

Klausur zur Lehrveranstaltung  
„Maschinelles Lernen 1 – Grundverfahren“  
(60 Minuten)

Nachname:	Vorname:
Matrikelnummer:	Studiengang & Semester:
Bekanntgabe-Code:	

## Anmerkungen

- Legen Sie Ihren Studierendenausweis und ein gültiges Ausweisdokument gut sichtbar bereit.
- Tragen Sie Nachname, Vorname, Matrikelnummer, Studiengang & Semester und Bekanntgabecode deutlich lesbar ein und unterschreiben Sie das Klausurexemplar unten.
- Die folgenden 6 Aufgaben sind vollständig zu bearbeiten.
- Als Hilfsmittel ist nur ein nicht programmierbarer Taschenrechner zugelassen.
- Täuschungsversuche führen zum Ausschluss von der Klausur.
- Unleserliche oder mit Bleistift geschriebene Lösungen können von der Korrektur bzw. der Wertung ausgeschlossen werden.
- Beim Ausfüllen von Lücken gibt die Größe der Kästen keinen Aufschluss über die Länge des einzufügenden Inhaltes.
- Die Bearbeitungszeit beträgt 60 Minuten.

**Ich bestätige, dass ich die Anmerkungen gelesen und mich von der Vollständigkeit dieses Klausurexemplars (Seite 1 - 14) überzeugt habe.**

Unterschrift

**Nur für den Prüfer**

Aufgabe	1	2	3	4	5	6	Gesamt	Note
Punkte	8	12	10	8	10	12	60	
Erreicht								

**Aufgabe 1 – Lerntheorie, induktives und unüberwachtes Lernen****\_\_\_ / 8P**

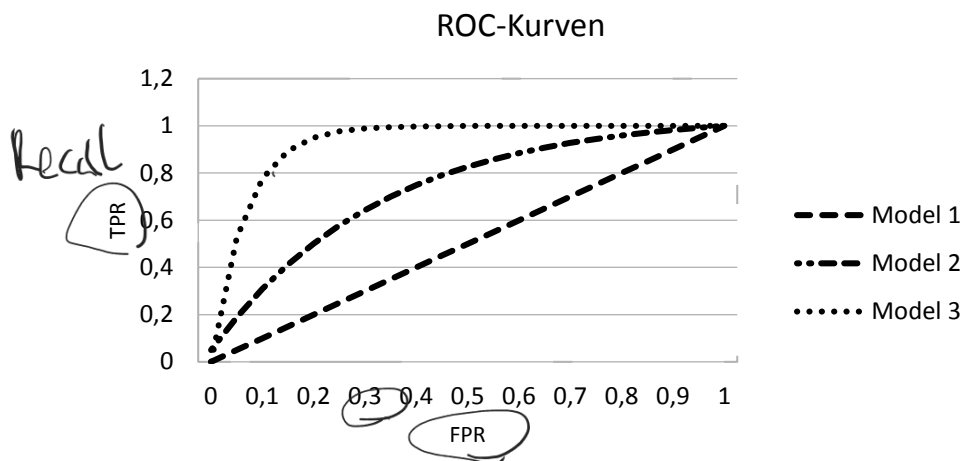
- a) Ausgehend von der lerntheoretischen Abschätzung des realen Fehlers, von welchen drei Kriterien hängt der Lernerfolg einer Lernmaschine ab?

(\_\_\_/1,5P)

Capacity of learning system  
 optimization method  
 learning instances in dataset

- b) Es wurden drei Modelle trainiert und anschließend mit der ROC-Kurve evaluiert. Welches Modell würde das womöglich beste Ergebnis liefern? Begründen Sie Ihre Entscheidung und geben Sie die Formeln für FPR und TPR an.

(\_\_\_/4,5P)



Model 3

TPR should be as high as possible  $TPR = \frac{TP}{TP + FN}$

FPR should be as small as possible

$$FPR = \frac{FP}{FP + FN}$$

- c) Welche zwei Parameter sind für das k-Means Clustering festzulegen?  
Welche Herausforderungen (Probleme) entstehen durch diese Parameter?

(\_\_\_/2P)

Number of clusters  $k$

initial cluster center  $c$

Problem: Wrong number of initial clusters

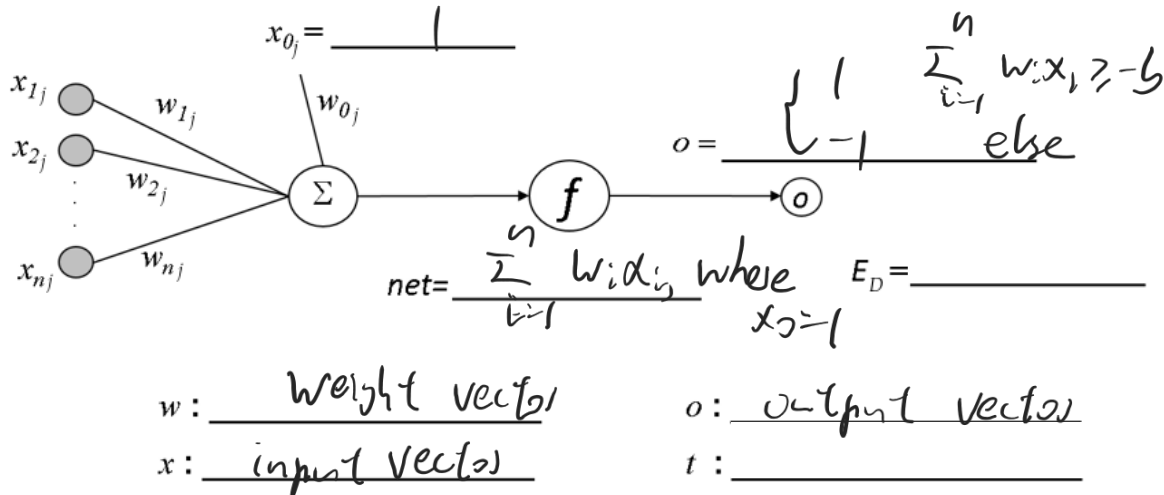
## Aufgabe 2 – Neuronale Netze

\_\_\_ / 12P

a) Beschriften Sie das unten abgebildete Neuron.

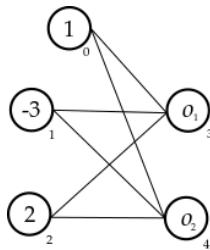
04, 10P

(\_\_\_/5P)



b) Führen Sie einen Forward-Pass auf dem nachfolgenden Neuronalen Netz mit einer ReLu-Aktivierungsfunktion aus, um die Ausgabe  $o_1$  und  $o_2$  zu berechnen. Mit  $w_{ij}$  wird hierbei das Gewicht zwischen Neuron  $i$  und Neuron  $j$  beschrieben.

(\_\_\_/3P)



Die Gewichte sind wie folgt vorgegeben:

$$w_{03} = -\frac{3}{2}$$

$$w_{04} = 1$$

$$w_{13} = \frac{1}{6}$$

$$w_{14} = -\frac{1}{2}$$

$$w_{23} = \frac{3}{4}$$

$$w_{24} = \frac{5}{4}$$

$$\begin{aligned}
 o_1 &= \max(0, 1 \times (-\frac{3}{2}) + (-3) \times \frac{1}{6} + 2 \times \frac{3}{4}) \\
 &= \max(0, -\frac{1}{2}) = 0
 \end{aligned}$$

$$\begin{aligned}
 o_2 &= \max(0, 1 \times 1 + (-3) \times (-\frac{1}{2}) + 2 \times \frac{5}{4}) \\
 &= \max(0, 5) \\
 &= 5
 \end{aligned}$$

- c) Wodurch treten verschwindende Gradienten (vanishing gradients) auf? (\_\_\_/2P)  
Welche Gegenmaßnahme gibt es ohne die Anzahl der Neuronen zu verändern?

Activation functions have saturation behavior.  
Use activation functions with less saturation.

- d) Nennen Sie ein Optimierungsverfahren, das dem einfachen Gradientenabstieg überlegen ist und nennen Sie zwei Gründe wieso es besser ist. (\_\_\_/1P)

SGD  
Faster update cycles; Computationally tractable with limited memory

- e) „Je höher die VC-Dimension, umso besser kann das Neuronale Netz aus einem bestehenden Datensatz lernen und generalisieren.“ (\_\_\_/1P)  
Ist diese Aussage wahr oder falsch? Begründen Sie Ihre Antwort.

Falsch. Because model with larger VC-dimension is easier to be overfitting.

**Aufgabe 3 – Convolutional Neural Networks****\_\_\_ / 10P**

- a) Nennen Sie drei Methoden der Gewichtsinitialisierung für CNNs.  
Was passiert, wenn alle Gewichte mit 1 initialisiert werden?

(\_\_\_/2P)

Xavier, Kaiming He, transfer learning

Vanishing Gradient

"Symmetry Breaking"

- b) Die Größe eines Eingabebildes für ein CNN beträgt 97x97x3 und es werden 128 Filteroperationen (Faltungen) durchgeführt. Dabei ist die angewendete Filtergröße 3x3, die Schrittweite ist 2 und es wird kein Zero Padding verwendet.

(\_\_\_/3P)

Berechnen Sie die resultierende Ausgabedimension. Geben Sie dazu die vollständige Formel an und benennen Sie die einzelnen Elemente.

$$m_{\text{new}} = \frac{(m-n)}{s} + 1$$

$$\frac{97-3}{2} + 1 = 48$$

$$48 \times 48 \times 128$$

input image:  $n \times n$

output image:  $m_{\text{new}} \times m_{\text{new}}$

S: Slide

filter size:  $n \times n$

- c) Was ist der Zweck einer 1x1 Faltung, wie sie im Inception Modul vorkommt?

(\_\_\_/1P)

47/47

- d) Das Eingabebild hat eine Dimension von  $5 \times 5 \times 2$ . Berechnen Sie die Ausgangsvolumenwerte beim Anwenden einer Faltungsschicht mit Zero Padding und Stride = 2 (vertikal und horizontal). Der Bias hat einen Wert von 1. (\_\_\_/4P)

Eingabebild  $[5 \times 5 \times 2]$ Filter  $W_0$  ( $3 \times 3 \times 2$ )Ausgabebild  $[3 \times 3 \times 1]$ 

$X[:, :, 0]$

1	0	2	1	2
1	2	1	1	1
1	2	0	0	0
0	0	1	2	0
0	0	0	1	2

$W_0[:, :, 0]$

0	1	0
-1	1	1
0	0	0

$(-1 \quad -2 \quad 1)$   
 $-1 \quad 1 \quad 1$

$1 \quad 1 \quad 2 \quad 1$   
 $2 \quad 1 \quad -1 \quad -2 \quad 1$   
 $-1 \quad 1 \quad 2 \quad 1$   
 $1 \quad 1 \quad 2 \quad 1 \quad 2 \quad 1 \quad 1$

5	1	2
8	1	-2
3	3	0

$X[:, :, 1]$

2	1	0	0	0
0	2	1	0	0
0	0	0	2	1
1	1	1	2	1
2	1	2	0	1

$W_0[:, :, 1]$

-1	0	1
-1	0	1
-1	0	1

**Aufgabe 4 – Support Vector Machines****\_\_\_ / 8P**

- a) Beschreiben Sie die grundlegende Intuition auf der eine Support Vector Machine (SVM) für die Klassifikation basiert. (\_\_\_/1P)

Size of the margin determines the generalization capacity

- b) Formulieren Sie das grundlegende Optimierungsproblem mit Bedingungen für einen Support Vector Klassifikator formal. (\_\_\_/2P)

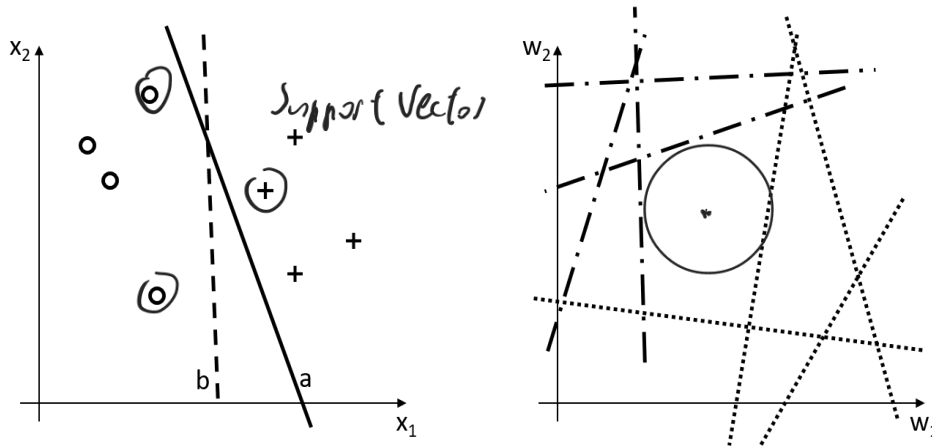
- c) Nennen Sie jeweils zwei Vor- und Nachteile von Support Vector Machines. (\_\_\_/2P)

Adv: Effective use of feature space; Robust against overfitting  
Dis: Sensitive to parameter choice; Computationally intensive



- d) Welche der beiden Hypothesen in der linken Grafik ist optimal?  
Markieren Sie die Supportvektoren für Ihre Auswahl und tragen Sie die  
Parameter  $w$  der optimalen Hypothese in den rechten Graphen ein.

(\_\_\_/3P)



b

**Aufgabe 5 – Reinforcement Learning****\_\_\_ / 10P**

- a) Durch welches Modell lässt sich die Problemstellung beim Reinforcement Learning formal darstellen? Welche vier Bestandteile werden für die Modellierung benötigt? (\_\_\_/3P)

Markov Decision Process

State Space Action Space

Reward function Transition function

- b) Was besagt die Markov-Bedingung? (\_\_\_/1P)

$$P(S_{t+1} | S_t, a_t) = P(S_{t+1} | S_0, a_0, \dots, S_t, a_t)$$

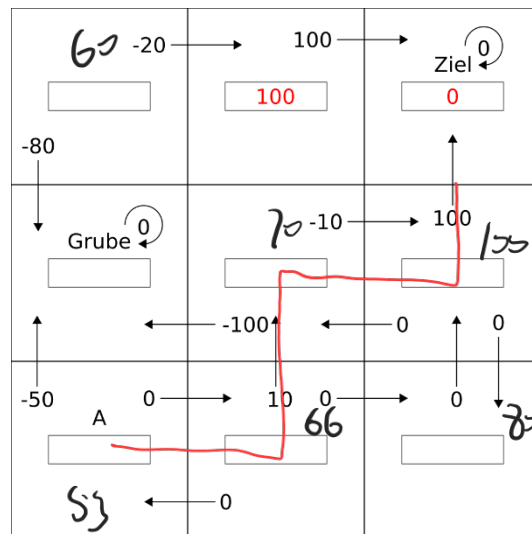
- c) Gegeben ist die Wertfunktion  $V^\pi(s)$ : Geben Sie die formale Beschreibung der optimalen Strategie  $\pi^*(s)$  an. Definieren Sie zudem die rekursive Form der Bellmann Optimalitätsgleichung in Abhängigkeit der optimalen Wertfunktion. (\_\_\_/2P)

$$\pi^*(s) = \underset{\pi}{\operatorname{argmax}} V^\pi(s)$$

$$V^*(s) = \max_{a \in A} (r(s, a) + \gamma \sum_{s' \in S} P(s' | s, a) V^*(s'))$$

- d) Betrachten Sie die untenstehende Welt. Ein Agent kann sich mit den angezeigten Zustandsübergängen von Zelle zu Zelle bewegen. Die Belohnung für einen Übergang entspricht der Zahl an den Pfeilen. Nehmen Sie an, dass die optimale Strategie gelernt wurde. Tragen Sie die Zustandswerte ( $V^*(s)$ ) dieser Strategie in die entsprechenden Kästen ein (Diskontierungsfaktor  $\gamma = 0,8$ ). Runden Sie ihre Ergebnisse auf ganze Zahlen. Zeichnen Sie den Pfad der optimalen Strategie von Zelle A zum Ziel ein.

(\_\_\_\_/4P)

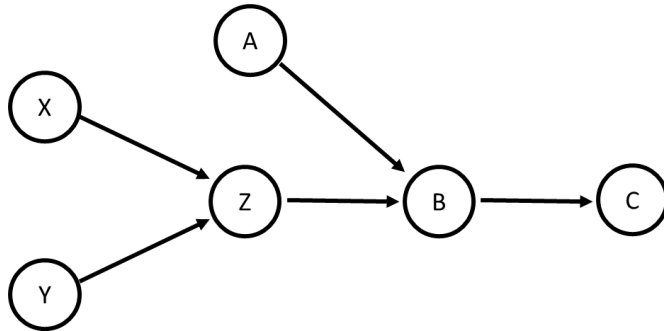


**Aufgabe 6 – Bayes'sche Netze und HMMs****\_\_\_ / 12P**

- a) Gegeben sei das folgende Bayes Netz mit binären Zufallsvariablen  $X, Y, Z, A, B, C \in \{0, 1\}$ .

(\_\_\_/2P)

Geben Sie die Faktorisierung der Verbundwahrscheinlichkeitsverteilung  $P(X, Y, Z, A, B, C)$  an, die das abgebildete Bayes-Netz repräsentiert.



$$P(X, Y, Z, A, B, C) = P(X) \cdot P(Y) \cdot P(Z|X, Y) \cdot P(A) \cdot P(B|Z, A) \cdot P(C|B)$$

b) Gegeben sei der folgende Datensatz für das Bayes'sche Netz aus Aufgabe a):

(\_\_\_\_/4P)

A	B	C	Z	X	Y
1	0	0	0	0	0
0	0	0	1	0	0
0	1	0	1	0	0
0	0	0	0	1	0
1	0	0	0	0	1
1	0	0	1	0	1
0	0	1	0	1	1
0	0	0	1	1	1
0	0	1	1	1	1
1	0	1	1	1	1
1	1	1	1	1	1

Schätzen Sie durch relative Häufigkeiten (ohne Laplace-Schätzer) die Werte der bedingten Wahrscheinlichkeitsverteilungen  $P(Z | X, Y)$  anhand der beobachteten Daten. Tragen Sie ihre Ergebnisse in folgende Tabelle ein:

X	Y	Z	$P(Z X,Y)$
0	0	0	1/11
0	0	1	6/77
0	1	0	1/11
0	1	1	15/77
1	0	0	1/11
1	0	1	8/77
1	1	0	1/11
1	1	1	20/77

$$P(Z=0) = \frac{4}{11} \quad P(Z=1) = \frac{7}{11}$$

$$P(X=0 | Z=0) = \frac{2}{4} = \frac{1}{2} \quad P(X=1 | Z=0) = \frac{1}{2}$$

$$P(Y=0 | Z=0) = \frac{2}{4} = \frac{1}{2} \quad P(Y=1 | Z=0) = \frac{1}{2}$$

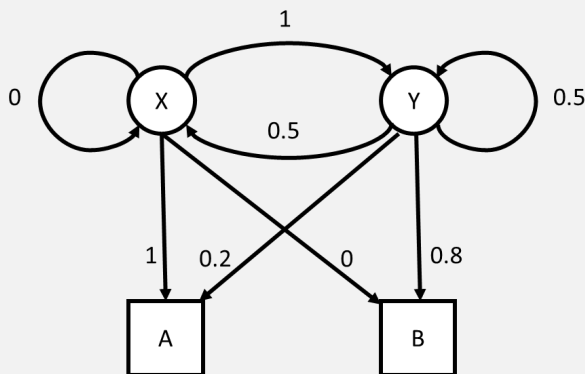
$$P(X=0 | Z=1) = \frac{3}{7} \quad P(X=1 | Z=1) = \frac{4}{7}$$

$$P(Y=0 | Z=1) = \frac{2}{7} \quad P(Y=1 | Z=1) = \frac{5}{7}$$

- c) Nennen Sie zwei Verfahren mit denen die Parameter eines Bayes Netzes gelernt werden können, wenn nicht alle Variablen beobachtbar sind. (\_\_\_/2P)

*Gradient ascent, expectation-maximization algorithm*

Gegeben sei das folgende Hidden Markov Modell mit der Zustandsmenge  $S = \{X, Y\}$  und der Menge der möglichen Beobachtungen  $O = \{A, B\}$ . Die Startverteilung bei  $t = 0$  sei  $S_0 = (0.5, 0.5)$ .



- d) Sagen Sie den (System-)Zustand des Hidden-Markov-Modells zum Zeitpunkt  $t = 2$  voraus (Prädiktion). (\_\_\_/2P)

- e) Sie erhalten zum Zeitpunkt  $t = 1$  die Messung  $O_1 = B$ . Schätzen Sie den Zustand des Hidden Markov Modells zum Zeitpunkt  $t = 1$ . (\_\_\_/2P)

