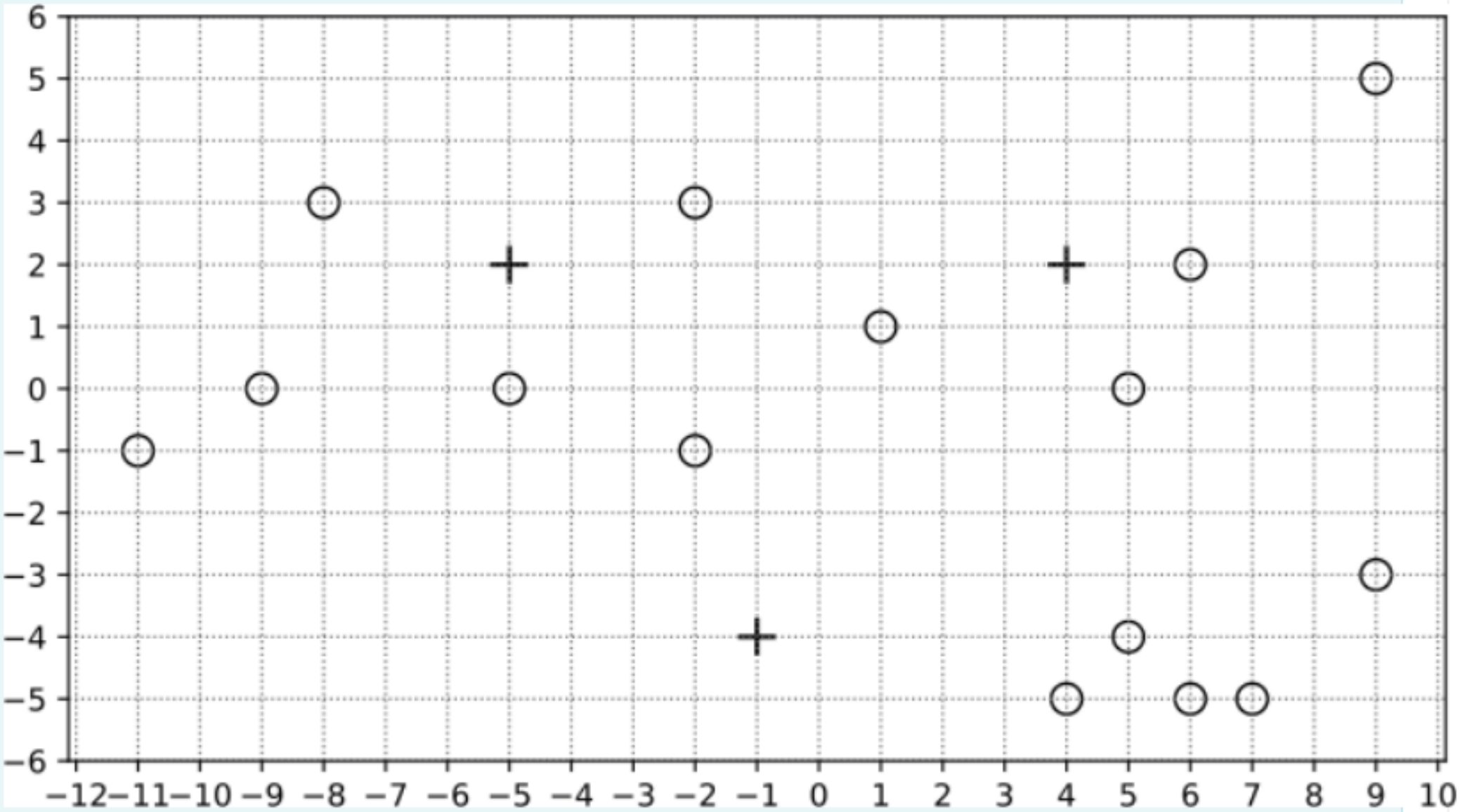
Frage **1**

Antwort gespeichert

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	-7	1
mitte	4	-4
rechts	6	1

Frage **2**

Antwort gespeichert

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 ab

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
$P(R=R_i)$	0,1	0,2	0,7

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
$P(\omega \mid R = R_1)$	0,45	0,55
$P(\omega \mid R = R_2)$	0,90	0,10
$P(\omega \mid R = R_3)$	0,99	0,01

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)
$R = R_1$	S	0,45
$R = R_2$	Ü	0,10
$R = R_3$	Ü	0,01

4.) Berechne abhängig von den obigen Entscheidungen die Fehlerrate P(error) = 0,07

Frage **3**

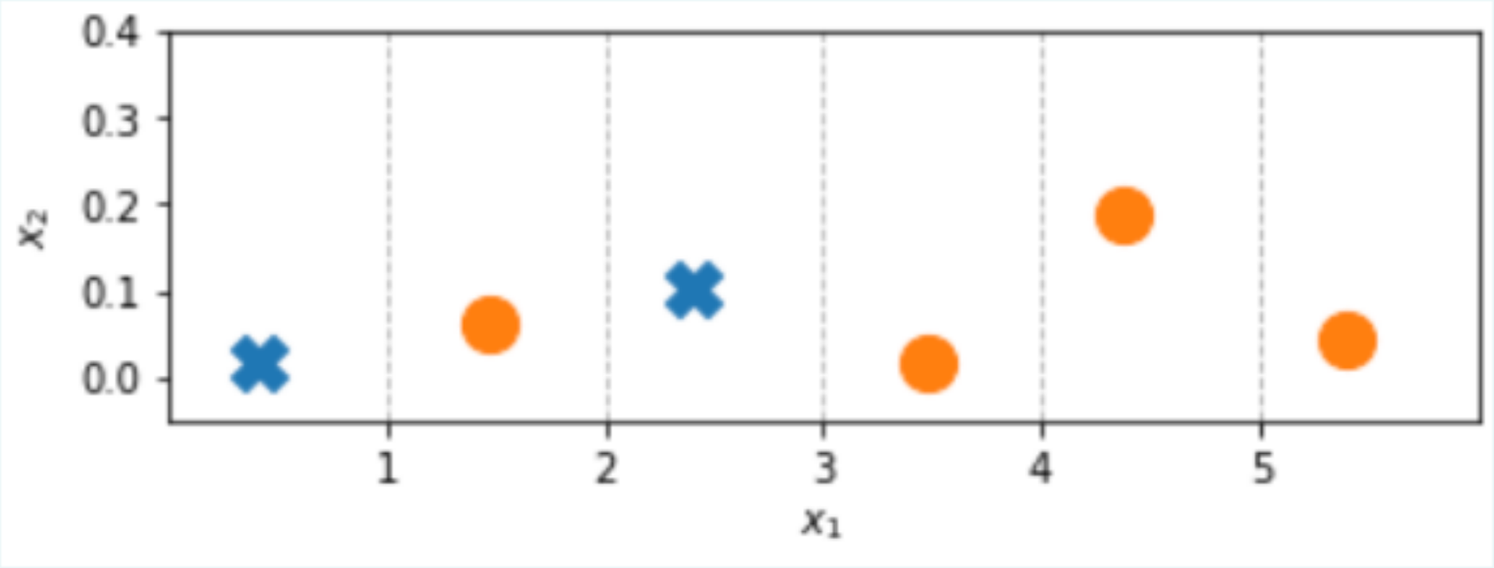
Antwort gespeichert

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	0,83	0	0,72
$x_1 = 2$	0,33	0,67	1	0,81
$x_1 = 3$	0,5	0,5	0,92	0
$x_1 = 4$	0,67	0,33	1,00	0,00
$x_1 = 5$	0,83	0,17	0,97	0

Entscheide anhand der Tabelle, welche Aufteilung die optimale ist:

○ x_1=1 ○ x_1= 2 ◉ x_1=3 ○ x_1= 4 ○ x_1= 5

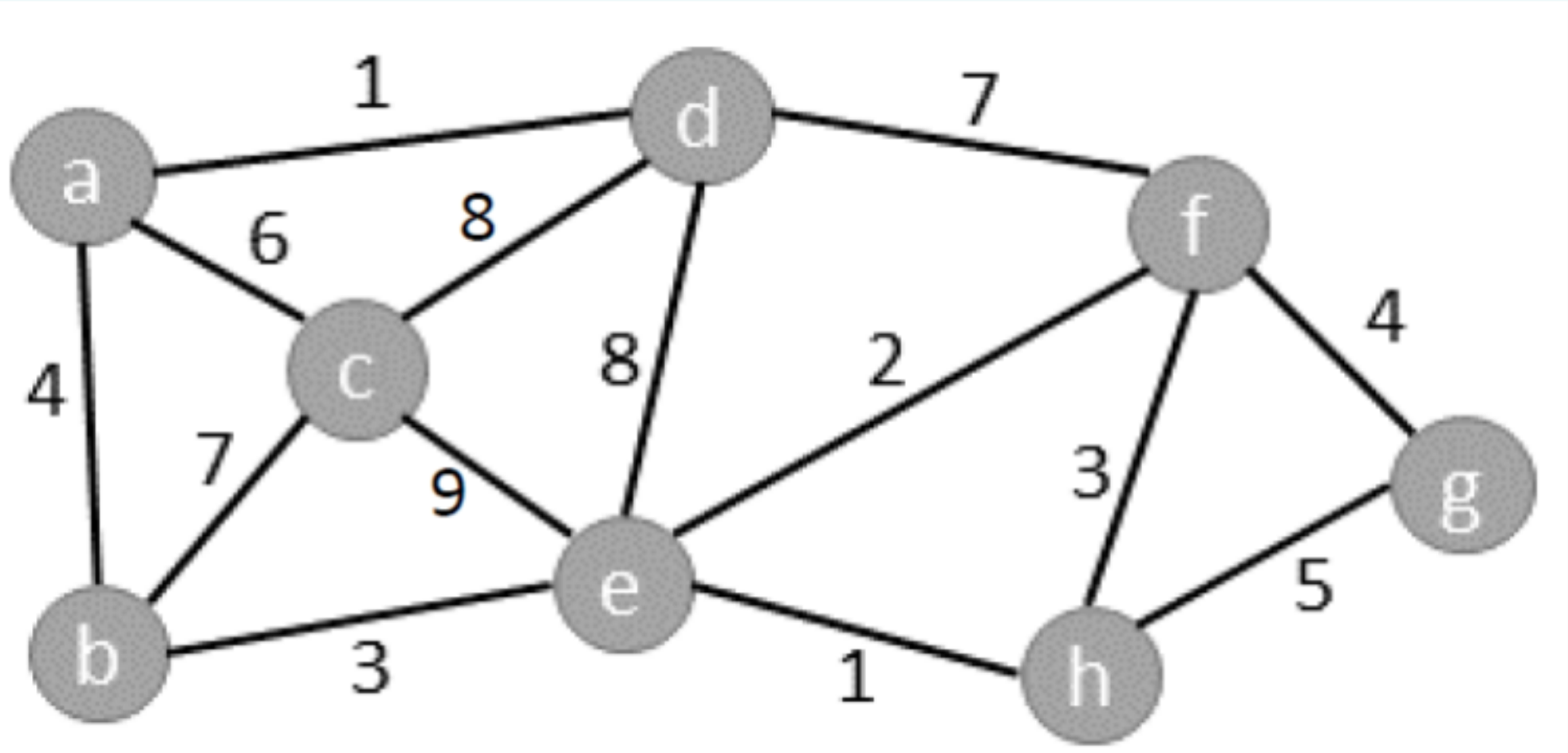
Frage **4**

Antwort gespeichert

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**

Antwort gespeichert

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.

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

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

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

D ist keine Kovarianzmatrix, weil sie nicht symmetrisch ist. (0.56 vs 0.65)
E ist keine Kovarianzmatrix, weil sie einen negativen Eigenwert hat (etwa -2.15)
F ist keine Kovarianzmatrix, weil sie nicht symmetrisch ist. (-0.29 vs 0.29)

Die anderen sind Kovarianzmatrizen.

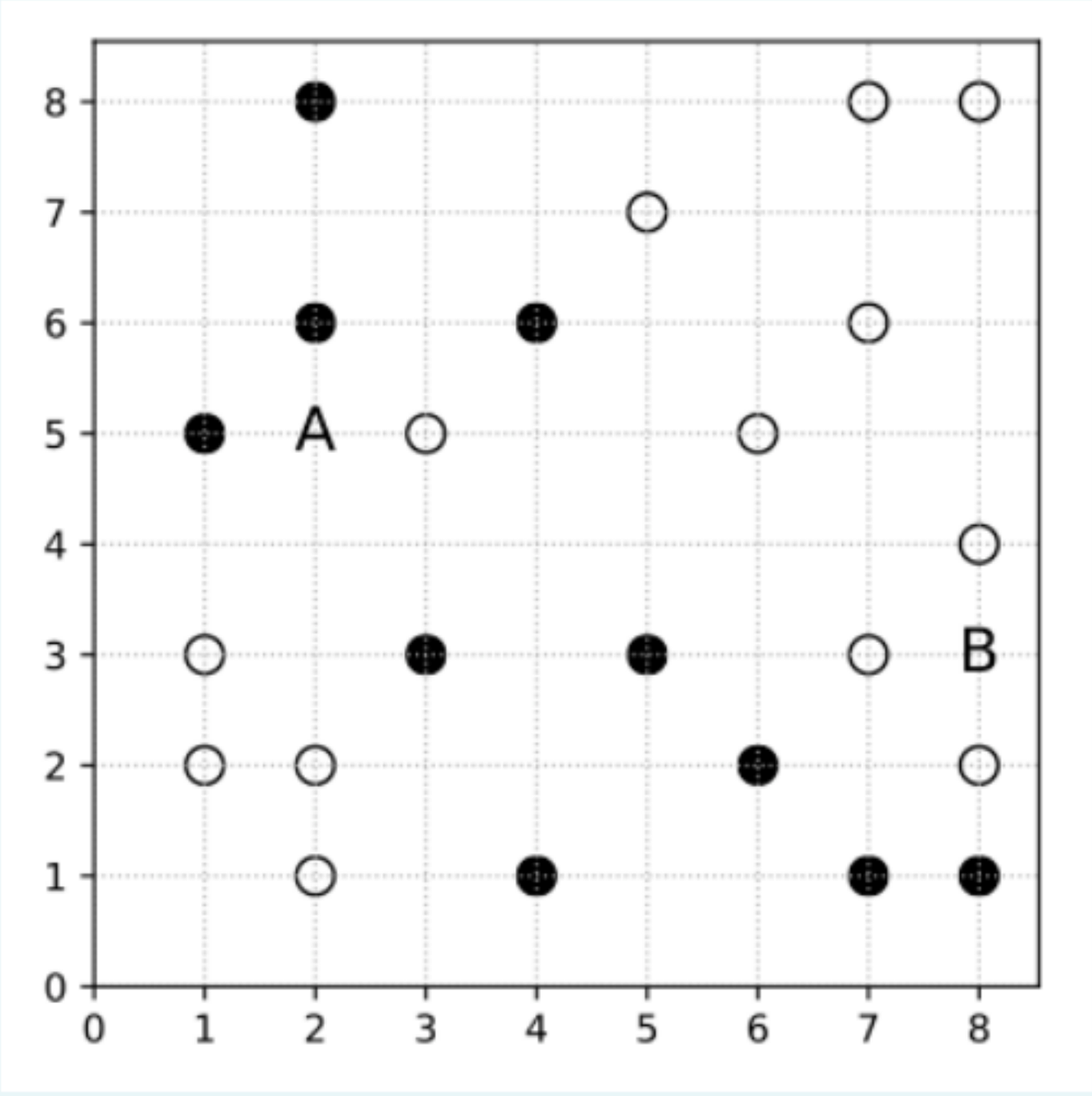
Frage **6**

Antwort gespeichert

Erreichbare Punkte: 6,00

Frage markieren

Im folgenden Raster wurden Datenpunkte der Klasse o und • eingetragen. Klassifiziere mit Hilfe des k-NN Klassifikators die beiden neuen Datenpunkte A und B. Die L1-Norm (Hinweis: Manhattan-Distanz, 4er-Nachbarschaft) soll als Metrik verwendet werden.



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

	k = 3	k = 7
A	•	•
B	o	•

Frage **7**

Antwort gespeichert

Erreichbare Punkte: 8,00

Frage markieren

Der Perceptron-Algorithmus hat einen Gewichtsvektor $w = [w_0, w_1, w_2]^T = [1, 5, -9]^T$ 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 "-".

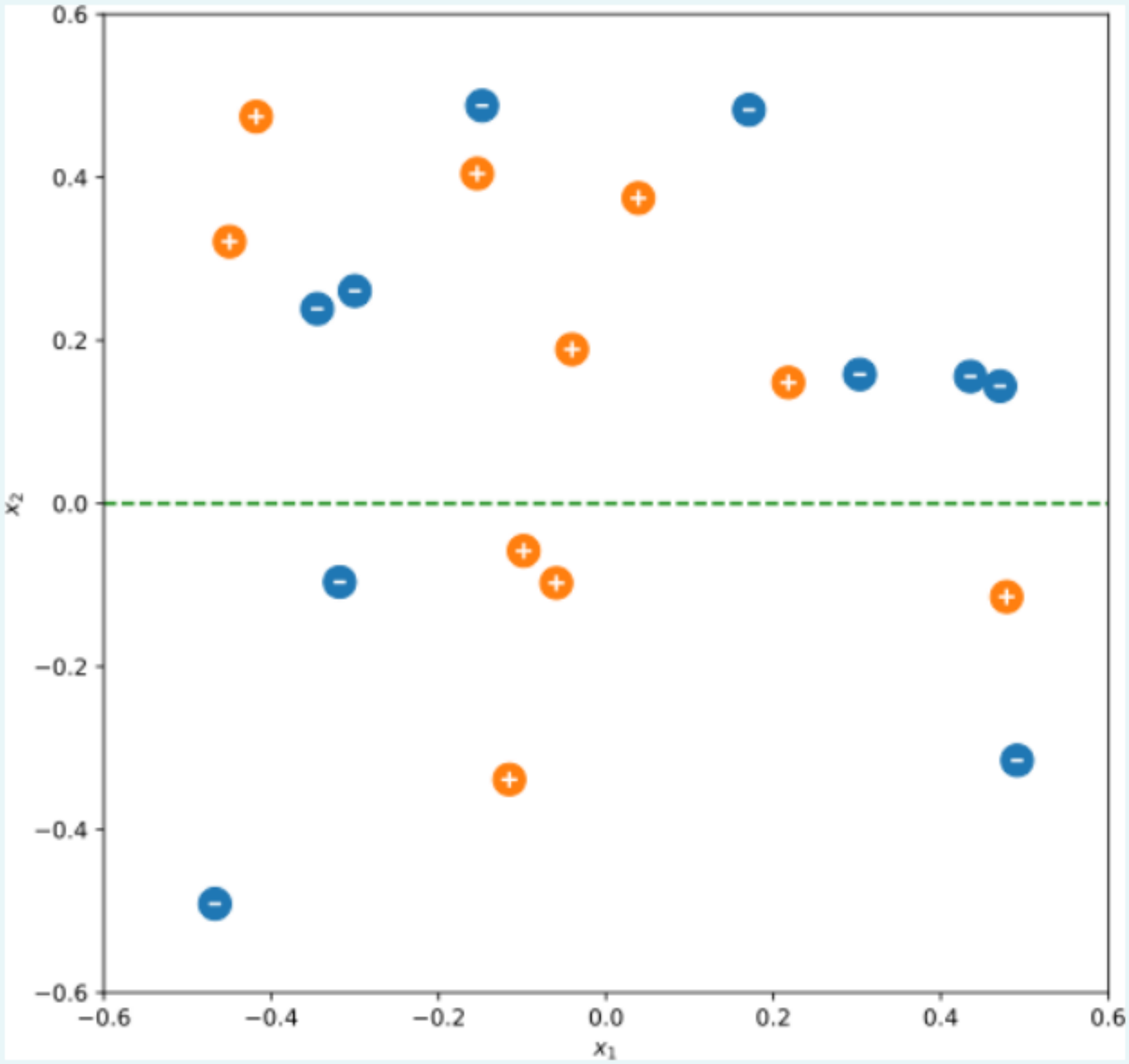
	x_1	x_2	Class
T_1	3	6	-
T_2	3	7	-
T_3	8	5	-
T_4	5	2	+

Frage **8**

Antwort gespeichert

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_2)** gegeben.

TP (Anzahl angeben, nicht Anteil):	6
FP (Anzahl angeben, nicht Anteil):	7
TN (Anzahl angeben, nicht Anteil):	3
FN (Anzahl angeben, nicht Anteil):	4
Precision pos. Klasse (2 Nachkommastellen):	0,46
Precision neg. Klasse (2 Nachkommastellen):	0,43
Recall pos. Klasse (2 Nachkommastellen):	0,6
Recall neg. Klasse (2 Nachkommastellen):	0,3
Overall Accuracy (2 Nachkommastellen):	0,45

Frage **9**

Antwort gespeichert

Erreichbare Punkte: 2,00

Frage markieren

Was trifft für den k-NN Klassifikator zu?

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

Frage **10**

Antwort gespeichert

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
- [x] C
- [ ] D
- [ ] E
- [ ] F

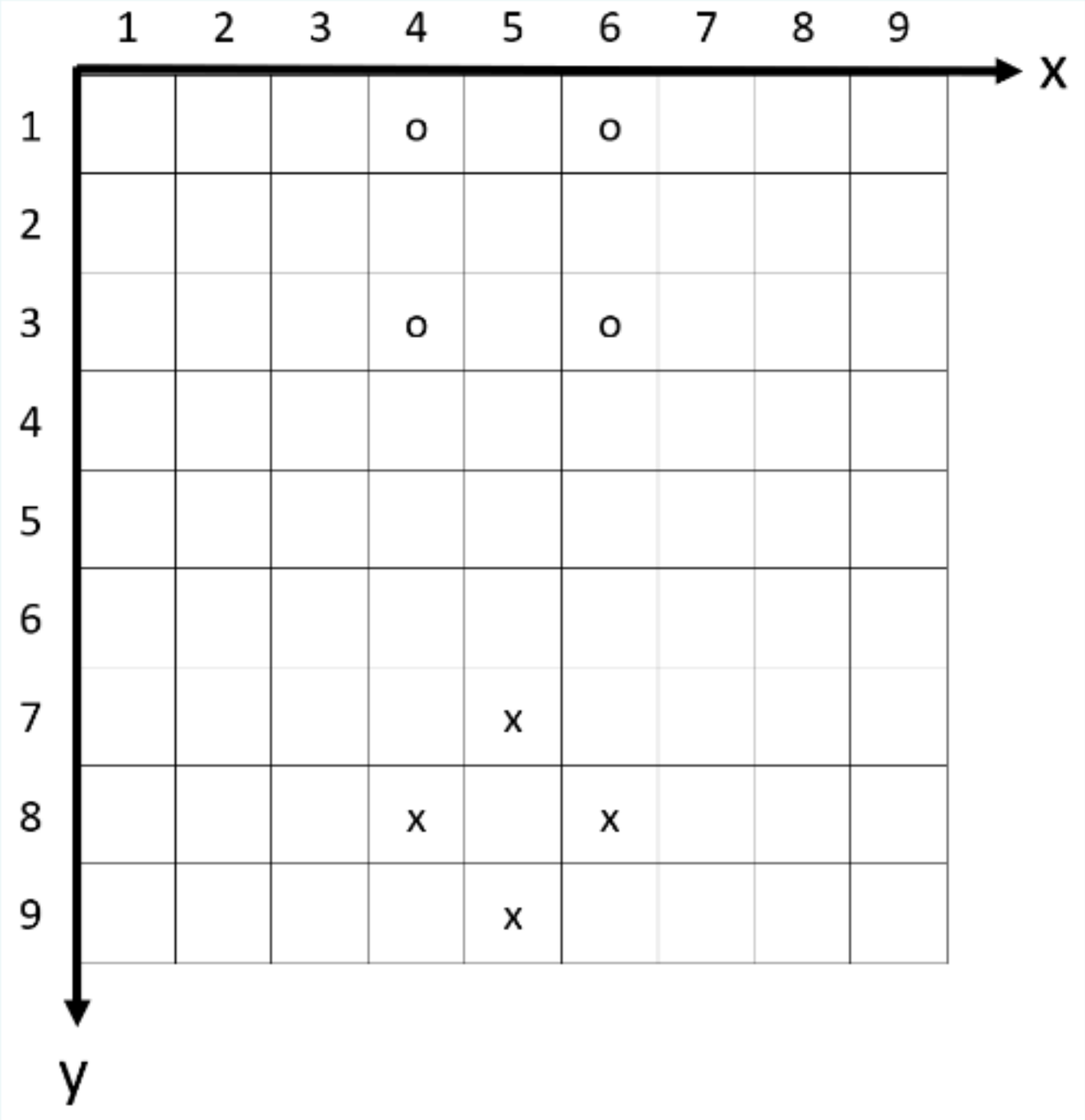
Frage **11**

Antwort gespeichert

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?

Bei Spezialfall 1 sind die Entscheidungsgrenzen linear und normal zur Verbindungsstrecke zwischen den beiden Klassenmitteln. Bei Gleichen Priors verläuft die Entscheidungsgrenze durch (μ_1 + μ_2)/2

Die Klassenmitteln liegen bei (5,2) und (5,8), die Verbindungsstrecke ist senkrecht, ihre Normale also waagerecht. Sie geht durch den Punkt (5,5).

Die Entscheidungsgrenze ist also eine waagerechte Linie durch den Punkt (5,5), anders ausgedrückt f(x)=y=5