

Übung 6

Übung zu Informationstechnik II und Automatisierungstechnik – Nathalie Brenner

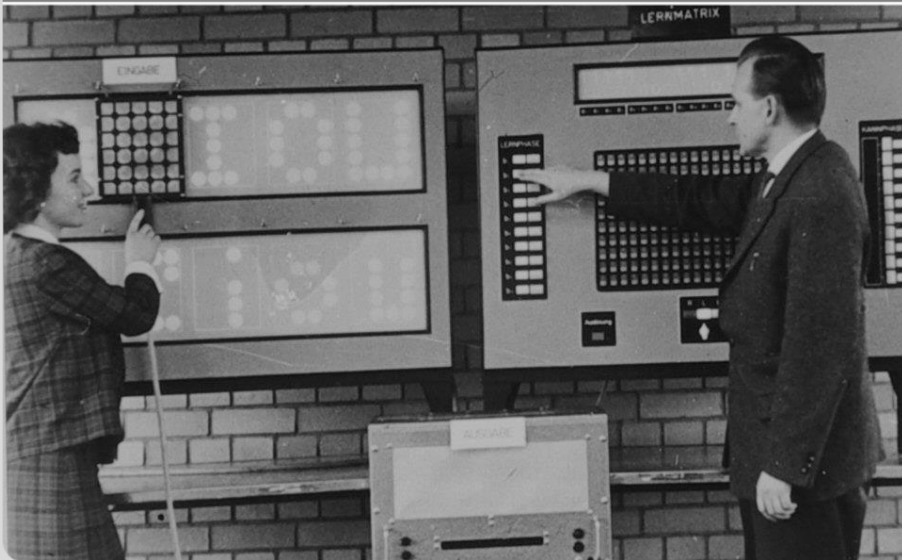
Prof. Dr.-Ing. Eric Sax

Institutsleitung

Prof. Dr.-Ing. J. Becker

Prof. Dr.-Ing. E. Sax

Prof. Dr. rer. nat. W. Stork



ORGANISTAROISCHES KLAUSUR SS2020

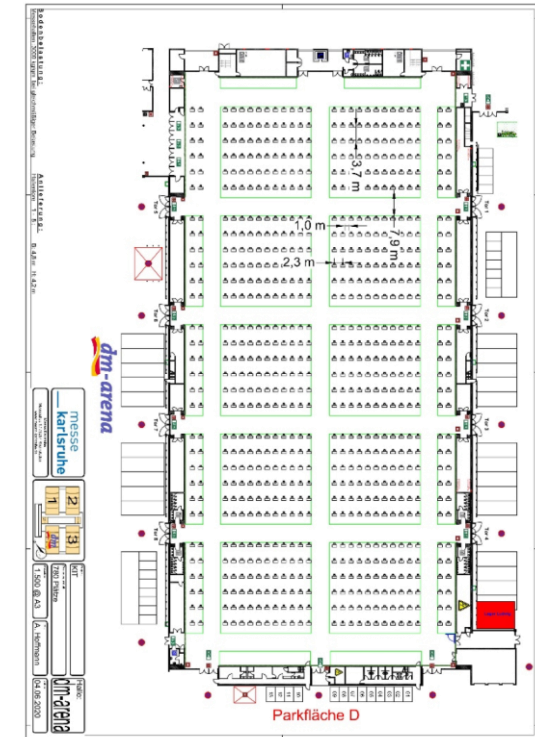
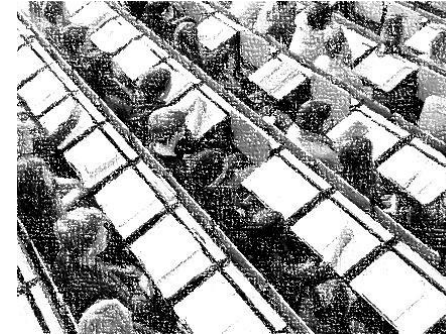


Organisatorisches

Informationen zur Klausur SS2020

- Inhalt: Vorlesung und Übung
- Termin: Freitag: **31.07 2020 09:00 bis 11:00 Uhr**
 - Anmeldefrist: 24.07.2020 23:59
 - Abmeldefrist: 30.07.2020 23:59
- Dauer: 2 Stunden
- Einlass: ab 30 Minuten vor Prüfungsbeginn (*Änderungen vorbehalten*)
- Ort: DM-Arena, Messehalle 1, Rheinstetten
- Hilfsmittel: 1x A4 Blatt (2x A4 Seite, handschriftlich, lesbar ohne Hilfsmittel)

Genaue Informationen zum Ablauf der Prüfung werden bis zum 24.07.2020 bekannt gegeben. Bitte sehen Sie bis dahin von weiteren Nachfragen ab!



Organisatorisches

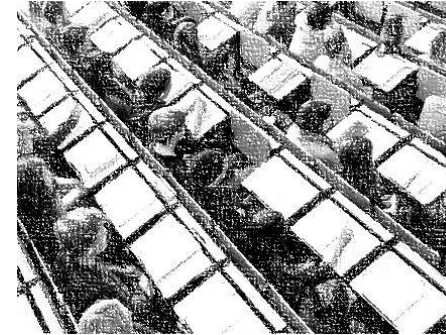
Informationen zur Klausur SS2020

■ Mitzubringen:

- KIT-card oder Lichtbildausweis
- Schreibzeug
- **Unterschiedene Erklärung**

„Erklärung über den fehlenden Verdacht einer Infektion mit dem Coronavirus“
(Verfügbar auf Ilias ab 24.07.2020
oder unter <https://intranet.kit.edu/downloads/erklaerung-praesenzpruefung.pdf>)

- Mund-Nasen-Bedeckung ist bei Betreten und Verlassen des Prüfungsraums Pflicht. Während der Abfassung der Klausur kann sie abgelegt werden.
- Weitere Informationen werden direkt nach Erhalt in Ilias veröffentlicht. Wir bitten Sie sich regelmäßig nach aktualisierten Informationen zu erkundigen!



- Organisatorisches zur Klausur



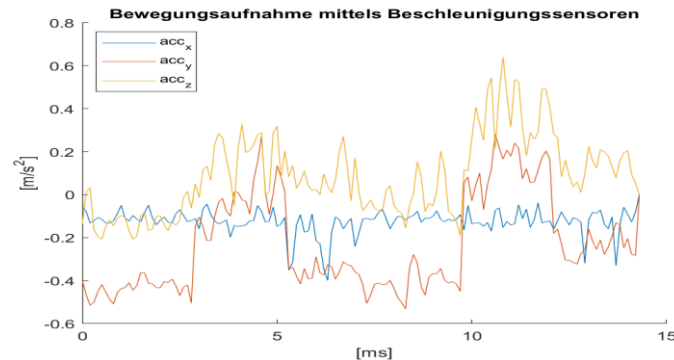
WIEDERHOLUNG ÜBUNG 5



Wiederholung Übung 5

Datenaufbereitung

Datenbereinigung



Datenaufbereitung – Anomaliedetektion über Anomaliemaß Zwischenübung

- Führen Sie eine Anomaliedetektion von Sensor 1 und 2 über die **Standardabweichung** und über die **Whisker eines Boxplots** durch:
 - Als Akzeptanzbereich für die **Standardabweichung** wurde $1,9\sigma$ festgelegt
 - Als **Whisker Grenze** wurde das 1,5 Fache des Interquartilsabstands (IQR) festgelegt

Zeit (h.)	Sensor 1	Sensor 2
3	13	125
6	8	145
9	255	99
11	n/a	278
15	15	133
19	54	155
21	9	n/a
24	5	136
1534	-13	-12
1537	-6	58

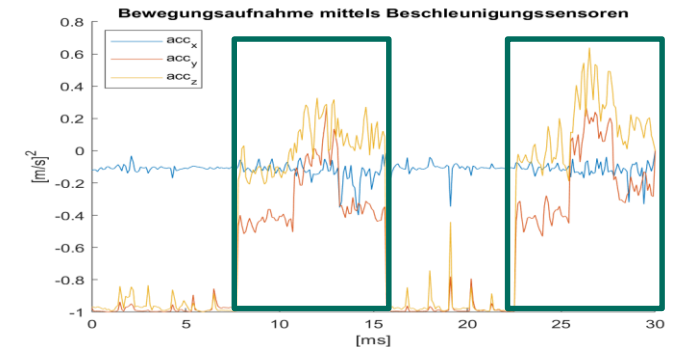
Mittelwert $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$
 Varianz $\sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$
 Standardabweichung $\sigma = \sqrt{\sigma^2}$
 Median = $\begin{cases} x_{n+1/2} & n \text{ ungerade} \\ 1/2 (x_{n/2} + x_{(n/2)+1}) & n \text{ gerade} \end{cases}$
 1. bzw. 3. Quartil ist der Median aus den Zahlen unterhalb bzw. oberhalb des 50%-Medians (bei gerade Anzahl wird der 50%-Median hinzugezogen)

- Sensor 1:
 - Mittelwert = 22
 - Standardabweichung = 79,79
 - 1. Quartil = 5
 - 3. Quartil = 15
- Sensor 2:
 - Mittelwert = 124,11
 - Standardabweichung = 73,65
 - 1. Quartil = 99
 - 3. Quartil = 145

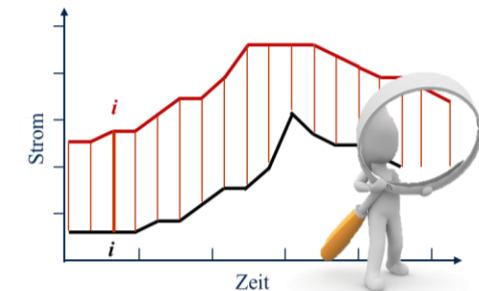


Zeitfensterauswahl

- Tag/Nacht?
- Sommer/Winter
- Event getriggert



- Reduktion der Daten auf relevante Situationen, um
 - invalide Ergebnisse zu vermeiden (Zusammenhang anderer Situationen)
 - Datenmenge für weitere Bearbeitung zu reduzieren
- Wiederfindung ähnlicher Situationen / Signalverläufe in Daten
 - Herausforderung: Streckung oder Stauchung gleicher Situationen



Euklidische Abstände → schlechtes Ähnlichkeitsmaß

Wiederholung Übung 5

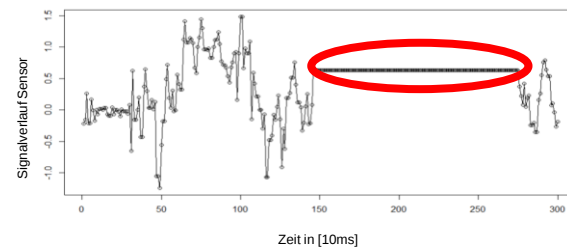
Datenaufbereitung

Datenmanipulation

1. Umgang mit Ausreißern
2. Konvertierung der Daten nach Use Case
3. Umgang mit fehlerhaften Werten
4. Qualitätsverbesserung
5. Merkmalsreduktion

Qualitätsverbesserung

- Entfernung fehlerhafter Messungen
- Glättung von Messwerten (durch Interpolation)

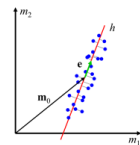


Merkmalsreduktion durch PCA

Datenaufbereitung

Merkmalsreduktion – Transformation durch PCA

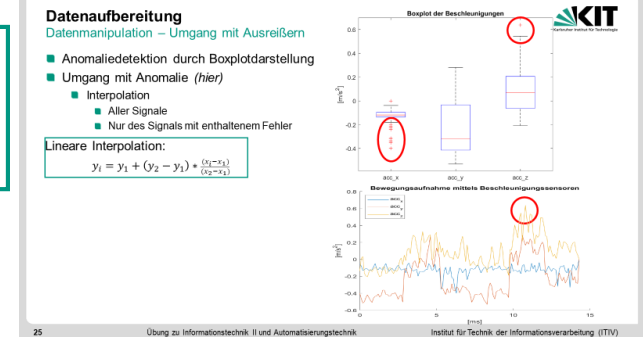
- Durch die Zeitsynchronisation kann der Datensatz inzwischen als $n \times m$ Matrix dargestellt werden, wobei n die jeweiligen Instanzen sind und m die Merkmale
- Der bereinigte und aufbereitete Datensatz kann so als Punktwolke in einem m -dimensionalen Raum dargestellt werden (jeder Punkt ist eine Instanz)
- Ziel der PCA ist es, die Datenpunkte so in einen q -dimensionalen Unterraum zu projizieren, dass dabei möglichst wenig Informationen verloren gehen ($\rightarrow$ Reduktion der Merkmale ohne Informationsverlust)
- Mathematisch: Hauptachsentransformation
 - Minimierung der Korrelation mehrdimensionaler Merkmale durch Überführung in Vektorraum neuer Basis
 - Orthogonale Matrix, bestehend aus Eigenvektoren der Kovarianzmatrix
- Dritter Schritt: Welche Gerade, welche senkrecht zu den ermittelten Hauptkomponenten ist, repräsentiert die Daten weiter mit minimalem quadratischen Fehler?
- Vierter Schritt: Schritt drei so oft wiederholen, bis alle bzw. q Hauptkomponenten bestimmt sind



Umgang mit Ausreißern

Lineare Interpolation:

$$y_i = y_1 + (y_2 - y_1) * \frac{(x_i - x_1)}{(x_2 - x_1)}$$



Konvertierung

- Standardisierung
- Normierung
- Anpassung von Einheiten

Datenaufbereitung

Datenmanipulation - Konvertierung

Standardisierung

- Werte an den benötigten Standard des Data Mining Algorithmus anpassen
- Beispiel: „verständliche“ Labels

Acc _x	Acc _y	Acc _z	Activity
0.33150518	-0.03658496	-0.10886677	SITTING
0.26663115	-0.043309471	-0.14096216	SITTING
0.28784253	-0.050273681	-0.13274829	SITTING
0.22715659	-0.022146672	-0.14521449	LAYING
0.23483919	0.0081011057	-0.14108314	LAYING
0.23820197	-0.0026928607	-0.12149269	LAYING
0.27873663	-0.048279124	-0.12092488	LAYING
0.26374279	-0.02958616	-0.050708835	LAYING
0.31615359	0.0012773598	-0.06545266	WALKING
0.15366105	-0.010077719	-0.04389438	WALKING
0.071035289	-0.015656183	-0.094263452	WALKING
0.33304231	-0.01019258	-0.1220055	WALKING

Anpassung von Einheiten (meist in SI-Einheiten)

Beispiel: hier schon gegeben mit m/s^2 und ms



Normierung

- Min/Max-Normierung
- Z-Score-Normierung
- Dezimal-Skalierung

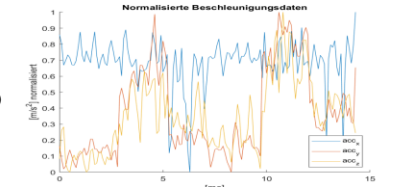
$$x^{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$x^{new} = \frac{x - \mu}{\sigma_x}$$

$$x^{new} = |x| * 10^a, \quad a = \max = i \in \mathbb{Z}, |x| * 10^i < 1$$

$$x^{new} = \log_a x$$

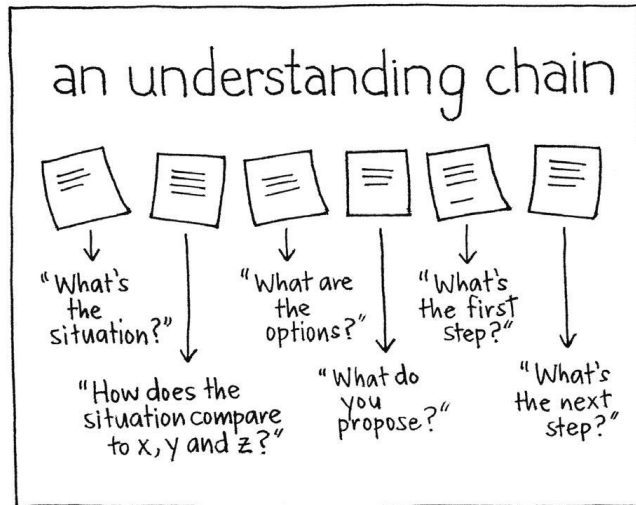
Beispiel: Beschleunigungen auf $[0, 1]$ normieren um Ausreißer durch Visualisierung zu erkennen



INHALT ÜBUNG 6



Data Understanding



Data Preparation



Modeling



Ziele der heutigen Übung



■ Nach der heutigen Übung können Sie....

- ...Ansätze zur Verwaltung und Analyse großer Datenbestände hinsichtlich ihrer Anwendbarkeit und Wirksamkeit einschätzen

1

- ... Merkmale und Eigenschaften von selbstlernenden Algorithmen benennen und abgrenzen

2

- ... Methoden des maschinellen Lernens einordnen, beschreiben und bewerten

3

- ... Modelle berechnen unter Anwendung überwachter maschineller Lernmethoden

4

- ... Verfahren zur Auswahl einer geeigneter Methode beschreiben und anwenden

MODELING

EINFÜHRUNG IN MASCHINELLES LERNEN

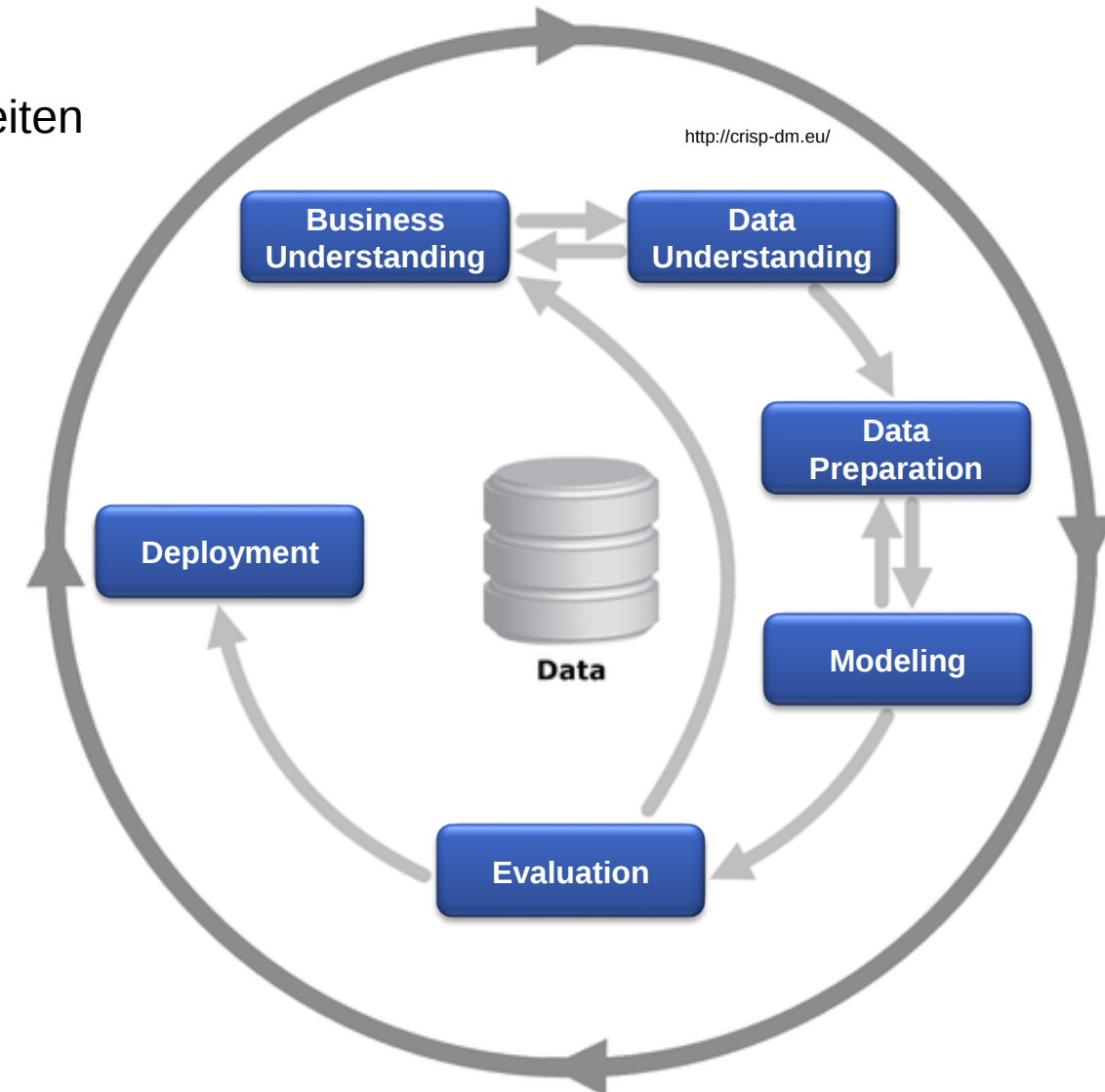
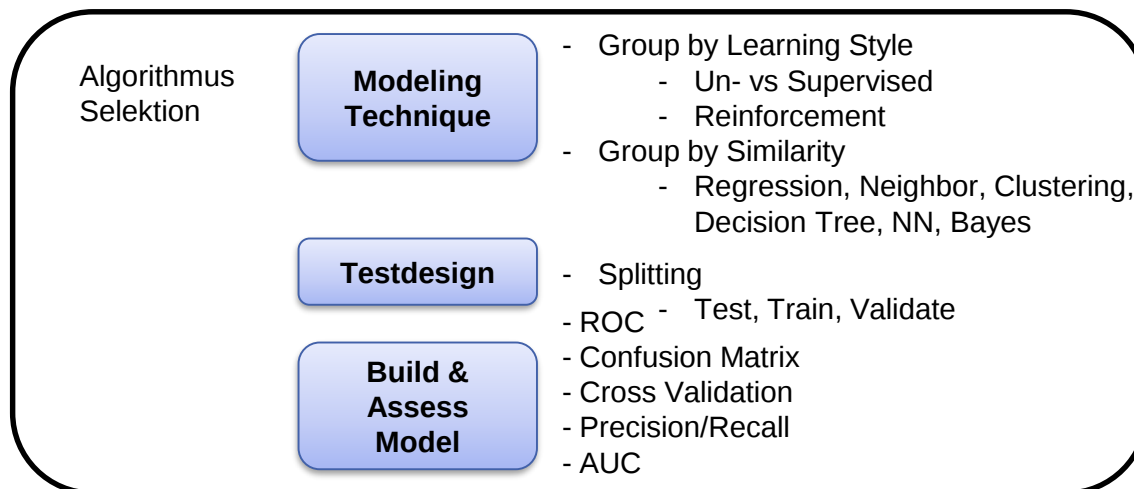


Cross Industry Standard Process for Data Mining (CRISP-DM)

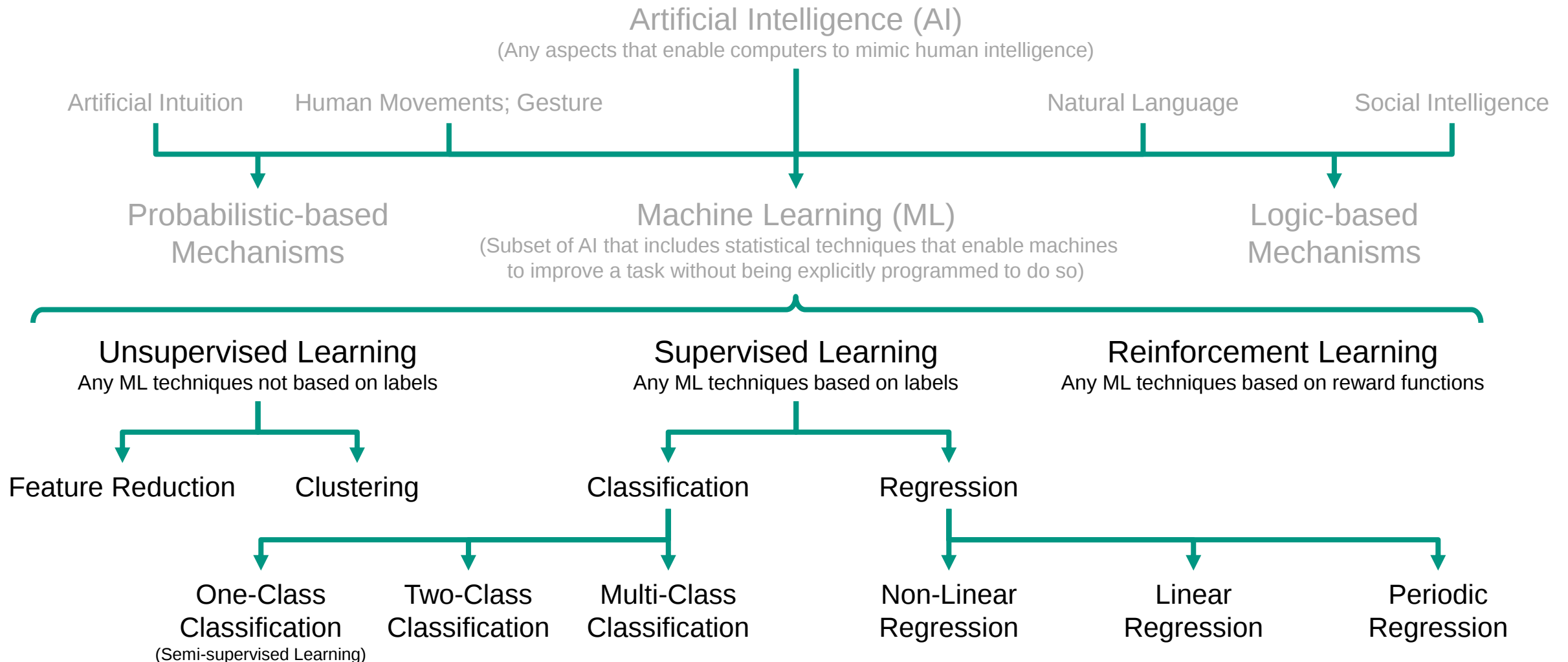
Modeling

■ Modellbildung

- Suche nach Modellen, Mustern oder Gesetzmäßigkeiten in den vorliegenden Daten



Modelbildung – Übersicht über Verfahren



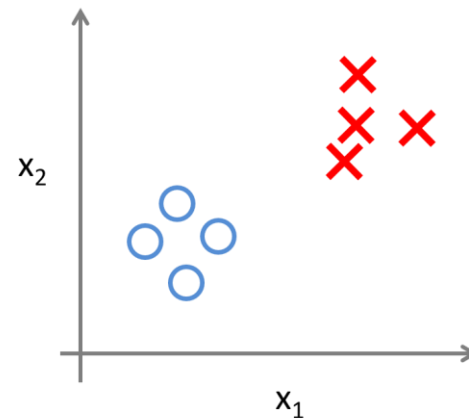
Modeling – Einteilung der Algorithmen

Group by Learning Style

Supervised Learning

- gelabelte Daten
- Lernen/Vorhersagen von Output aus Input-Daten
- Herausforderung:
 - extrapolieren
 - generalisieren
- Beispiele
 - Klassifizierung
 - Regression

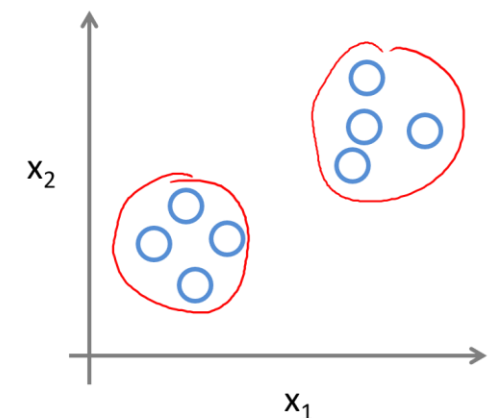
Supervised Learning



Unsupervised Learning

- ungelabelte Daten
- Auffinden von versteckten Strukturen in Daten
- Herausforderung:
 - Subjektiver als SL
 - Validierung
- Beispiele
 - Clustering
 - Dimensionsreduktion

Unsupervised Learning



SUPERVISED LEARNING REGRESSION



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

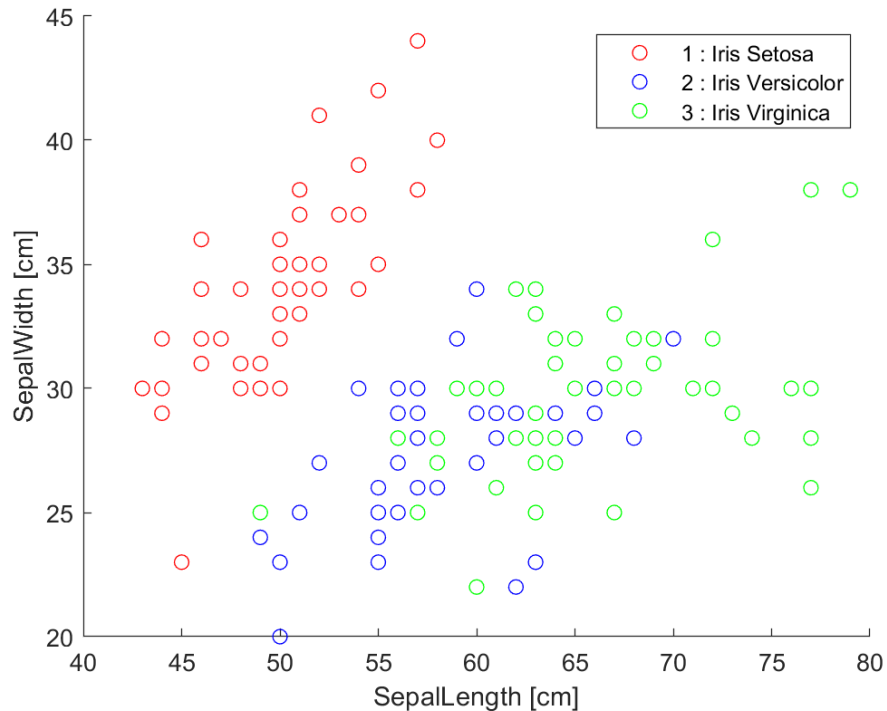
Regression

Berechnung der Linearen Regression

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

$$S_{xx} = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2$$

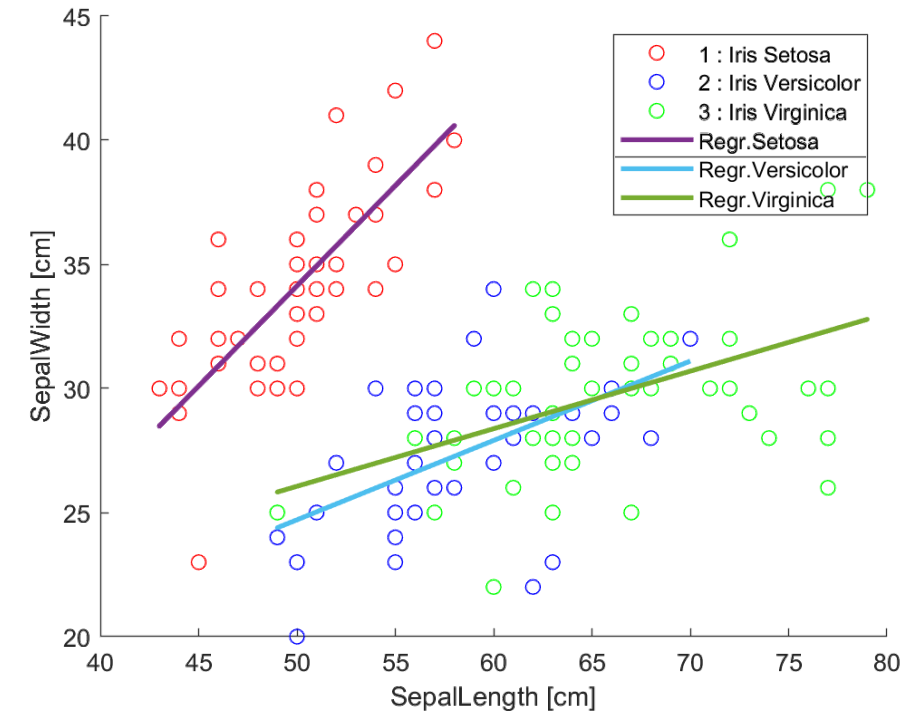
$$S_{xy} = ((x_1 - \bar{x}) * (y_1 - \bar{y})) + ((x_2 - \bar{x}) * (y_2 - \bar{y})) + \dots + ((x_n - \bar{x}) * (y_n - \bar{y}))$$



Regressionsgerade:
 $y = 0.8072x - 6.2301$

Regressionsgerade:
 $y = 0.3197x + 8.7215$

Regressionsgerade:
 $y = 0.2319x + 14.4631$



Regression

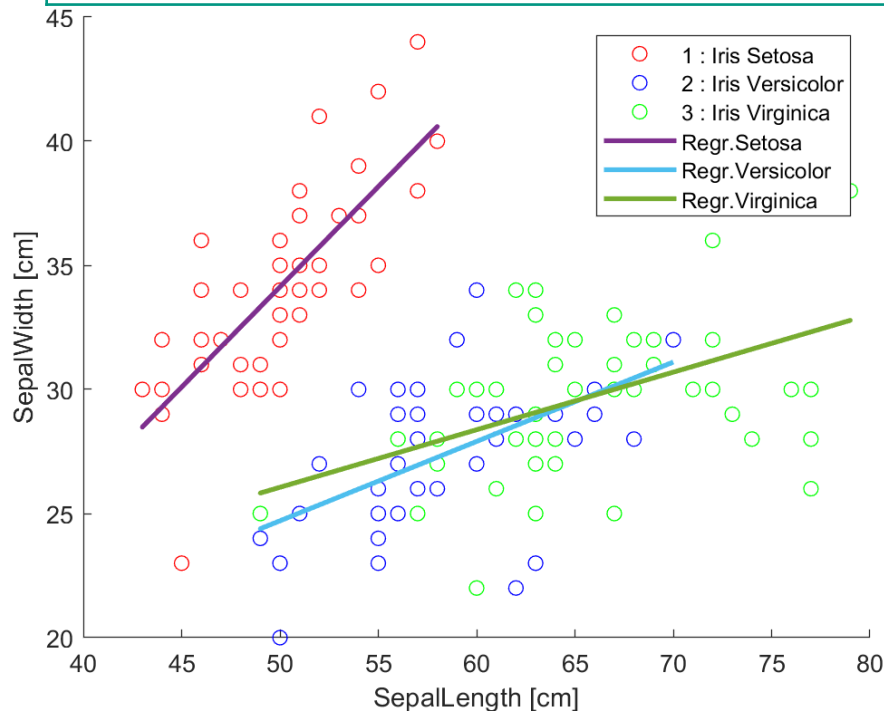
Quadratische Regression

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

$$a = \frac{(\overline{y_i x_i^2} - \overline{y_i} \cdot \overline{x_i^2}) \cdot (\overline{x_i^2} - (\overline{x_i})^2) - (\overline{y_i x_i} - \overline{y_i} \cdot \overline{x_i}) \cdot (\overline{x_i^3} - \overline{x_i} \cdot \overline{x_i^2})}{(\overline{x_i^4} - (\overline{x_i^2})^2) \cdot (\overline{x_i^2} - (\overline{x_i})^2) - (\overline{x_i^3} - \overline{x_i} \cdot \overline{x_i^2})^2}$$

$$b = \frac{\overline{y_i x_i} - \overline{y_i} \cdot \overline{x_i} - a \cdot (\overline{x_i^3} - \overline{x_i} \cdot \overline{x_i^2})}{\overline{x_i^2} - (\overline{x_i})^2}$$

$$c = \overline{y_i} - a \cdot \overline{x_i^2} - b \cdot \overline{x_i}$$



Regressionskurve:

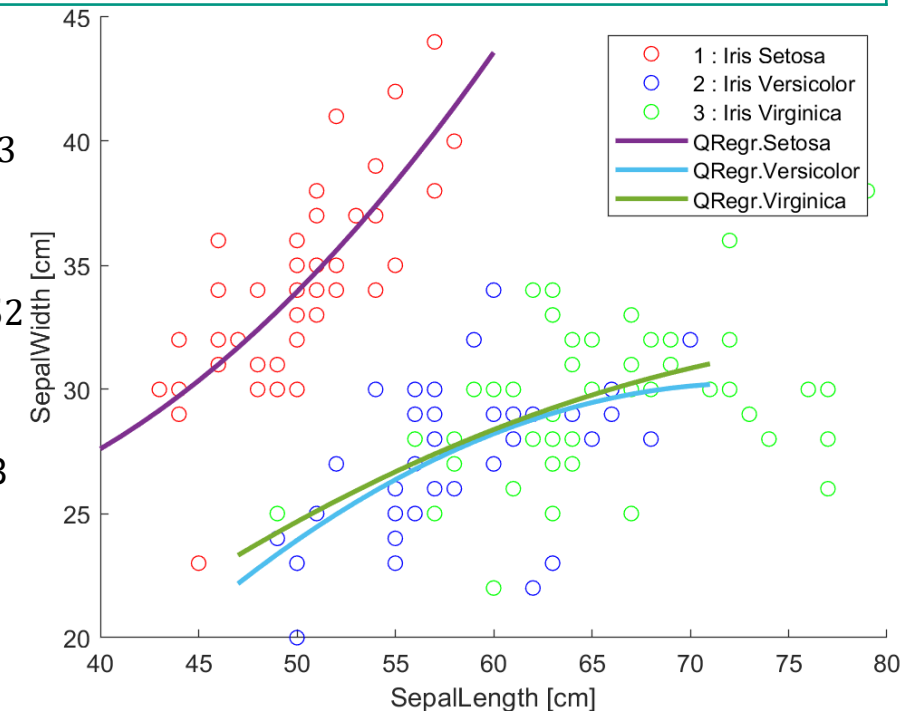
$$y = 0.0166x^2 - 0.8598x + 35.4623$$

Regressionskurve:

$$y = -0.0118x^2 - 1.7305x - 33.0252$$

Regressionskurve:

$$y = -0.0062x^2 + 1.0516x - 12.4383$$

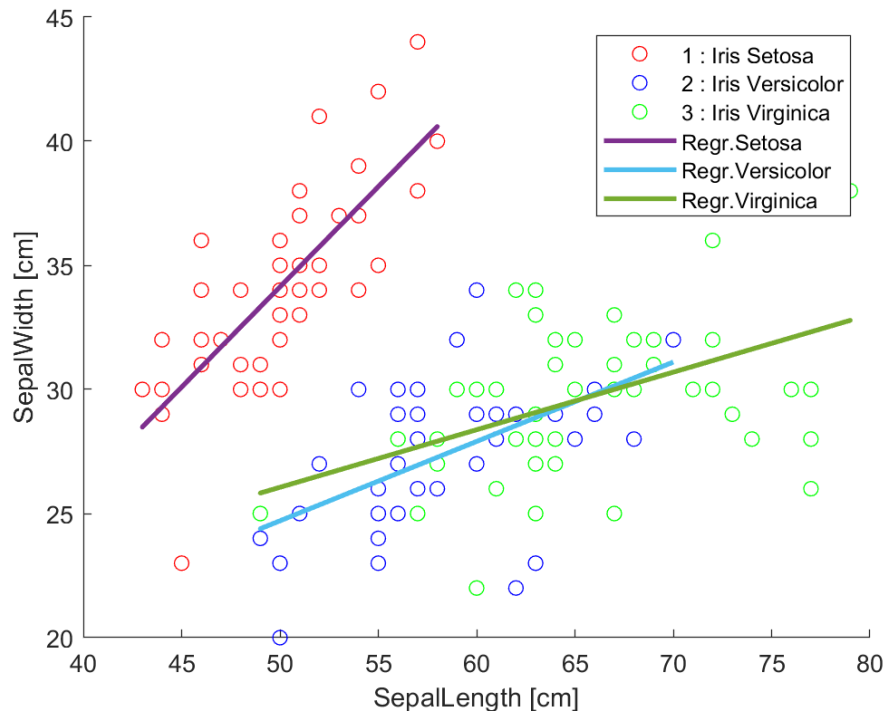


Regression

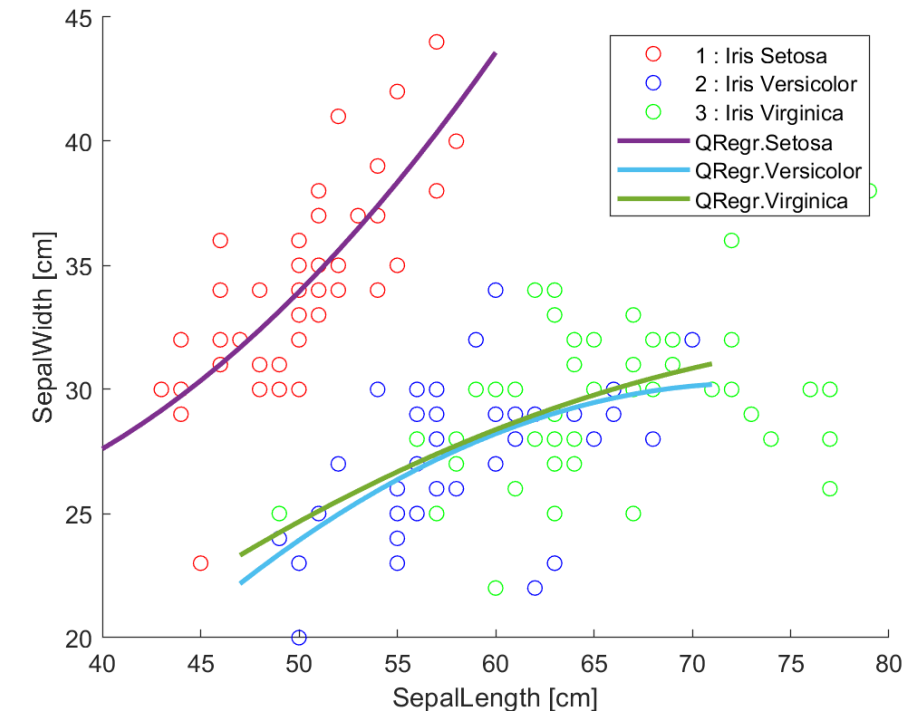
Methodenauswahl

Es sei ein neuer Datenpunkt für die „Iris-Analyse“ gegeben
Wie wird nun entschieden, welche Methode sich am besten eignet um eine Vorhersage zu treffen, zu welcher Klasse der neue Punkt gehört?

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	51	35	14		2 Iris-setosa
2	49	30	14		2 Iris-setosa
3	47	32	13		2 Iris-setosa
4	46	31	15		2 Iris-setosa
5	50	36	14		2 Iris-setosa
51	70	32	47		14 Iris-versicolor
52	64	32	45		15 Iris-versicolor
53	69	31	49		15 Iris-versicolor
54	55	23	40		13 Iris-versicolor
55	65	28	46		15 Iris-versicolor
152	52	50	30		6 ?????

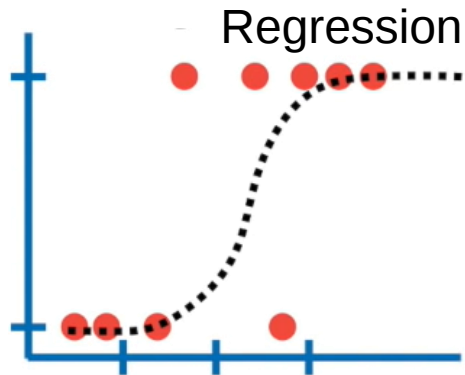


k-fold Cross Validation

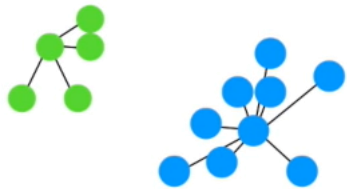


Supervised Learning

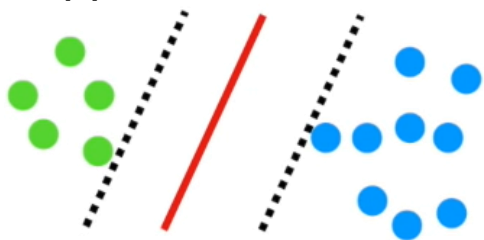
Methodenauswahl – k-fold Cross Validation



K-nearest neighbors



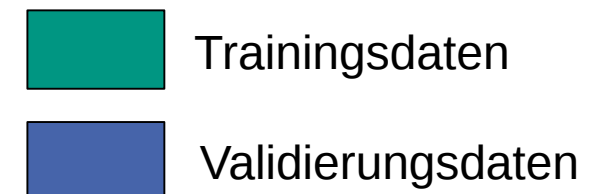
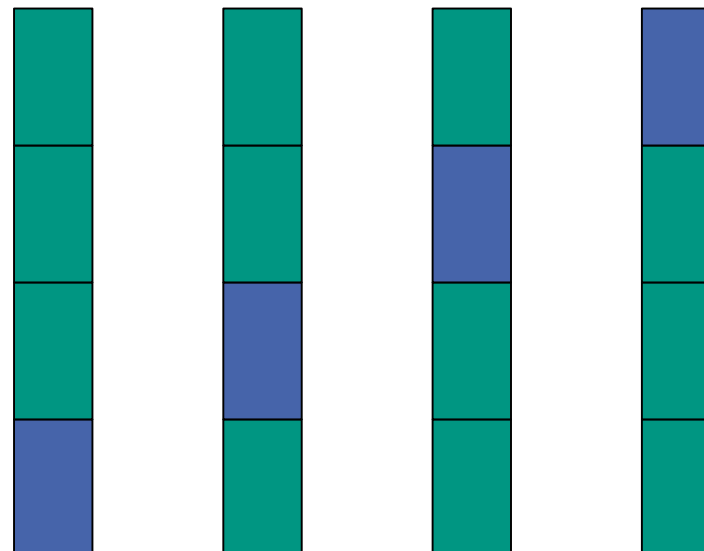
Support Vector machines



■ Kompletten Datensatz teilen in 70% Trainings- und 30% Testdaten

1. Die Trainingsdaten werden in k Teile unterteilt
2. Auf Basis von k-1 Teilen wird das Modell berechnet und im Anschluss mit dem verbleibenden Teil validiert.
3. Wiederholen bis alle Teile einmalig der Validierung dienen
4. MSE berechnen und merken
5. Wiederholung der Punkte 1.-4. für jede Methode
6. Auswahl der Methode mit dem geringsten MSE

■ Modellbildung auf Basis der gesamten 70% anhand der ausgewählten Methode



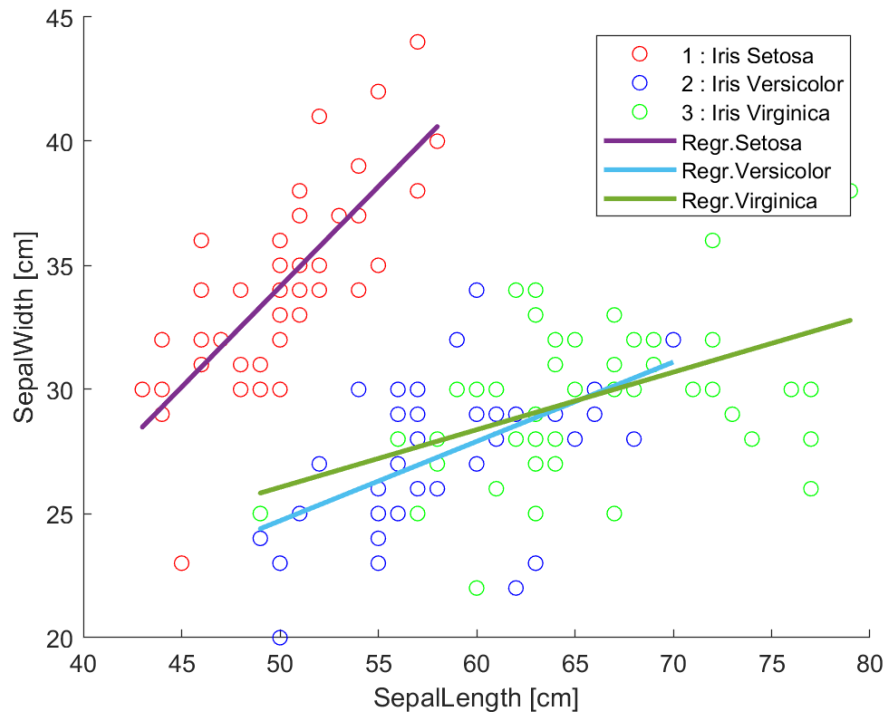
Supervised Learning

Methodenauswahl – k-fold Cross Validation

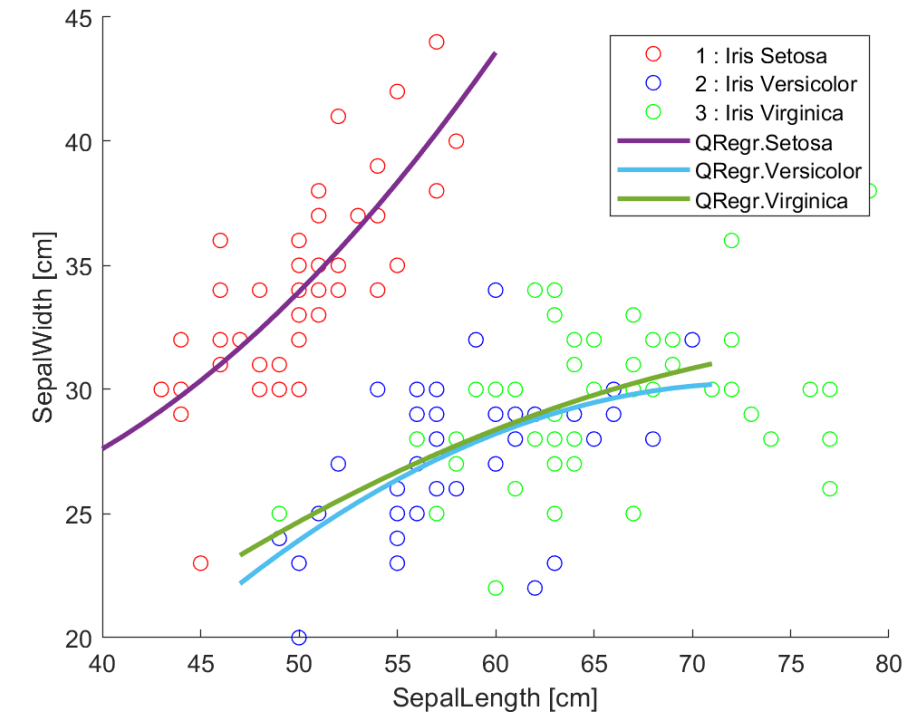
Mean Squared Error

$$\text{MSE} = \frac{1}{n} \sum (y(x) - y^*(x))^2$$

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	51	35	14		2 Iris-setosa
2	49	30	14		2 Iris-setosa
3	47	32	13		2 Iris-setosa
4	46	31	15		2 Iris-setosa
5	50	36	14		2 Iris-setosa
51	70	32	47		14 Iris-versicolor
52	64	32	45		15 Iris-versicolor
53	69	31	49		15 Iris-versicolor
54	55	23	40		13 Iris-versicolor
55	65	28	46		15 Iris-versicolor
152	52	50	30		6 ?????



Iris Setosa	
Regr.art	MSE
L. Regr	6.5554
Q. Regr.	6.2262



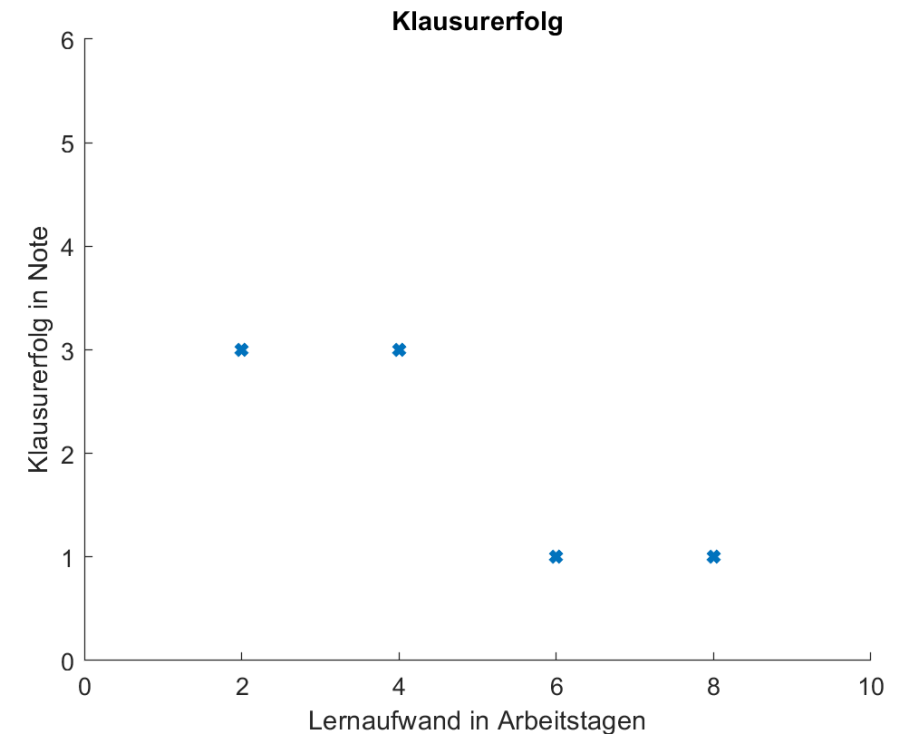
Regression

Zwischenübung



Es ist der Lernaufwand x_i von vier Personen, sowie der Klausurerfolg y_i (in Note) gegeben. Berechnen Sie die Regressionsgerade $y = \frac{s_{xy}}{s_{xx}}x + \bar{y} - \frac{s_{xy}}{s_{xx}}\bar{x}$

x_i	y_i
8	1
4	3
6	1
2	3
$\bar{x}=5$	$\bar{y}=2$



Regression

Zwischenübung - Lsg

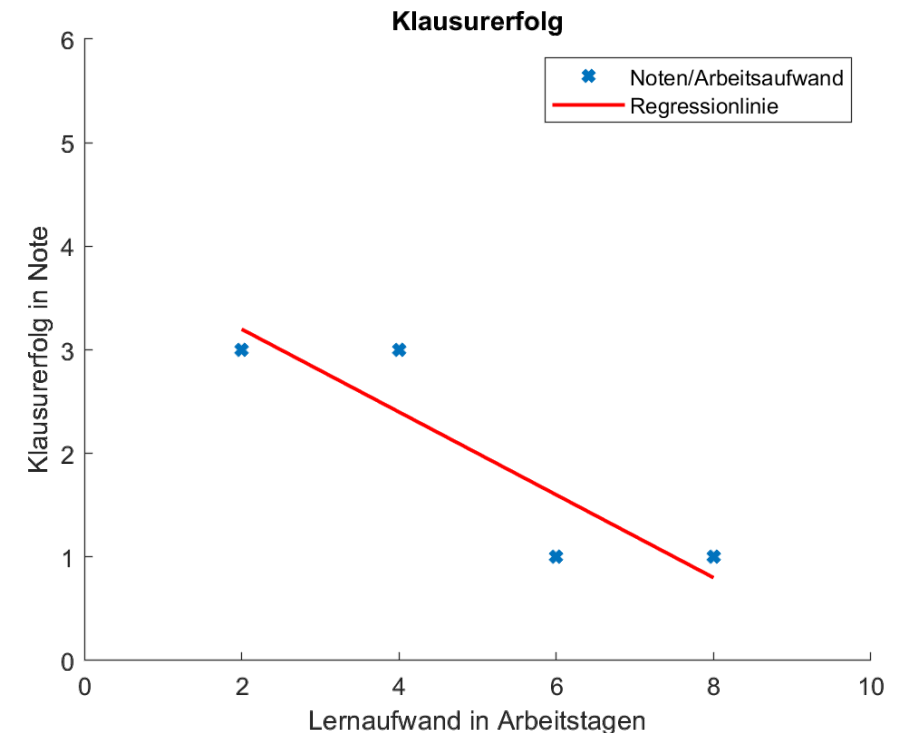
Es ist der Lernaufwand x_i von vier Personen, sowie der Klausurerfolg y_i (in Note) gegeben. Berechnen Sie die Regressionsgerade $y = \frac{s_{xy}}{s_{xx}} x + \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
8	1	3	-1	9	1	-3
4	3	-1	1	1	1	-1
6	1	1	-1	1	1	-1
2	3	-3	1	9	1	-3
$\bar{x}=5$	$\bar{y}=2$			$\sum 20$	$\sum 4$	$\sum -8$

$$y_{regr} = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} x + \bar{y} - \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \bar{x}$$

$$y_{regr} = -2/5 x + 4$$

$$MSE = 1/n \sum (y_{regr}(x_i) - y_i)^2 \text{ (hier: MSE=0.2)}$$



- Supervised Learning
 - Regression



Fragen?

SUPERVISED LEARNING KLASSIFIKATION



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

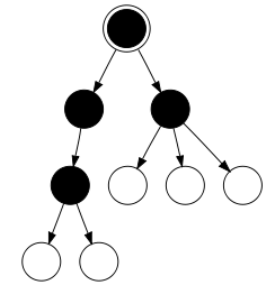
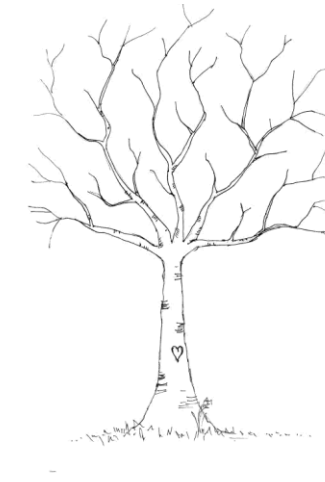
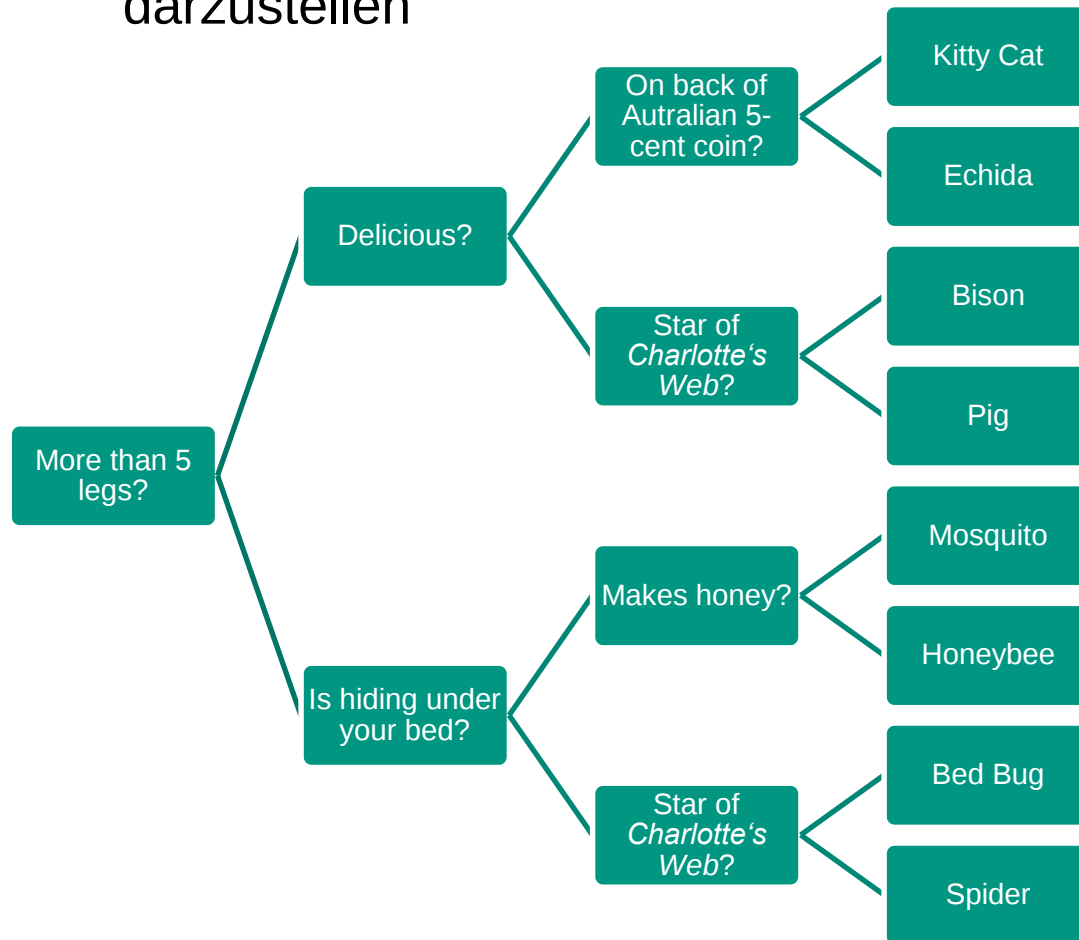
Klassifikationsprobleme

- 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

Klassifikation

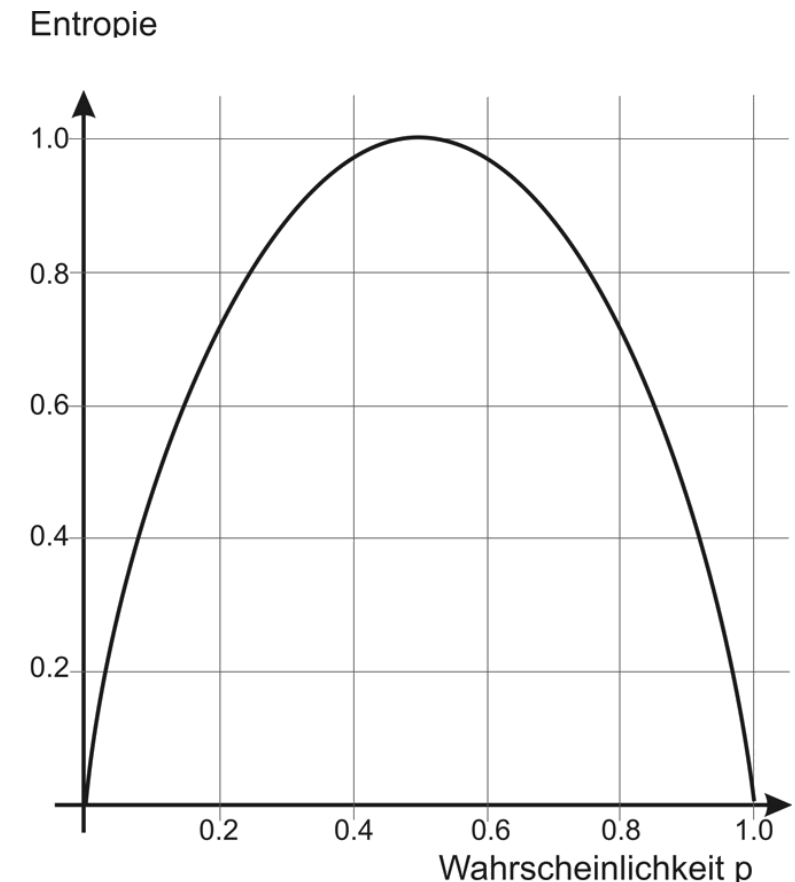
Decision Tree

- Ein Decision Tree verwendet eine Baumstruktur, um eine Reihe möglicher Entscheidungspfade und das Ergebnis für jeden Pfad darzustellen



Aufbau?

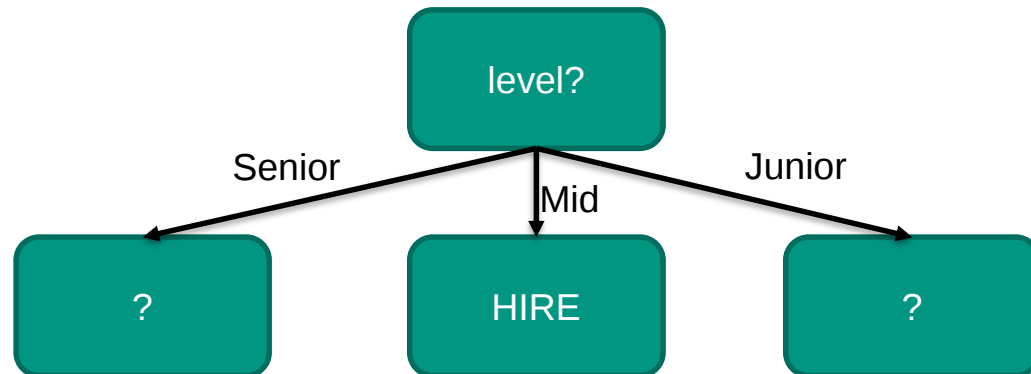
- Für den Entwurf des Decision Tree muss entschieden werden
 - Welche „Fragen“ werden an die Daten gestellt (*Auswahl der genutzten Attribute*)
 - In welcher Reihenfolge werden die „Fragen“ an die Daten gestellt (*Reihenfolge der Attributsabfrage*)
- Abhilfe schafft Entropie
 - Entropie $H(S)$: Aussage über Informationsgehalt
$$H(S) = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_n \log p_n$$
mit p_i als Wahrscheinlich des Ereignis i
 - Je eher p_i gegen 0 oder 1 geht desto $H(S_i) \rightarrow 0$



Decision Tree Entwurf

„Should you hire the candidate?“

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
level	Senior	Senior	Mid	Junior	Junior	Junior	Mid	Senior	Senior	Junior	Senior	Mid	Mid	Junior
language	Java	Java	Python	Python	R	R	R	Python	R	Python	Python	Python	Java	Python
tweets	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
PhD	No	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes
Hire?	False	False	True	True	True	False	True	False	True	True	True	True	True	False



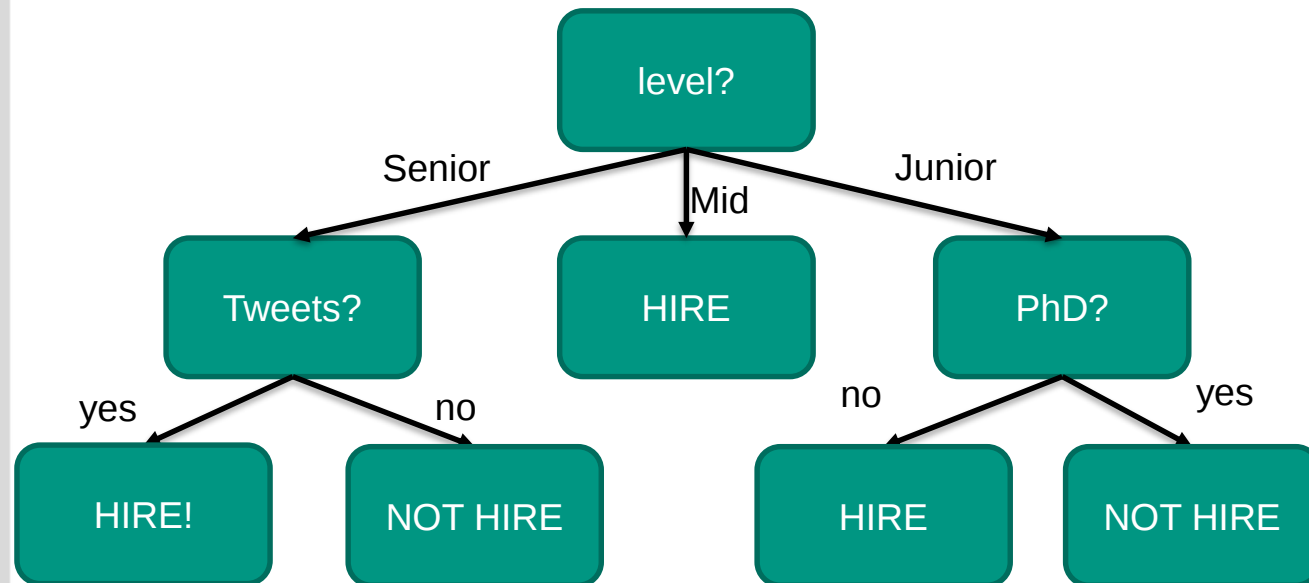
Entropie:

level:	0.693536
language:	0.860131
tweets:	0.788450
PhD:	0.892158

Decision Tree Entwurf

„Should you hire the candidate?“

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
level	Senior	Senior	Mid	Junior	Junior	Junior	Mid	Senior	Senior	Junior	Senior	Mid	Mid	Junior
language	Java	Java	Python	Python	R	R	R	Python	R	Python	Python	Python	Java	Python
tweets	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
PhD	No	Yes	No	No	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes
Hire?	False	False	True	True	True	False	True	False	True	True	True	True	True	False



Entropie:

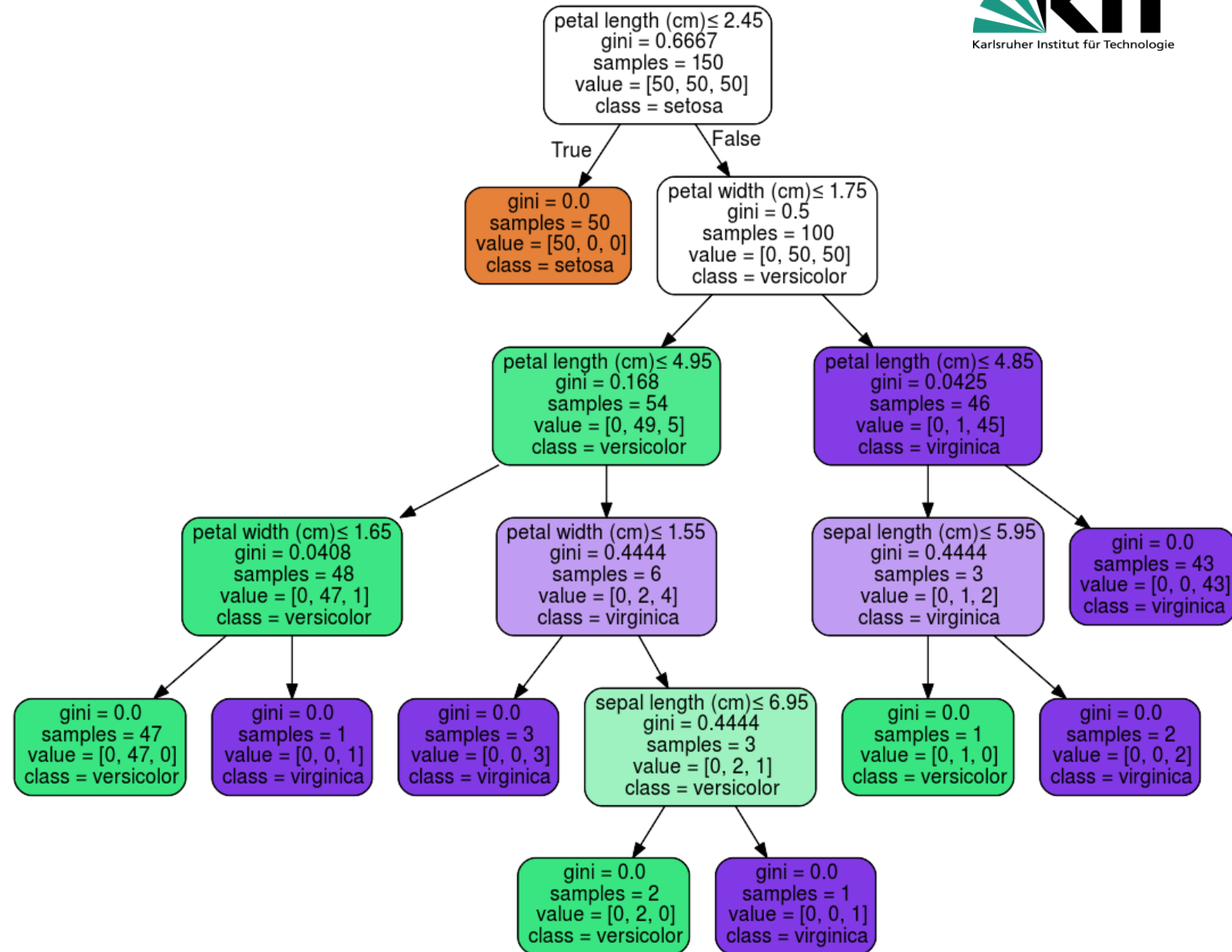
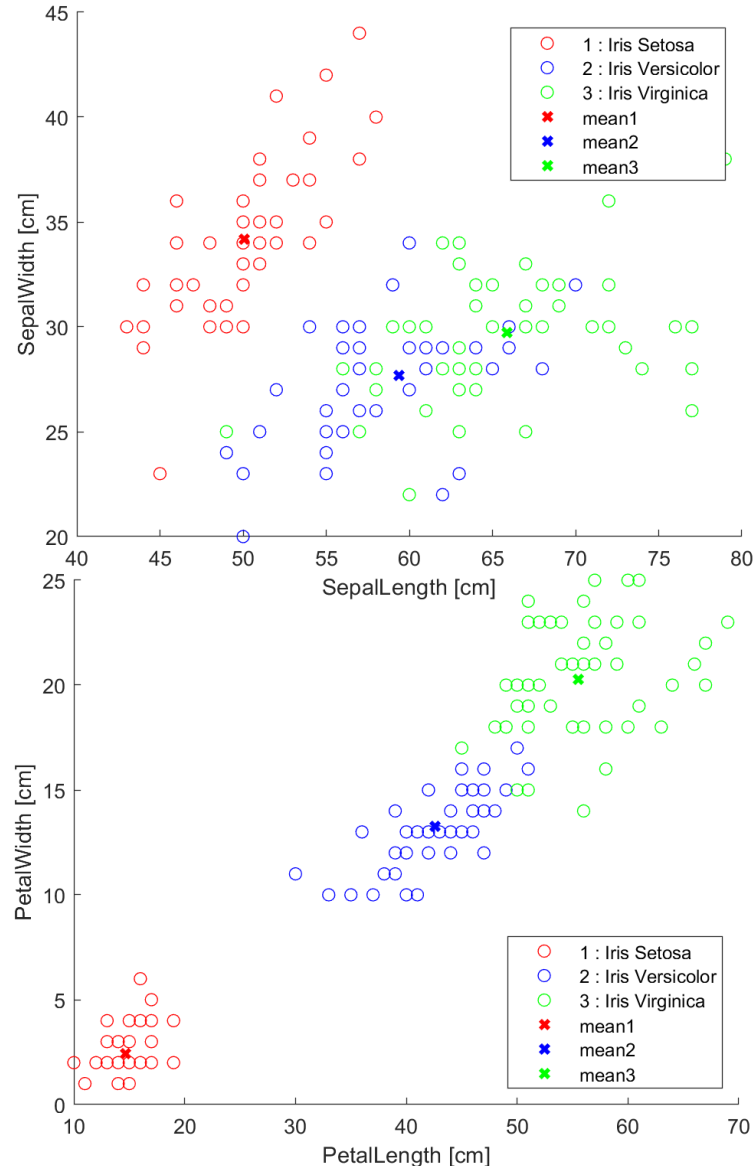
level: 0.693536
 language: 0.860131
 tweets: 0.788450
 PhD: 0.892158

Entropie „Senior“

language: 0.4
 tweets: 0.0
 PhD: 0.950977

Klassifikation

Decision Tree – Iris Datensatz



Decision Tree und Entropie

Zwischenübung

Gegeben seien folgende Datenpunkte:

Label	x	x	x	o	o	o	o
x	1	4	4	1	3	3	4
y	1	1	3	3	2	4	4

Ziel: Erhalt eines Entscheidungsbaumes für x und o

Vorgehen:

1. Skizzieren Sie die Datenpunkte
2. Fügen Sie zwei Trennlinien ein, welche die Datenpunkte möglichst „gut“ trennen
3. Entscheiden und begründen Sie rechnerisch nach welche Trennlinie die erste Ebene des Entscheidungsbaumes bilden sollte



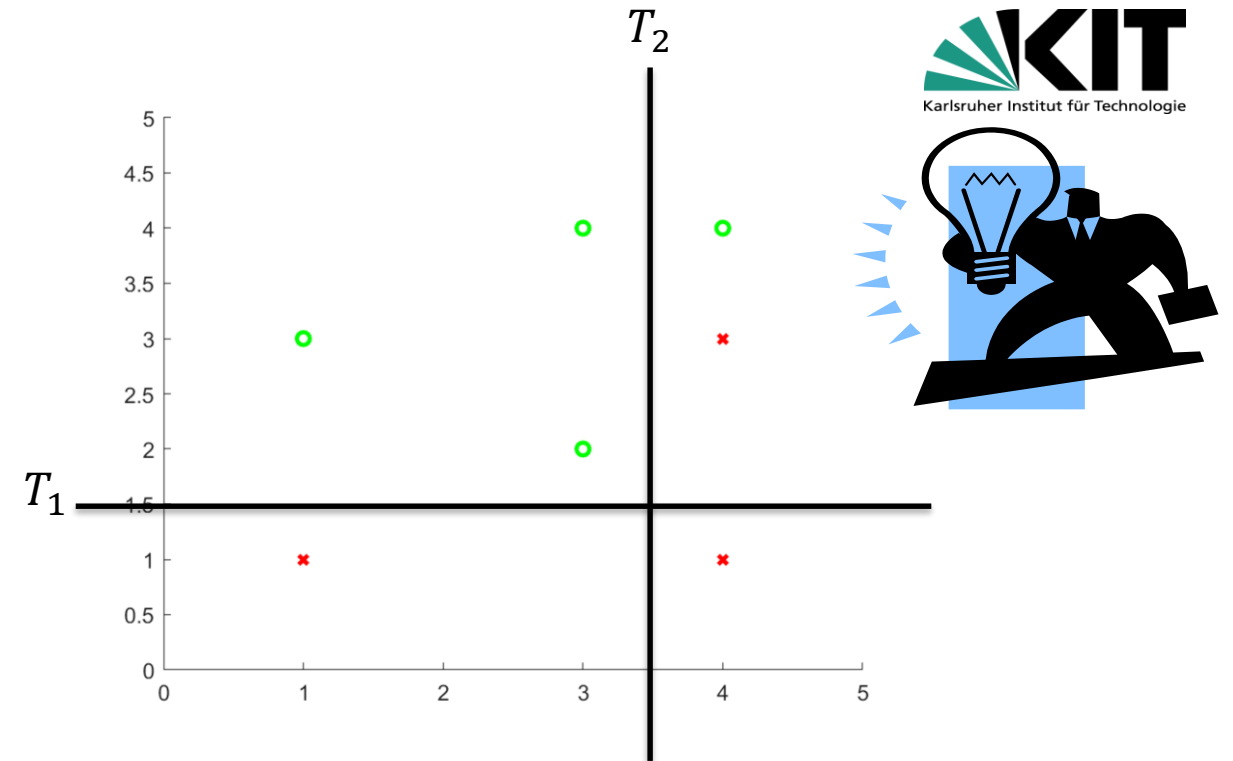
Decision Tree und Entropie

Zwischenübung -Lsg

Gegeben seien folgende Datenpunkte:

Label	x	x	x	o	o	o	o
x	1	4	4	1	3	3	4
y	1	1	3	3	2	4	4

1. Skizzieren Sie die Datenpunkte



Decision Tree und Entropie

Zwischenübung -Lsg

Gegeben seien folgende Datenpunkte:

Label	x	x	x	o	o	o	o
x	1	4	4	1	3	3	4
y	1	1	3	3	2	4	4

3. Entscheiden und begründen Sie rechnerisch nach welcher Trennlinie die erste Ebene des Entscheidungsbaumes gebildet werden sollte

Einzel-Entropie: $H(S) = -p_1 \log p_1 - p_2 \log p_2$

Gesamtentropie: $H_{Ges}(X) = \sum_{c \in X} G_i(c) * H_i(c)$, mit G_i als Gewichtungen der Einzel-Entropien

$$H(T_1) = \frac{2}{7} H(T_{1unten}) + \frac{5}{7} H(T_{1oben}) = 0,51$$

$$H(T_{1unten}) = -1 * \log_2 1 - 0 * \log_2 0 = 0$$

$$H(T_{1oben}) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} * \log_2 \frac{4}{5} = 0,72$$

$$H(T_1) = \frac{2}{7} H(T_{1unten}) + \frac{5}{7} H(T_{1oben}) = 0,51$$

$$H(T_{1unten}) = -1 * \log_2 1 - 0 * \log_2 0 = 0$$

$$H(T_{1oben}) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} * \log_2 \frac{4}{5} = 0,72$$

$$H(T_2) = \frac{4}{7} H(T_{2links}) + \frac{3}{7} H(T_{2rechts}) = 0,85$$

$$H(T_{2links}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0,81$$

$$H(T_{2rechts}) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0,91$$



Decision Tree und Entropie

Zwischenübung -Lsg

Gegeben seien folgende Datenpunkte:

Label	x	x	x	o	o	o	o
x	1	4	5	1	3	3	4
y	1	1	3	3	2	4	4

3. Entscheiden und begründen Sie rechnerisch nach welcher Trennlinie die erste Ebene des Entscheidungsbaumes bilden sollte

Einzel-Entropie: $H(S) = -p_1 \log p_1 - p_2 \log p_2$

Gesamtentropie: $H_{Ges}(X) = \sum_{c \in X} G_i(c) * H_i(c)$, mit G_i als Gewichtungen der Einzel-Entropien

$$H(T_1) = \frac{2}{7} H(T_{1unten}) + \frac{5}{7} H(T_{1oben}) = 0,51$$

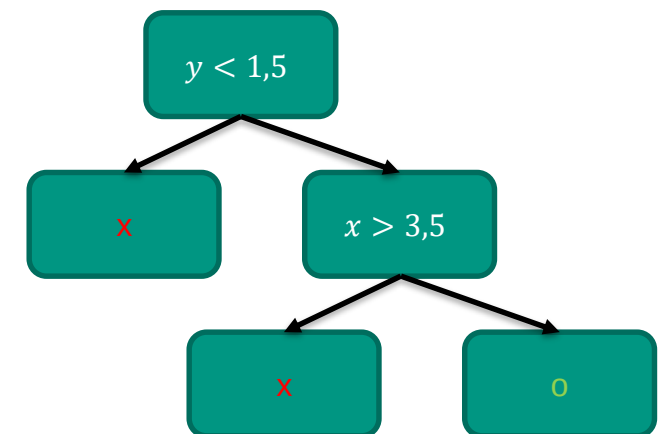
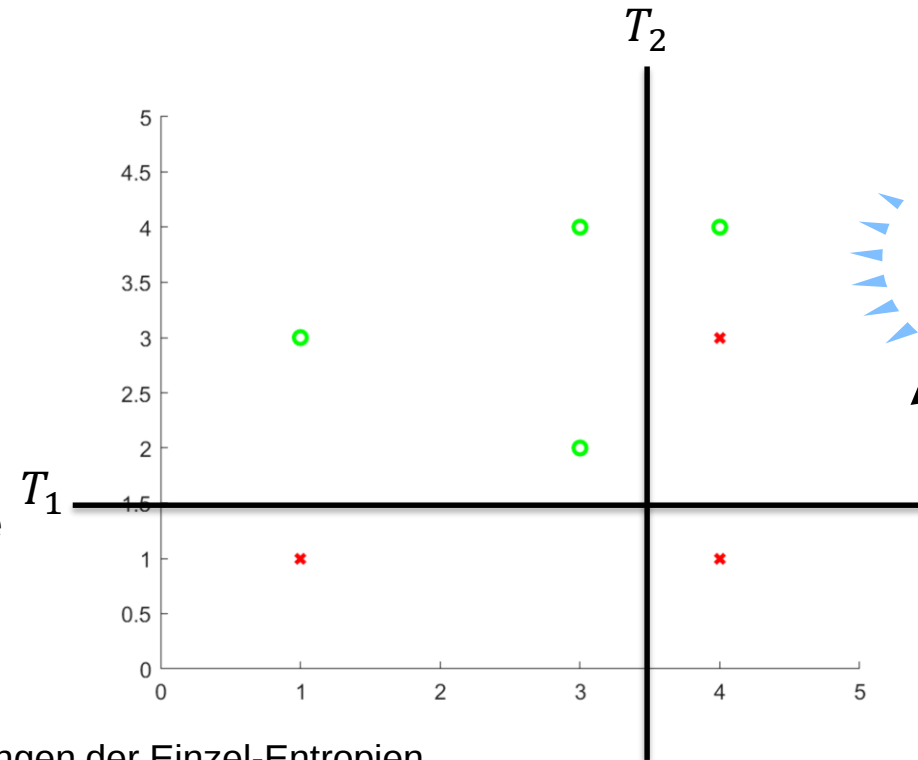
$$H(T_{1unten}) = -1 * \log_2 1 - 0 * \log_2 0 = 0$$

$$H(T_{1oben}) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} * \log_2 \frac{4}{5} = 0,72$$

$$H(T_2) = \frac{4}{7} H(T_{2links}) + \frac{3}{7} H(T_{2rechts}) = 0,85$$

$$H(T_{2links}) = -\frac{1}{4} * \log_2 \frac{1}{4} - \frac{3}{4} * \log_2 \frac{3}{4} = 0,81$$

$$H(T_{2rechts}) = -\frac{1}{3} * \log_2 \frac{1}{3} - \frac{2}{3} * \log_2 \frac{2}{3} = 0,91$$



- Supervised Learning
 - Decision Tree

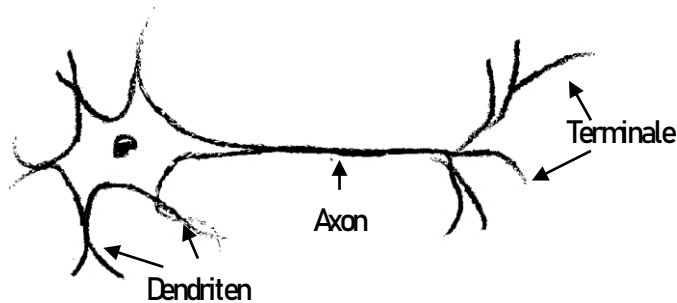


Fragen?

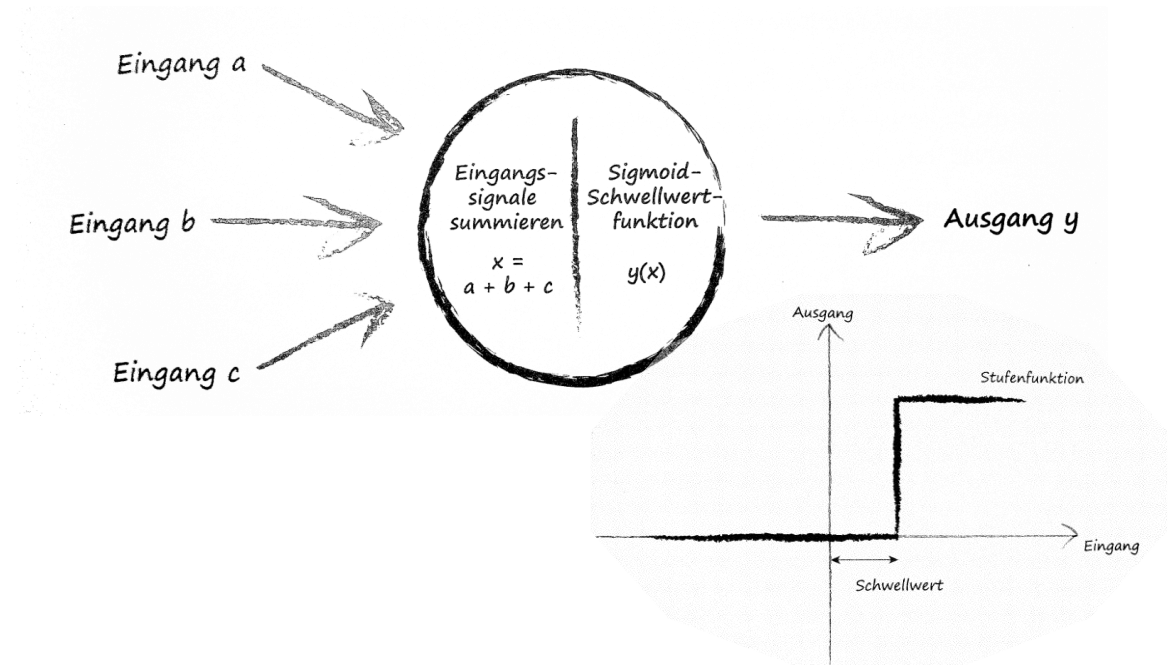
Klassifikation

Künstliche Neuronale Netze - Einführung

- Die Idee von Künstlichen Neuronalen Netzen beruht auf den Neuronennetzen des menschlichen Gehirns und finden Anwendung in zahlreichen Gebieten, wie z.B. Handschrifterkennung, Gesichtsdetektion, uvm.
- Aufteilung in zwei Teilbereiche
 - KNN, die modelliert werden um menschliches Verhalten nachzuahmen/ zu verstehen
 - KNN, modelliert um ein konkretes Anwendungsproblem aus Bereichen, wie beispielsweise der Statistik, zu lösen



- KNN bestehen aus
 - Input-Units: Signale von der „Außenwelt“
 - Hidden-Units: interne Repräsentation der „Außenwelt“, Eingang berechnet sich aus Übertragungsfunktion, Ausgang wird mit Aktivierungsfunktion belegt
 - Output-Units: Signale an die „Außenwelt“

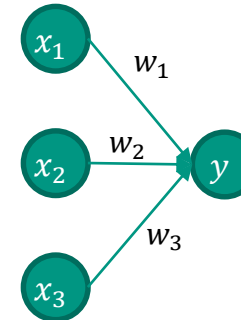


Künstliche Neuronale Netze

Übertragungs- & Aktivierungsfunktionen

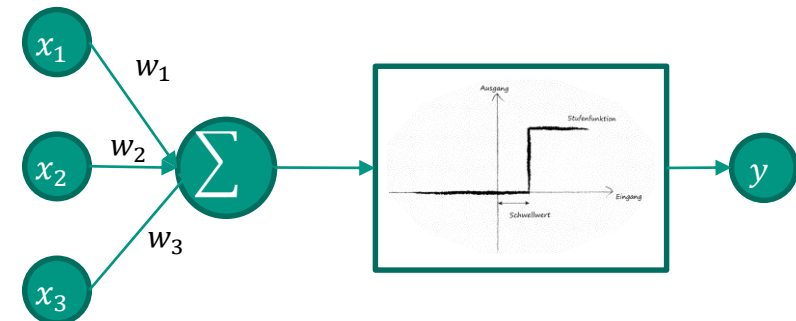
- Perceptron:
einzelnes künstliches Neuron mit anpassbaren Gewichten und Schwellenwerten
- Übertragungsfunktion Perceptron, ohne Schwellenwert:

$$y = \sum_{i=1}^n w_i x_i$$



- Übertragungsfunktion Perceptron, mit Schwellenwert:

$$y = \phi\left(\sum_{i=1}^n w_i x_i\right)$$



Künstliche Neuronale Netze

Übertragungs- & Aktivierungsfunktionen

- Aktivierungsfunktion stellt einen Hyperparameter dar
 - Hyperparameter: „ein Parameter, der zur Steuerung des Trainingsalgorithmus verwendet wird und dessen Wert im Gegensatz zu anderen Parametern nicht im eigentlichen Training des Modells gelernt wird“.
 - Aktivierungsfunktion: „eine Funktion, die das Eingangssignal übernimmt, ein Ausgangssignal generiert und dabei aber eine Art Schwellwert berücksichtigt“

- Verschiedene Arten von Aktivierungsfunktionen:

- Stufenfunktion

$$y = \begin{cases} 0: x < 0 \\ 1: x \geq 0 \end{cases}$$

- Sigmoid-Funktion

$$y = \frac{1}{1+e^{-x}}$$

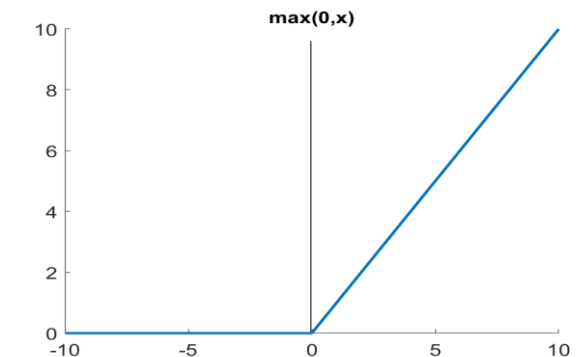
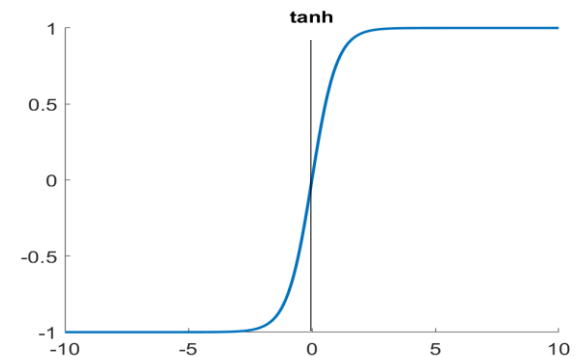
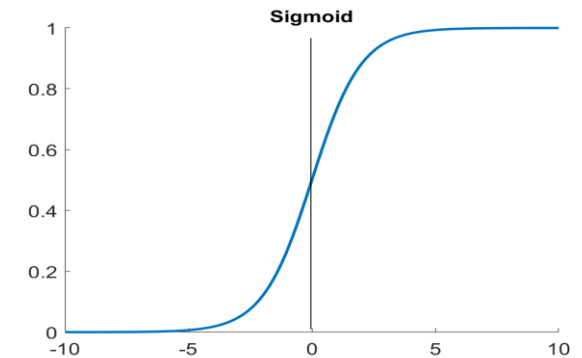
- Tangens-Hyperbolicus-Funktion

$$y = \tanh(x)$$

- Maximum-Funktion

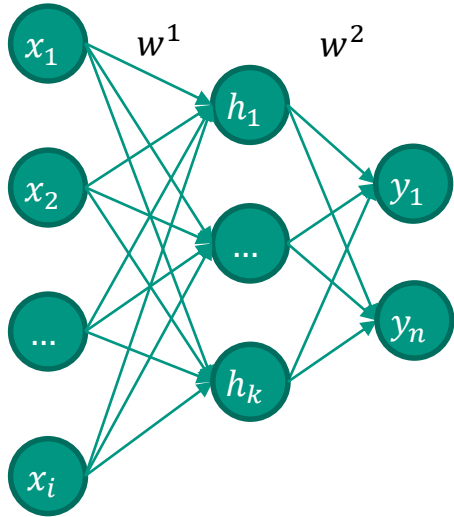
$$y = \max(0, x)$$

- uvm



Künstliche Neuronale Netze

Feedforward NN



$$\bar{x} = [x_1 \ x_2 \ \dots x_i]$$

$$W^1 = \begin{bmatrix} w_{11}^1 & \dots & w_{1k}^1 \\ \dots & \dots & \dots \\ w_{i1}^1 & \dots & w_{ik}^1 \end{bmatrix}$$

$$\bar{h}' = [h'_1 \ h' \ \dots h'_k] = \bar{x} * W^1 = [x_1 \ x_2 \ \dots x_i] * \begin{bmatrix} w_{11}^1 & \dots & w_{1k}^1 \\ \dots & \dots & \dots \\ w_{i1}^1 & \dots & w_{ik}^1 \end{bmatrix}$$

$$\bar{h} = \phi(\bar{h}') = \phi(\bar{x} * W^1), \text{ mit } \phi \text{ als Aktivierungsfunktion}$$

$$y = [y_1 \ \dots y_n] = \bar{h} * W^2 = \phi([x_1 \ x_2 \ \dots x_i] * \begin{bmatrix} w_{11}^1 & \dots & w_{1k}^1 \\ \dots & \dots & \dots \\ w_{i1}^1 & \dots & w_{ik}^1 \end{bmatrix}) * \begin{bmatrix} w_{11}^2 & \dots & w_{1n}^2 \\ \dots & \dots & \dots \\ w_{k1}^2 & \dots & w_{kn}^2 \end{bmatrix}$$

Künstliche Neuronale Netze

CNN - Faltung

Originalmatrix Kernel Resultierende Matrix

7	5	8
5	3	6
2	0	1

×

-1	1
1	0

=

3	6
0	3

$$\begin{array}{rcl} 7*(-1)+5*1+5*1+3*0 & = & 3 \\ \hline 5*(-1)+8*1+3*1+6*0 & = & 6 \\ \hline 5*(-1)+3*1+2*1+0*0 & = & 0 \\ \hline 3*(-1)+6*1+0*1+1*0 & = & 3 \end{array}$$

■ Faltung

Kernelmatrix wird mit vordefinierter Schrittgröße („stride“) über die Originalmatrix geschoben

■ Padding

Je nach Schrittgröße kann die Kernelmatrix über den Rand hinaus geschoben werden. Der Umgang mit dieser Situation wird „Padding“ genannt.

VALID-Padding: kein Überschreiten der Ränder

SAME-Padding: z.B. mit zero-Padding: Der Originalmatrix wird außerhalb der Ränder „0“ zugewiesen“

- Supervised Learning
 - Neuronale Netze

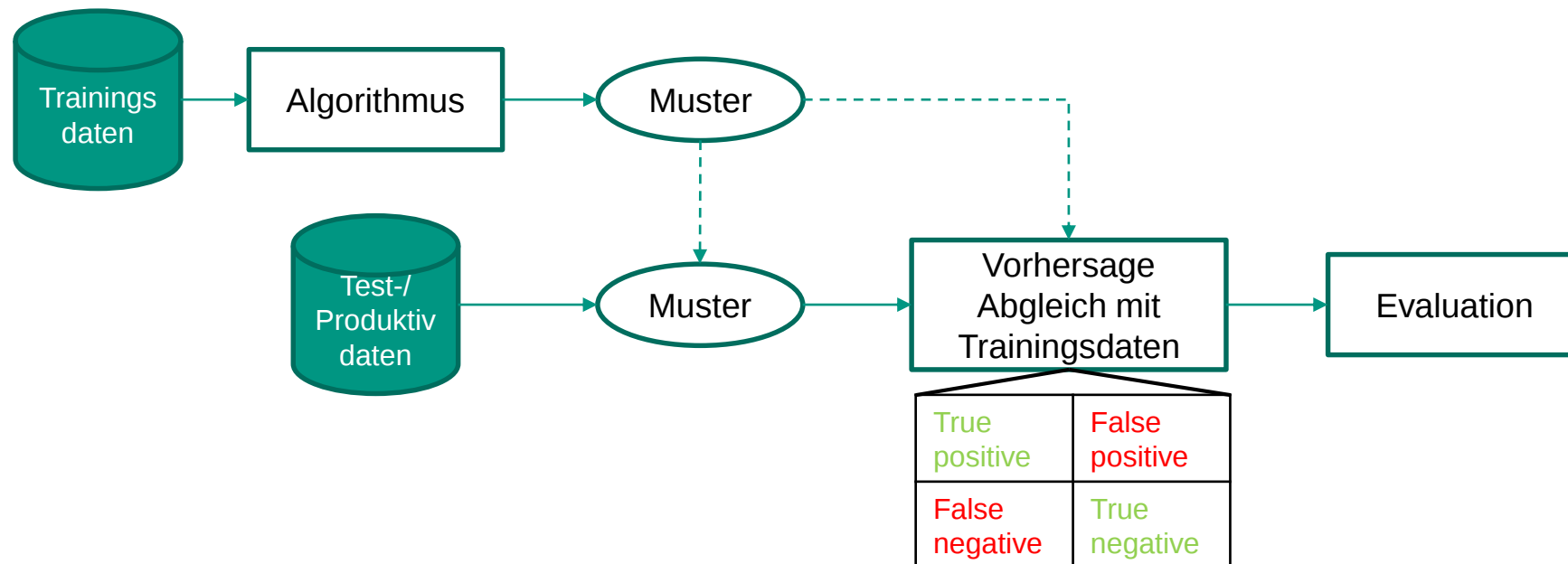


Fragen?

Klassifikation

Bewertung des trainierten Modells - Confusion Matrix

- Klare Zielvorgabe, eventueller Hypothesenansatz
- Über Trainingsphase auf ein Problem abgerichtet
- Produktiv als Assistenzsystem (bis hin zum Automated Decision Maker)
- Richtige Antwort muss während Trainingsphase vorliegen → Annotation, Label
- Lücke zwischen Input und definiertem Output füllen



Klassifikation

Bewertung des trainierten Modells - Confusion Matrix

$$TPR = \frac{\sum \text{True positives}}{\sum \text{Condition positive}}$$

$$FPR = \frac{\sum \text{False positives}}{\sum \text{Condition negative}}$$

$$FNR = \frac{\sum \text{False negatives}}{\sum \text{Condition positive}}$$

$$TNR = \frac{\sum \text{True negatives}}{\sum \text{Condition negative}}$$

		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“ 

„Legende“	
True	Algorithmus Vorhersage richtig
False	Algorithmus Vorhersage falsch
Predictive value:	<u>Entscheidung</u> des Algorithmus
	Pos Gesuchtes Objekt (Husky)
	Neg Nicht das gesuchte Objekt (kein Husky)
Actual value	Yes: gesuchtes Objekt dargestellt No: Gesuchtes Objekt nicht dargestellt

Gütekriterium:

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

True positive	False positive
False negative	True negative

Confusion Matrix

Zwischenübung



$$\text{TPR} = \frac{\sum \text{True positives}}{\sum \text{Condition positive}}$$

$$\text{FPR} = \frac{\sum \text{False positives}}{\sum \text{Condition negative}}$$

$$\text{FNR} = \frac{\sum \text{False negatives}}{\sum \text{Condition positive}}$$

$$\text{TNR} = \frac{\sum \text{True negatives}}{\sum \text{Condition negative}}$$

True positive (TP) „Vorhersage richtig, Objekt vorhanden“ $t_p = 80$	False positive (FP) „Vorhersage falsch, Objekt nicht vorhanden“ $f_p = 5$
False negative (FN) „Vorhersage falsch, Objekt vorhanden“ $f_n = 10$	True negative (TN) „Vorhersage richtig, Objekt nicht vorhanden“ $t_n = 50$

TPR = ?

FPR = ?

Untersuchte Bilder	n_{ges}	145
Tatsächlich Husky	n_{Husky}	90
Tatsächlich kein Husky	$n_{noHusky}$	55
Alg behauptet Husky	n_{AHusky}	85
Alg behauptet kein Husky	$n_{AnoHusky}$	60

Berechnen Sie:

- TPR
- FPR
- FNR
- TNR

FNR = ?

TNR = ?

Confusion Matrix

Zwischenübung - Lsg



$$\text{TPR} = \frac{\sum \text{True positives}}{\sum \text{Condition positive}}$$

$$\text{FPR} = \frac{\sum \text{False positives}}{\sum \text{Condition negative}}$$

$$\text{FNR} = \frac{\sum \text{False negatives}}{\sum \text{Condition positive}}$$

$$\text{TNR} = \frac{\sum \text{True negatives}}{\sum \text{Condition negative}}$$

True positive (TP) „Vorhersage richtig, Objekt vorhanden“ $t_p = 80$	False positive (FP) „Vorhersage falsch, Objekt nicht vorhanden“ $f_p = 5$
False negative (FN) „Vorhersage falsch, Objekt vorhanden“ $f_n = 10$	True negative (TN) „Vorhersage richtig, Objekt nicht vorhanden“ $t_n = 50$

TPR =

FPR =

FNR =

TNR =

Untersuchte Bilder

n_{ges}

145

Tatsächlich Husky

n_{Husky}

90

Tatsächlich kein Husky

$n_{noHusky}$

55

Alg behauptet Husky

n_{AHusky}

85

Alg behauptet kein Husky

$n_{AnoHusky}$

60

Berechnen Sie:

- TPR
- FPR
- FNR
- TNR

- Supervised Learning
 - Regression
 - Decision Tree
 - Neuronale Netze
 - Konfusionsmatrix



Ziele der heutigen Übung



■ Nach der heutigen Übung können Sie....

- ...Ansätze zur Verwaltung und Analyse großer Datenbestände hinsichtlich ihrer Anwendbarkeit und Wirksamkeit einschätzen

1

- ... Merkmale und Eigenschaften von selbstlernenden Algorithmen benennen und abgrenzen

2

- ... Methoden des maschinellen Lernens einordnen, beschreiben und bewerten

3

- ... Modelle berechnen unter Anwendung überwachter maschineller Lernmethoden

4

- ... Verfahren zur Auswahl einer geeigneter Methode beschreiben und anwenden