Frage **1**

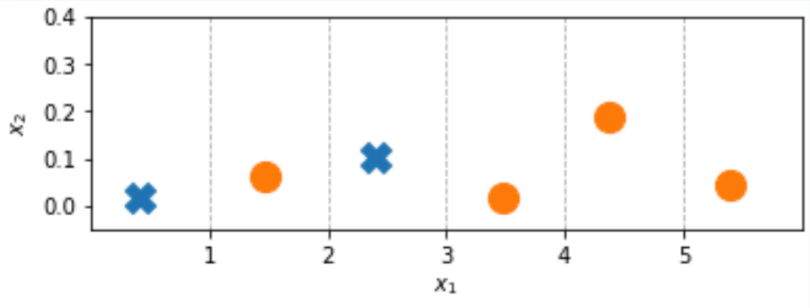
Bisher nicht beantwortet

Erreichbare Punkte: 7,00

Frage markieren

Entscheidungsbäume

Bestimme auf der Basis des Entropie-Unreinheitskriteriums die optimale vertikale Aufteilung für den folgenden Datensatz!



Fülle zunächst die folgende Tabelle aus. Jeder der 5 Kandidaten unterteilt den Datensatz in eine linke und eine rechte Teilmenge, N_L bzw. N_R.

- P_L und P_R sind die Proportionen der beiden Teilmengen.
- $i(N_L)$ und $i(N_R)$ sind die jeweiligen Unreinheiten (Impurities).

Bitte runde auf 2 Nachkommastellen!

Aufteilung in	P_L	P_R	$i\ (N_L)$	$i\ (N_R)$
$x_1 = 1$	0,17			
$x_1 = 2$		0,67		
$x_1 = 3$				
$x_1 = 4$	0,67	0,33	1,00	0,00
$x_1 = 5$	0,83	0,17		

Entscheide anhand der Tabelle, welche Aufteilung die optimale ist:

- x_1=1
- x_1= 2
- x_1=3
- x_1= 4
- x_1= 5

Frage **2**

Bisher nicht beantwortet

Erreichbare Punkte: 3,00

Frage markieren

Wähle alle Matrizen aus die keine Kovarianzmatrizen sind und begründe weshalb. Schreibe dazu im folgenden Textfeld für jede ausgewählte Matrix den Buchstaben und eine kurze Erklärung.

$$A = \begin{pmatrix} 0{,}80 & 0 \\ 0{,}1 & 1{,}2 \end{pmatrix} \qquad D = \begin{pmatrix} 2{,}1 & 0{,}56 \\ 0{,}65 & 3{,}4 \end{pmatrix}$$

$$B = \begin{pmatrix} 4{,}3 & 0{,}4 \\ 0{,}4 & 0{,}92 \end{pmatrix} \qquad E = \begin{pmatrix} 0{,}12 & 2{,}5 \\ 2{,}5 & 0{,}61 \end{pmatrix}$$

$$C = \begin{pmatrix} 1{,}8 & -0{,}29 \\ -0{,}29 & 6{,}8 \end{pmatrix} \qquad F = \begin{pmatrix} 0{,}18 & -0{,}29 \\ 0{,}29 & 0{,}68 \end{pmatrix}$$

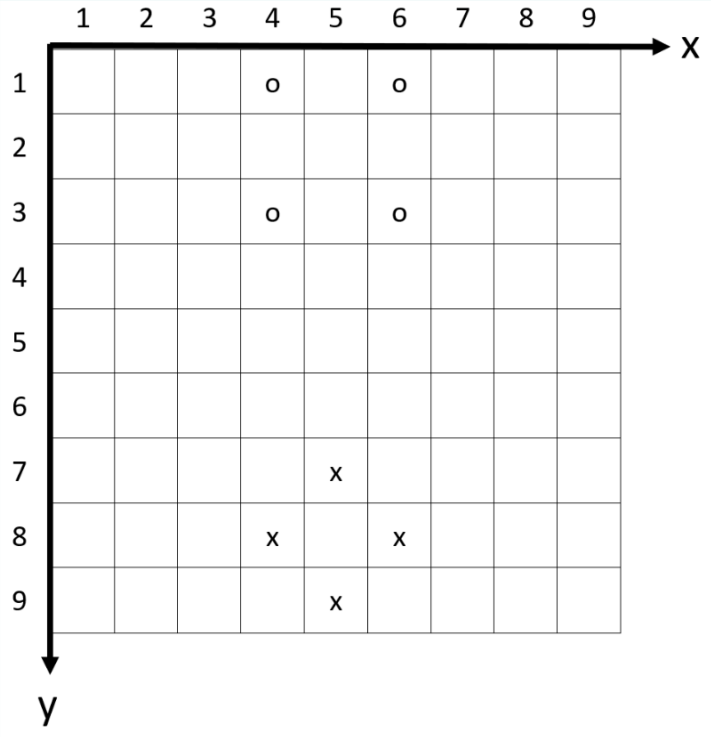
Frage **3**

Bisher nicht beantwortet

Erreichbare Punkte: 5,00

Frage markieren

Wir nehmen an, dass die Klassen o und x normalverteilt sind und ihre Priors gleich sind. Die Datenpunkte (o und x) der Klassen sind wie folgt verteilt:



Unter der Annahme, dass Spezialfall 1 gilt: Erkläre wie die Entscheidungsgrenze ohne Berechnungen bestimmbar ist. Wie verläuft sie und warum weiß man das ohne Berechnung?

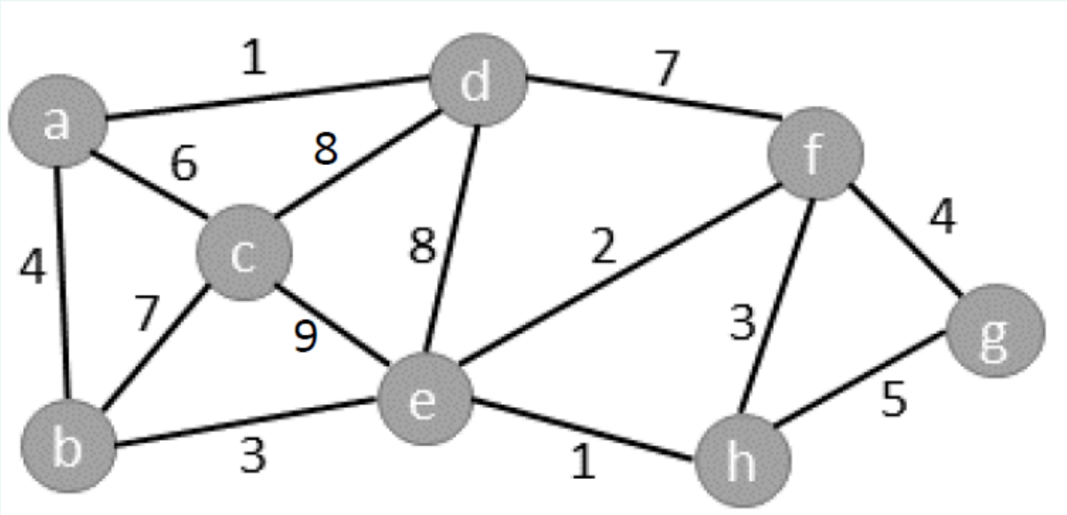
Frage **4**

Bisher nicht beantwortet

Erreichbare Punkte: 7,00

Frage markieren

Minimum Spanning Trees (MST) werden beim graph-basierten Clustering eingestzt. Welche Kanten werden vom Kurskal-Algorithmus für den MST in die Kanten-Liste eingetragen?



Kante	Gewicht	In der Kanten-Liste?	
ad	1	ja	nein
eh	1	ja	nein
ef	2	ja	nein
fh	3	ja	nein
be	3	ja	nein
ab	4	ja	nein
fg	4	ja	nein
gh	5	ja	nein
ac	6	ja	nein
bc	7	ja	nein
df	7	ja	nein
de	8	ja	nein
cd	8	ja	nein
ce	9	ja	nein

Frage **5**

Bisher nicht beantwortet

Erreichbare Punkte: 8,00

Frage markieren

Der Perceptron-Algorithmus hat einen Gewichtsvektor $w = [w_0, w_1, w_2]^T = [0, -1, 1]^T$ ergeben. Das erste Element dieses Vektors ist die homogene Koordinate.

Wende den Vektor w an, um die folgenden vier Testpunkte T_i als positive oder negative Klasse zu klassifizieren. Wähle je nach Bedarf "+" oder "-".

	x_1	x_2	Class
T_1	1	2	
T_2	2	1	
T_3	0	1	
T_4	1	0	

Frage **6**

Bisher nicht beantwortet

Erreichbare Punkte: 20,00

Frage markieren

Patienten mit COVID-42 Infektion überleben (Klasse ω = Ü) in 91,8% und sterben (Klasse ω = S) in 8,2% der Fälle. Diese Patienten können in 3 Risikogruppen R ∈ {R_1, R_2, R_3} eingeteilt werden. Mit Hilfe eines Trainingsdatensatzes wurden folgende Verteilungen für beide Klassen ermittelt:

	R_1	R_2	R_3
$P(R = R_i \mid \omega = \text{Ü})$	0,049	0,196	0,755
$P(R = R_i \mid \omega = \text{S})$	0,671	0,244	0,085

Basierend auf der Risikogruppe R_i sollte ein Mustererkennungssystem die Wahrscheinlichkeiten abs

1.) Ermittle mit Hilfe der Priors P(ω) und der Merkmalsverteilungen P(R|ω) die Randwarscheinlichkeiten P(R), runde sie auf 2 Dezimalstellen, und trage die Ergebnisse in die folgende Tabelle ein:

	R_1	R_2	R_3
$P(R=R_i)$			

2.) Anschließend sollen die a posteriori Wahrscheinlichkeiten für beide Klassen (Ü und S) und alle Risikogruppen ermittelt werden. Ergänze die folgende Tabelle um die fehlenden Werte. Runde die Ergebnisse auf 2 Dezimallstellen.

	ω=Ü	ω=S
$P(\omega \mid R = R_1)$		
$P(\omega \mid R = R_2)$		
$P(\omega \mid R = R_3)$		

3.) Wie soll man auf Grund der Bayes-Entscheidungsregel in den folgenden Fällen klassifizieren? Welche bedingte Fehlerwahrscheinlichkeit ergibt sich dadurch? Ergänze die folgende Tabelle:

	Klasse Ü oder S	P(error|x)
$R = R_1$		
$R = R_2$		
$R = R_3$		

4.) Berechne abhängig von den obigen Entscheidungen die Fehlerrate P(error) =

Frage **7**

Bisher nicht beantwortet

Erreichbare Punkte: 2,00

Frage markieren

Was trifft für den k-NN Klassifikator zu?

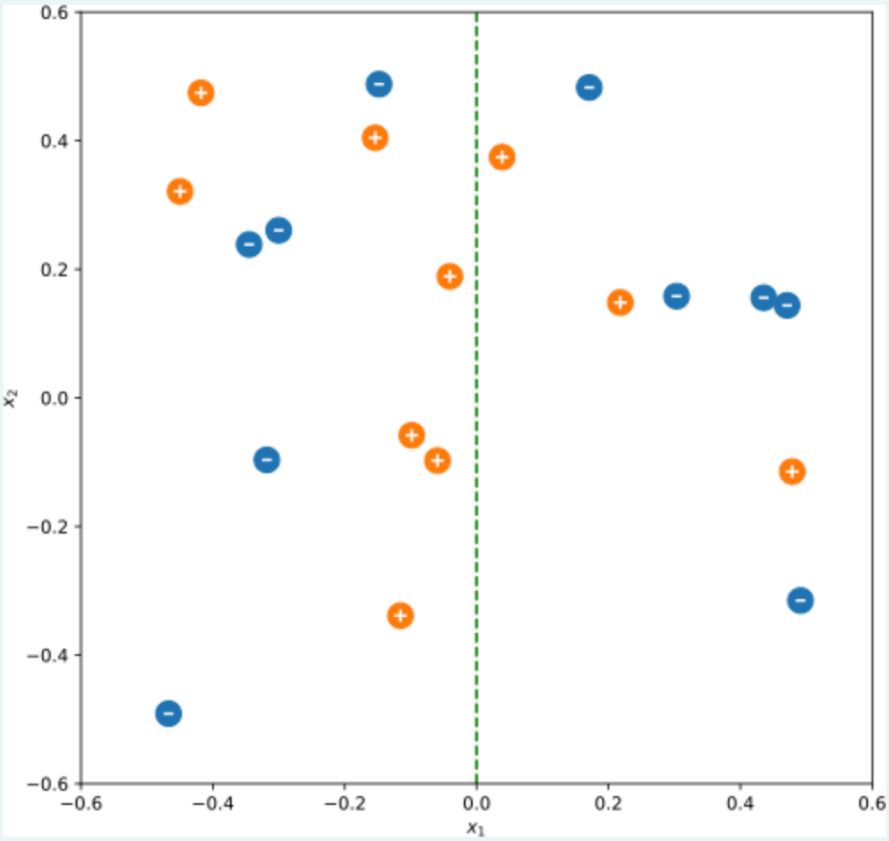
- [ ] a. Supervised Learning
- [ ] b. nicht-parametrisch
- [ ] c. Unsupervised Learning
- [ ] d. parametrisch

Frage **8**

Bisher nicht beantwortet

Erreichbare Punkte: 10,00

Frage markieren



Berechne für die obige Abbildung die gefragten Kennzahlen in der Tabelle. Die Entscheidungsregionen für die positive und die negative Klasse sind durch **sign(x_1)** gegeben.

TP (Anzahl angeben, nicht Anteil):	
FP (Anzahl angeben, nicht Anteil):	
TN (Anzahl angeben, nicht Anteil):	
FN (Anzahl angeben, nicht Anteil):	
Precision pos. Klasse (2 Nachkommastellen):	
Precision neg. Klasse (2 Nachkommastellen) :	
Recall pos. Klasse (2 Nachkommastellen):	
Recall neg. Klasse (2 Nachkommastellen):	
Overall Accuracy (2 Nachkommastellen):	

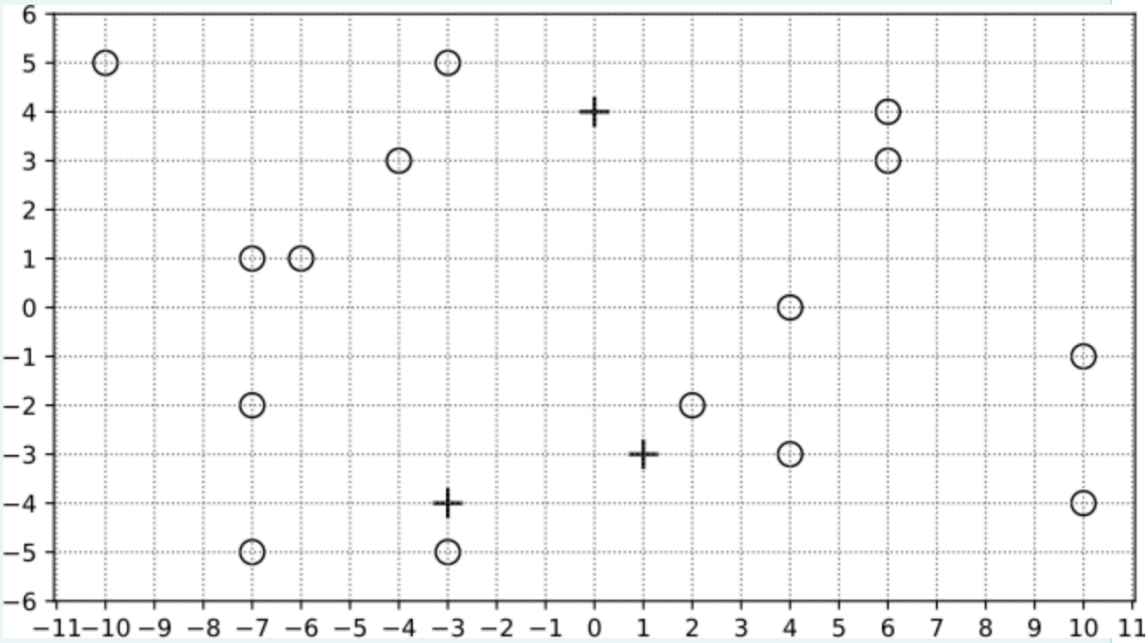
Frage **9**

Bisher nicht beantwortet

Erreichbare Punkte: 10,00

Frage markieren

Im Anschluss sieht man einen Raster in dem Datenpunkte (◦) eingetragen sind. Diese sollen mit Hilfe des k-Means Algorithmus in Cluster unterteilt werden. Die Anzahl der Cluster wurde auf k = 3 festgelegt und drei Start-Mittelpunkte sind bereits eingetragen (siehe +).



Führe **eine** Iteration des k-Means Algorithmus durch. Als Distanzmaß soll die L1-Norm (Manhattan-Distanz) verwendet werden.

Aufgaben:

1. Ordne jeden Punkt einem der drei Cluster zu
2. Berechne die neuen Koordinaten der Cluster-Mittelpunkte für die nächste Iteration. Sortiere die Mittelpunkte nach x von links nach rechts und **trage ihre Koordinaten** in der folgenden Tabelle ein.

Mittelpunkt	x	y
links		
mitte		
rechts		

Frage **10**

Bisher nicht beantwortet

Erreichbare Punkte: 6,00

Frage markieren

Im folgenden Raster wurden Datenpunkte der Klasse o und + eingetragen. Klassifiziere mit Hilfe des k-NN Klassifikators die neuen Datenpunkte A, B und C. Die Manhatten-Distanz (Hinweis: 4er-Nachbarschaft) soll als Metrik verwendet werden.

	1	2	3	4	5	6	7	8
1		+		+				+
2	o	A			+	+	o	
3								
4		+				o	B	o
5		C	+	o				
6		o						+
7								
8		o			o			+

Trage das Klassifikationsergebnis in die folgende Tabelle ein. Trage „o“ oder „+“ ein, je nachdem welcher Klasse der Datenpunkt zugeordnet wird.

	k = 3	k = 5
A		
B		
C		

Frage **11**

Bisher nicht beantwortet

Erreichbare Punkte: 2,00

Frage markieren

Bei welchen gültigen Kovarianzmatrizen gibt es einen negativen linearen Zusammenhang?

$$A = \begin{pmatrix} 0{,}80 & 0 \\ 0{,}1 & 1{,}2 \end{pmatrix} \qquad D = \begin{pmatrix} 2{,}1 & 0{,}56 \\ 0{,}65 & 3{,}4 \end{pmatrix}$$

$$B = \begin{pmatrix} 4{,}3 & 0{,}4 \\ 0{,}4 & 0{,}92 \end{pmatrix} \qquad E = \begin{pmatrix} 0{,}12 & 2{,}5 \\ 2{,}5 & 0{,}61 \end{pmatrix}$$

$$C = \begin{pmatrix} 1{,}8 & -0{,}29 \\ -0{,}29 & 6{,}8 \end{pmatrix} \qquad F = \begin{pmatrix} 0{,}18 & -0{,}29 \\ 0{,}29 & 0{,}68 \end{pmatrix}$$

- [ ] A
- [ ] B
- [ ] C
- [ ] D
- [ ] E
- [ ] F