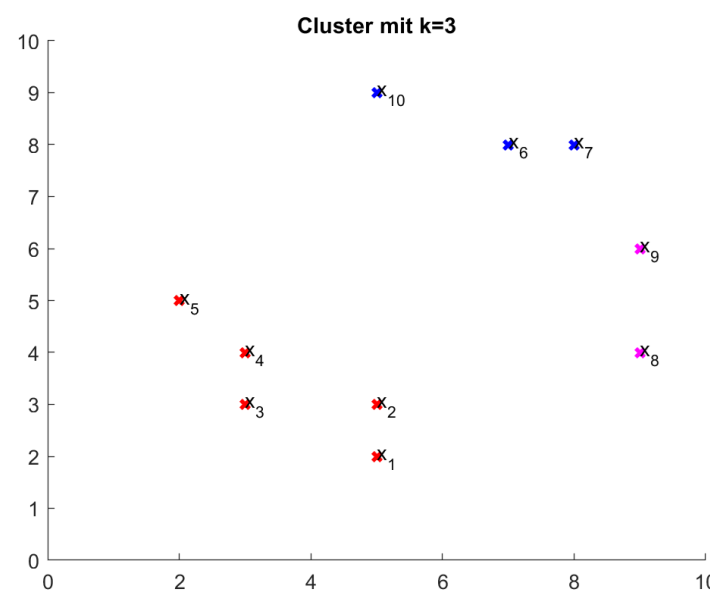
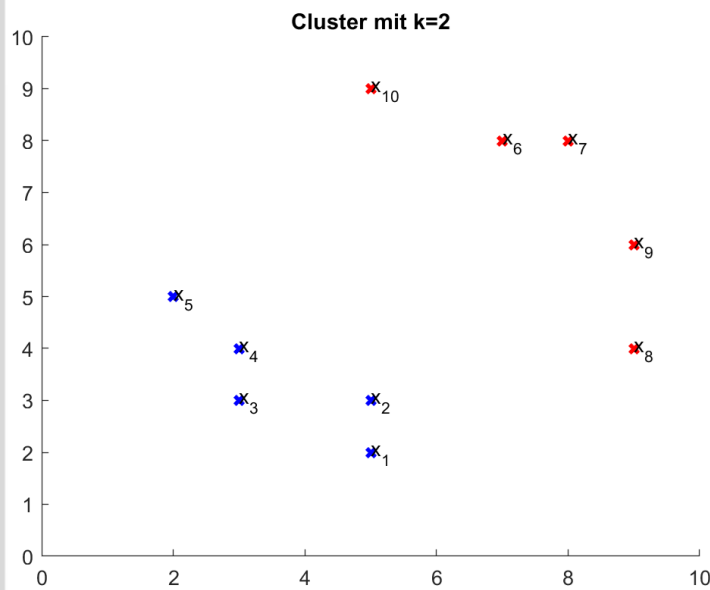
Dendrogramm zu den Datenpunkte x_1 bis x_{10}



Clustering

K-means mit unterschiedlichen k

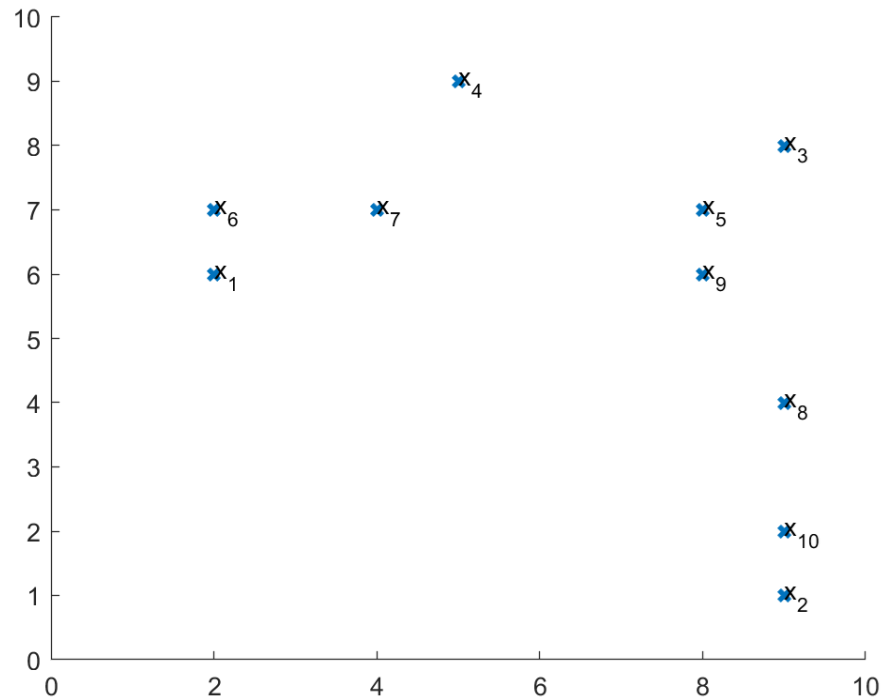
- Dendrogramm gibt Aufschluss über die zu wählende Anzahl der Cluster



K-means und Dendrogramm

Zwischenübung

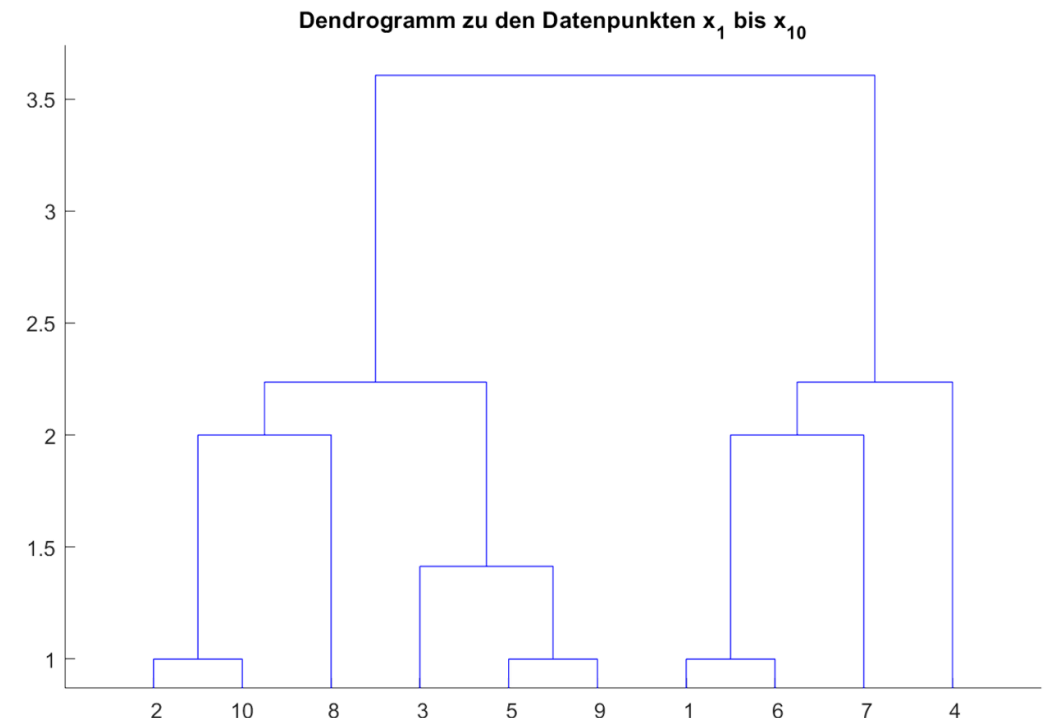
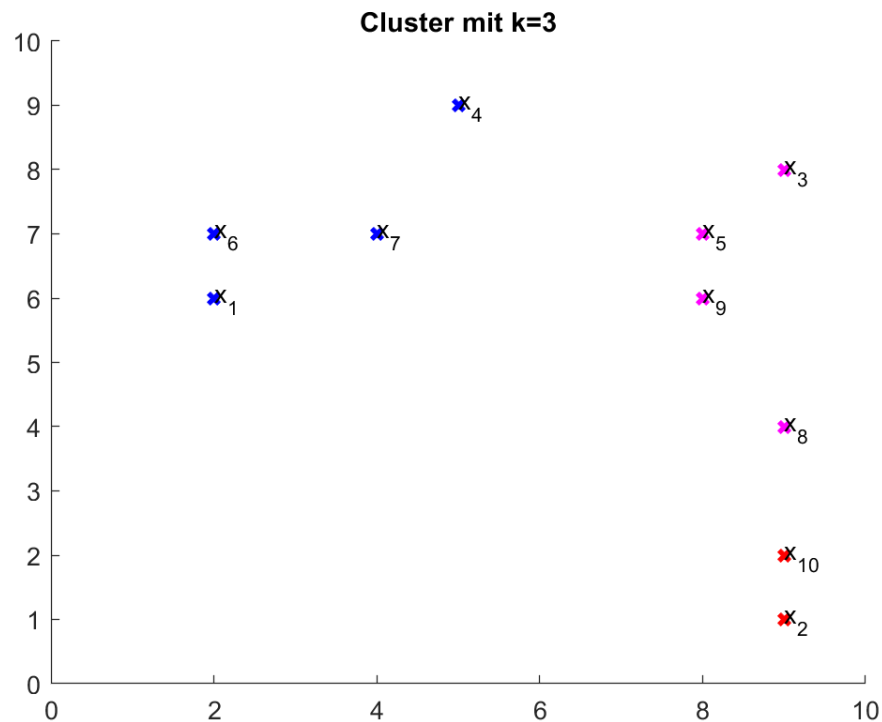
- Zeichnen Sie das zugehörige Dendrogramm zu den gegebenen Datenpunkten und geben Sie auf Basis dessen ein geeignetes k für einen k-means Algorithmus an



K-means und Dendrogramm

Zwischenübung - Lsg

- Zeichnen Sie das zugehörige Dendrogramm zu den gegebenen Datenpunkten und geben Sie auf Basis dessen ein geeignetes k für einen k-means Algorithmus an



Ziele der heutigen Übung



■ Nach der heutigen Übung können Sie....

- ...Ansätze der unüberwachten maschineller Lernverfahren auf ihre Anwendbarkeit und Wirksamkeit einschätzen

1

- ... Unterschiede zwischen unüberwachten und überwachten Lernverfahren benennen

2

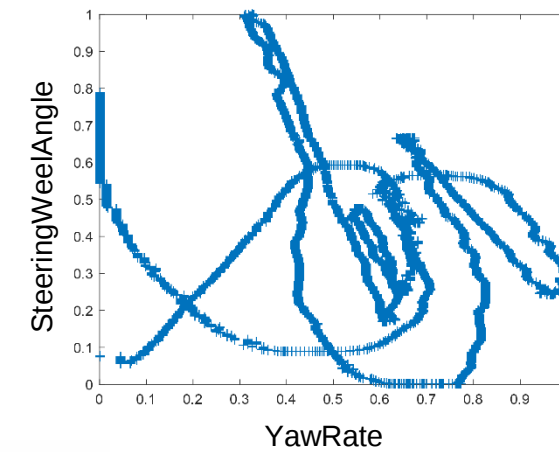
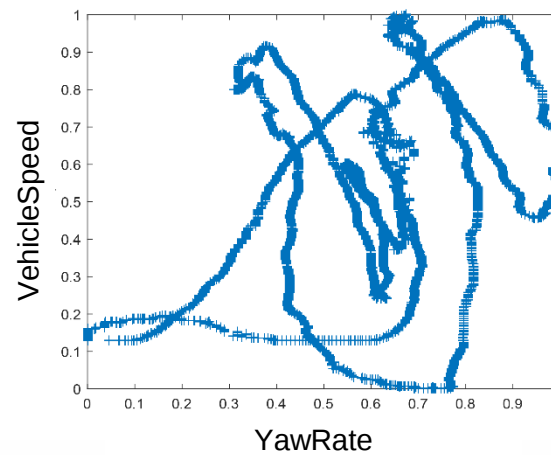
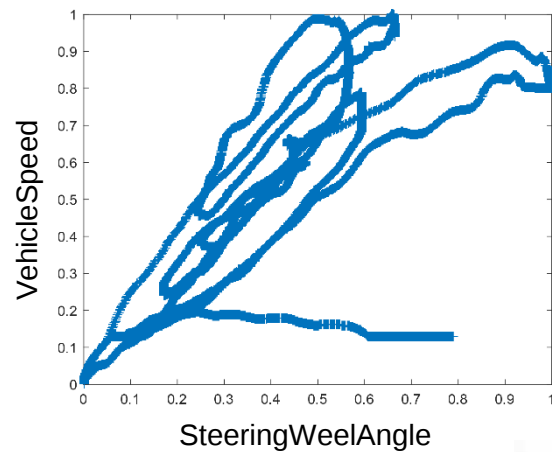
- ... die Familien der Clusterverfahren mit Vor- und Nachteilen benennen

3

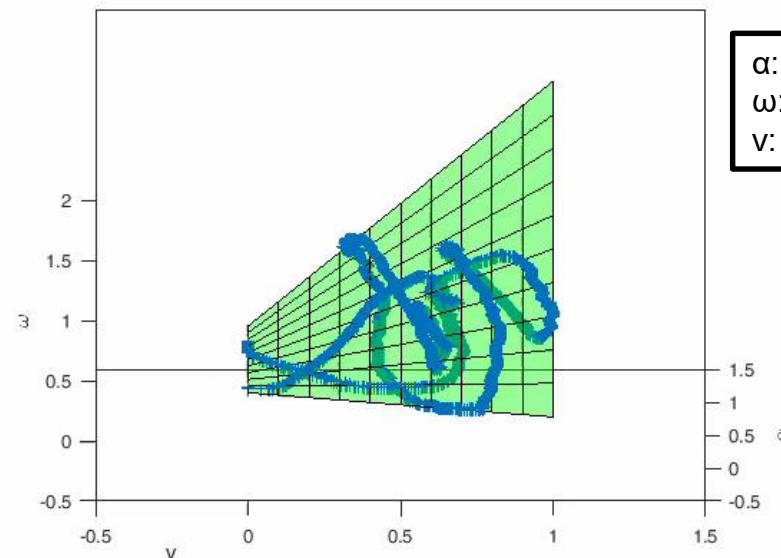
- ... Modelle unter Anwendung unüberwachter maschineller Lernverfahren aufstellen

Beispiel eines Musters

■ 3 Signale als 2D Plot



■ Als 3D Plot



α : Lenkradwinkel
 ω : Gierrate
 v : Geschwindigkeit