

**Klausur über den Stoff der Lehrveranstaltung:**  
**„Maschinelles Lernen 1 - Grundverfahren“**  
 (60 Minuten)

Name:	Vorname:
Matrikelnr.:	Studiengang:

Anmerkungen:

- Legen Sie Ihren Studierendenausweis gut sichtbar bereit.
- Tragen Sie Nachname, Vorname, Matrikelnummer und Studiengang deutlich lesbar ein und unterschreiben Sie das Klausurexemplar unten.
- Die folgenden 6 Aufgaben sind vollständig zu bearbeiten. Jede Antwort muss entweder in deutscher oder englischer Sprache formuliert sein.
- Als Hilfsmittel sind ausschließlich folgende zugelassen:
  - ein nicht programmierbarer Taschenrechner
  - ein nicht beschriftetes Wörterbuch
- Täuschungsversuche führen zum Ausschluss von der Klausur.
- Unleserliche oder mit Bleistift geschriebene Lösungen können von der Korrektur bzw. der Wertung ausgeschlossen werden.
- Die Bearbeitungszeit beträgt 60 Minuten.

Ich bestätige, dass ich die Anmerkungen gelesen und mich von der Vollständigkeit dieses Klausurexemplars (Seite 1 - 19 ) überzeugt habe.

---

Unterschrift

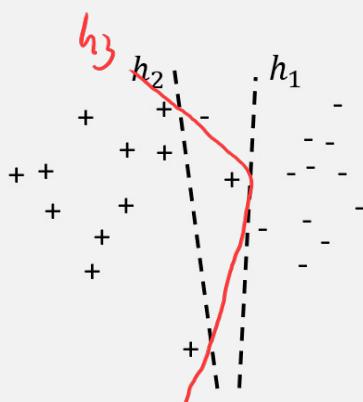
Nur für den Prüfenden:

Aufgabe	1	2	3	4	5	6	Gesamt
Punkte	9	6	14	8	11	12	60
Erreicht							

# Aufgabe 1 Lerntheorie & Unüberwachtes Lernen \_\_\_\_\_/9 Punkte

a)

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



Die Hypothesen  $h_1$  und  $h_2$  aus der Gleichung sind nicht Teil eines Versionsraums, da sie nicht konsistent und vollständig sind.

JL. PY

Welche von ihnen ist **konsistent** aber nicht **vollständig**?

Welche ist **vollständig** aber nicht **konsistent**?

Zeichnen sie eine vollständige und konsistente Hypothese in das Schaubild ein.

Welche Eigenschaft ergibt sich für den Versionsraum eines linearen Modells mit linearem Hypthesenraum  $H_{lin}$  für dieses Schaubild?

Hypothese ist **konsistent** und **nicht vollständig**:

 $h_2$ 

Hypothese ist **nicht konsistent** und **vollständig**:

 $h_1$ 

Eigenschaft  $H_{lin}$ :

*consistent and complete*

*Search space of the separation hypersphere*

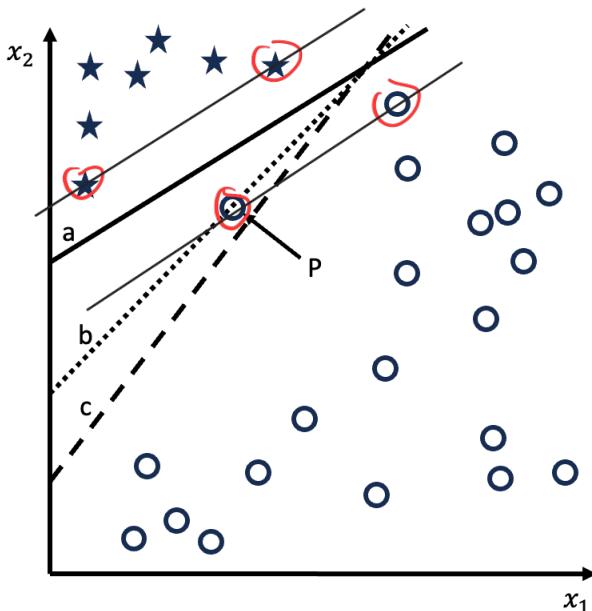
*is null*

- b) Sie haben einen gelabelten Datensatz. Sie teilen ihn auf in einen Test- und Trainingsdatensatz. (\_\_\_/2P)
- Sie trainieren sehr viele unterschiedliche Modelle auf dem Trainingdatensatz und evaluieren auf dem Testdatensatz. Sie verwenden immer die besten Modelle des Testdatensatzes um sie weiter zu verbessern. Außerdem verwenden sie den Testdatensatz um Early-Stopping zu implementieren.
- Wenn sie die Modelle in der echten Welt einsetzen, fällt ihnen auf, dass das beste Modell des Testdatensatzes nicht die beste Genauigkeit liefert sondern ein Modell, das im Testdatensatz schlechter abgeschnitten hatte. Was könnten die Gründe dafür sein? GPT

- ~~1. Overfitting to the test dataset~~ Data-dependent
- 2. Data leakage ~~not correctly split~~ the data
- ~~3. Early-stopping on test data~~
- 4. Unrepresentative test data
- 5. Randomness

- c) Sie verwenden K-means um ihre Daten zu clustern, aber Ihnen fällt auf, dass die Cluster nur sehr schlecht die zugrundeliegenden Klassen abbilden. Nennen Sie drei Fehler- bzw. Problemquellen die dazu führen könnten (\_\_\_/3P) 08. 113. fSpz

Wrong number of initial clusters  
 Nonlinear structures Curse of dimensionality  
 Sensitivity to Outliers.  
The distribution of the data is not suitable to use k-means.

**Aufgabe 2 Support Vector Machine**      /6 Punkte

Gegeben ist ein Datensatz mit den zwei Klassen Stern und Kreis. Dieser soll mit einer linearen SVM korrekt klassifiziert werden. In der oberen Abbildung finden Sie eine grafische Darstellung der Datenpunkte und verschiedenen Hypothesen im Musterraum/Merkmalsraum.

7, 108

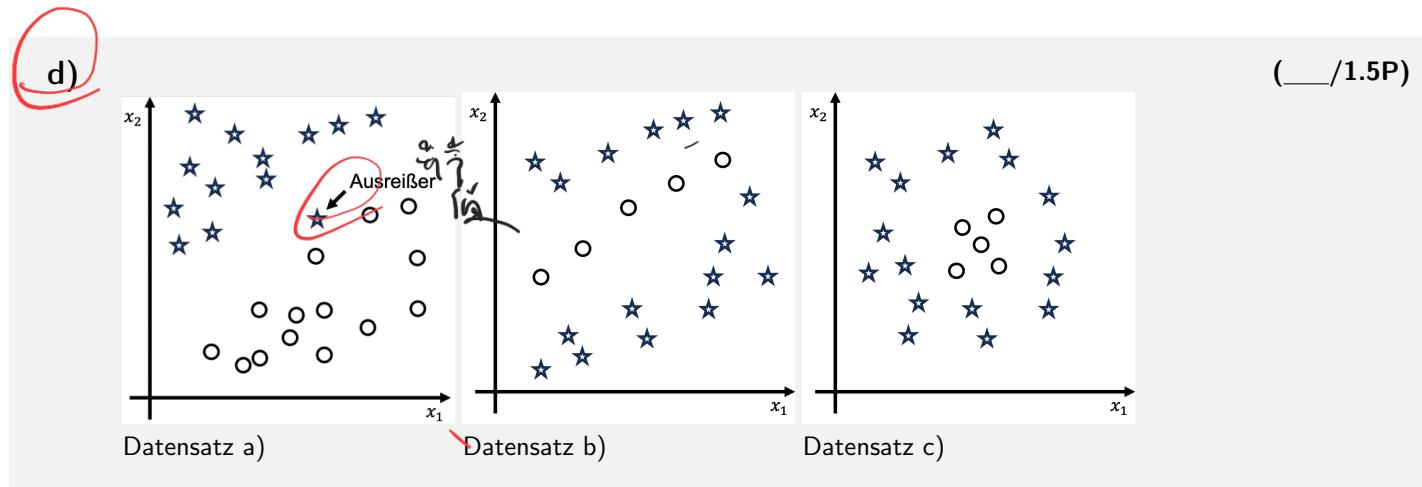
- a) Geben Sie an welche der Hypothesen (a, b, c) das optimale Ergebnis des SVM (      /1P) Algorithmus auf diesem Datensatz ist.

a

- b) Markieren Sie die 4 Stützvektoren (Support Vectors) in der oberen Abbildung durch Einkreisen. (Info: Falsch markierte Stützvektoren führen zu Punkteverlust. Streichen Sie falsche Ergebnisse durch.) (      /2P)

- c) Wie verändert sich das Ergebnis aus a), wenn der Datenpunkt P aus dem Datensatz entfernt wird? (      /0,5P)

*After removed p the hyperplane will change  
it will be close to b*



Zu sehen sind drei verschiedene Datensätze (a, b, c). Geben Sie für jeden Datensatz an, welche SVM Methode (lineare SVM, soft margin SVM, nicht-lineare SVM) sich für diesen am besten eignet.

- a: soft margin SVM  
 b: nicht-lineare SVM  
 c: nicht-lineare SVM

e) Erläutern Sie kurz den Kernel-Trick und welcher Vorteil sich daraus ergibt. ( \_\_\_ /1P)

07. P28

Instead of calculating the complex nonlinear transformation  $\phi(\cdot)$  and the scalar product, we use a kernel that does this implicitly, thus saving a lot of computational operations.

**Aufgabe 3 Neuronale Netze****\_\_\_\_\_ /14 Punkte**

- a) Optimierungsmethoden zweiter Ordnung sollten den Loss besser minimieren. Dennoch (\_\_\_\_/2P) werden sie bei neuronalen Netzen nur sehr selten verwendet. Wieso?

Nennen Sie zusätzlich eine Optimierungsmethode erster Ordnung, die eine Optimierungsmethode zweiter Ordnung approximiert.

05 PB

Because second-order optimization methods need to calculate the Hessian matrix, which is quadratic and demands a substantial amount of memory.

Quasi-Newton method Adam

- b) Geben Sie die in der Vorlesung vorgestellte ReLU Funktion und deren Ableitung an. (\_\_\_\_/1P)

04, 141

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

$$\frac{df}{dx} = \begin{cases} 1 & x > 0 \\ 0, & \text{else} \end{cases}$$

- c) Was ist der Vorteil von ReLU gegenüber einer Sigmoid- oder Tanh- (\_\_\_\_/1.5P) Aktivierungsfunktion?

Die genannten Aktivierungsfunktionen sind nicht-linear, was würde passieren wenn Sie nur lineare Aktivierungsfunktionen für ein neuronales Netz verwenden würden? 5pt

No saturation behavior for  $x > 0$ , very easy to calculate  
 A network with only linear activation functions  
 wouldn't be able to capture complex  
 non-linear relationships in the data

*The augmentation should be realistic and must not change the meaning of the original data*

- d) Wozu wird Datenaugmentierung durchgeführt und was ist dabei zu beachten? Nennen ( \_\_\_/2P)  
Sie drei Beispiele wie man Bilddateien augmentieren kann.

05 P47 P48

Reduce overfitting

increase the size and diversity  
of the dataset

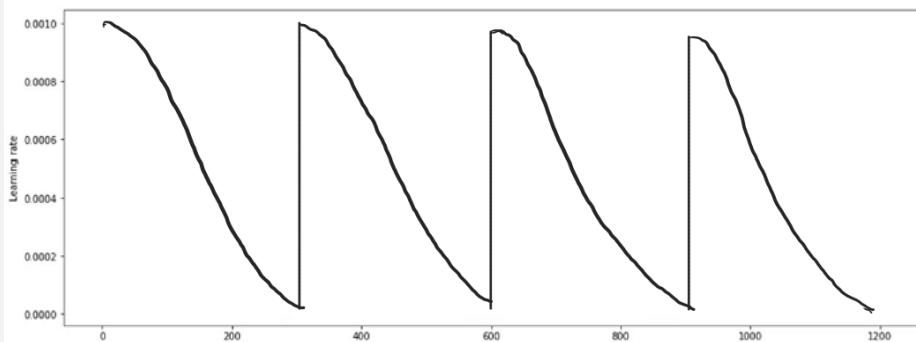
Brightness

Data Augmentation must be  
adapted to the training  
dataset

Rotation

Distortion

- e)
1. Warum werden meist dynamische Lernraten statt statischen Lernraten verwendet? ( \_\_\_/2P)
  2. Welchen Vorteil können zyklische Lernraten gegenüber monoton absinkenden Reduktionsverfahren haben? ( \_\_\_/2P)
  3. Skizzieren Sie den Verlauf einer zyklischen Lernrate mit Cosinus-Reduktion (cosine-annealing)



05 P29

1. Because dynamic learning rates allow for adjustments to the learning process during training, improving the model's convergence
2. Better escaping local minima  
Avoidance of too rapid convergence

- f) Gegeben ist ein Neuron mit Inputvektor  $\vec{x}$ , Gewichten  $\vec{w}$  und dem Bias  $b$ . Das Neuron verwendet eine Sigmoid Aktivierungsfunktion  $\sigma()$  und gibt für die Eingabe  $\vec{x}$ , die Ausgabe  $a$  aus. Führen Sie einen Backpropagation-Schritt mit der Fehlerfunktion  $L$  und Label  $\hat{y}$  durch. (\_\_\_/4P)

Errechnen Sie die Gradienten der Gewichte und des Bias  $\frac{\partial L}{\partial w_0}$ ,  $\frac{\partial L}{\partial w_1}$  und  $\frac{\partial L}{\partial b}$ .

Geben Sie zusätzlich die Zwischenergebnisse  $\frac{\partial L}{\partial a}$ ,  $\frac{\partial a}{\partial z}$ ,  $\frac{\partial z}{\partial w_0}$ ,  $\frac{\partial z}{\partial w_1}$  und  $\frac{\partial z}{\partial b}$  an.

Eingabevektor  $\vec{x} = [-3, 2]$ , Gewicht  $\vec{w} = \left[ \frac{1}{5}, \frac{4}{5} \right]$  und Bias  $b = \frac{2}{5}$  und Label  $\hat{y} = 1$ .

Als Hilfestellung geben wir Ihnen die Neuronenformel an und die jeweiligen Ausgaben  $a$ , und  $z$ . Weiter geben wir die Ableitung der Sigmoid Funktion an, sodass keine explizite Sigmoid-Berechnung notwendig ist.

$$\begin{aligned} z &= (w_0 x_0 + w_1 x_1) + b & z &= \sum_k w_k \cdot x_k + b \\ && a &= \sigma(z) \\ && L &= \frac{1}{2} (\hat{y} - a)^2 \\ &z = \frac{7}{5}, \quad \sigma(z) = a = 0.8, \quad \sigma'(z) = \sigma(z) \cdot (1 - \sigma(z)) & \frac{\partial f}{\partial x} &= f(x)(1-f(x)) \end{aligned}$$

$$\frac{\partial L}{\partial a} = -(1 - a) = -(1 - 0.8) = -0.2$$

$$\frac{\partial a}{\partial z} = \sigma'(z) \cdot (1 - \sigma(z)) = 0.8 \cdot 0.2 = 0.16$$

$$\frac{\partial z}{\partial w_0} = x_0 = -3$$

$$\frac{\partial z}{\partial w_1} = x_1 = 2$$

$$\frac{\partial z}{\partial b} = 1$$

$$\frac{\partial L}{\partial w_0} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w_0} = -0.2 \cdot 0.16 \cdot (-3) = 0.096$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w_1} = -0.2 \cdot 0.16 \cdot 2 = -0.064$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial b} = -0.2 \cdot 0.16 \cdot 1 = -0.032$$

- g) Dasselbe Neuron aus der vorherigen Aufgabe (Gewicht  $\vec{w} = \left[ \frac{1}{5}, \frac{4}{5} \right]$  und Bias  $b = \frac{2}{5}$ ) (\_\_\_/1.5P)  
 erhält nun eine andere Eingabe und eine andere Bezeichnung. Dabei errechnete es im Backpropagation-Schritt folgende Gradienten:

$$\frac{\partial L}{\partial w_0} = -1$$

$$\frac{\partial L}{\partial w_1} = 2$$

$$\frac{\partial L}{\partial b} = 1$$

Führen Sie mit den vorgegebenen Gradienten, der Lernrate  $\eta = \frac{1}{5}$  und den Gewichten und Bias aus der vorhergegangenen Aufgabe einen Gewichts-Update-Schritt durch.  
 Geben Sie  $w_0^{n+1}$  und  $b^{n+1}$  an.

$$w_0^{n+1} = w_0^n + (-y \cdot \frac{\partial L}{\partial w_0^n}) = \frac{1}{5} - \frac{1}{5} \times (-1) = \frac{2}{5}$$

$$w_1^{n+1} = w_1^n + (-y \cdot \frac{\partial L}{\partial w_1^n}) = \frac{4}{5} - \frac{1}{5} \times 2 = \frac{2}{5}$$

$$b^{n+1} = b^n + (-y \cdot \frac{\partial L}{\partial b^n}) = \frac{2}{5} - \frac{1}{5} \times 1 = \frac{1}{5}$$

**Aufgabe 4 Convolutional Neural Networks**      /8 Punkte

- a) Gegeben ist ein Input-Bild mit den Dimensionen  $32 \times 24 \times 3$  ( $H \times W \times C$ ). (\_\_\_\_/3P)

Gewünscht ist eine Output-Featuremap mit der Dimension  $24 \times 22 \times 4$ .

Geben Sie die Dimensionen  $H_k \times W_k \times C_{in} \times C_{out}$  eines Convolutional Layers an, das in Kombination mit einem Padding von je 1px auf allen Seiten, und einem Stride von 1px die gewünschten Output-Dimensionen erzeugt.

06 PL

$$M_{new} = \frac{(H - W_k)}{S} + 1 \quad 24 = \frac{32 + 2 - W_k}{1} + 1 \quad W_k = 1$$

$$N_{new} = \frac{(W - W_k)}{S} + 1 \quad 22 = \frac{24 + 1 - W_k}{1} + 1 \quad W_k = 5$$

~~11x5x3x4~~

- b) Berechnen Sie die folgende Convolution mit Stride=2 und ohne Padding. (\_\_\_\_/2P)

Input (4x4x1)

1	1	0	2
2	0	1	2
-3	1	2	1
1	1	0	0

Kernel (2x2x1x1)

\*

1	-1
-1	1

=

$$\begin{aligned}
 1 \times 1 + 1 \times (-1) + 2 \times (-1) + 0 \times 1 &= -2 \\
 0 \times 1 + 2 \times (-1) + 1 \times (-1) + 2 \times 1 &= -1 \\
 -3 \times 1 + 1 \times (-1) + 1 \times (-1) + 1 \times 1 &= -4 \\
 2 \times 1 + 1 \times (-1) + 0 \times (-1) &= 1
 \end{aligned}$$

-2	-1
-4	1

c) Bei CNNs ist die Größe der Kernel ein wichtiger Hyperparameter.

06.144 /3P

1. Welche Vorteile haben  $3 \times 3$  Kernel gegenüber größeren z.B.  $7 \times 7$  Kernel hinsichtlich des rezeptiven Felds?
2. Wie verhält sich das rezeptive Feld bei CNNs die nur  $1 \times 1$  Kernels verwenden?
3. Die Ausgabe eines  $3 \times 3$  Kernels hat eine kleinere Auflösung als seine Eingabe. Erklären Sie eine Art von Padding und eine dazu passende Größe, um dies zu verhindern.

GPT

$3 \times 3$  kernels contribute to building a larger receptive field, aiding in capturing complex patterns without significantly increasing the number of parameters.

the receptive field remains unchanged

when applying a  $3 \times 3$  kernel to an input with Zero Padding of size 1, a row and a column of zeros are added around each side of the input. This maintains the output size, as the kernel now interacts with the newly added zeros at the borders.

Same Padding which maintaining the feature map's size

**Aufgabe 5 Reinforcement Learning****\_\_\_\_\_ /11 Punkte**

- a) Was ist das primäre Ziel von Reinforcement Learning? Geben Sie eine kurze Beschreibung an. (\_\_\_\_/1P)

11, P5

find a policy

to Maximize rewards  $\sum_t r^t R(s_t, a_t)$ 

- b) Was ist die Hauptbedingung, die ein Prozess erfüllen muss, um als Markov-Entscheidungsprozess bezeichnet zu werden? (\_\_\_\_/1P)

11, P16

$$p(s_{t+1} | s_t, a_t) = p(s_{t+1} | s_0, a_0, \dots, s_t, a_t)$$

12, P16

- c) Definieren Sie kurz den Begriff „modellfrei“ im Kontext von Reinforcement Learning. (\_\_\_\_/1P)

No knowledge of MDP  
 Without the true model of the environment  
 the agent interacts with the  
 environment to gather experience

Samples,

- d) Warum werden Funktionsapproximatoren (wie z.B. neuronale Netze) in RL benötigt? (  /1P)  
 Geben Sie dazu auch ein kurzes Beispiel an, in dem ein Funktionsapproximator benutzt werden muss.

p42

1L. p 33

Because RL is often applied in more complex problems and impossible to store all values in a table. With approximation can value functions interpolate between different inputs.  
 Game Go :  $10^{170}$  states

- e) Ein RL-Agent hat in einer neuen Umgebung zwei Trajektorien gesammelt, die alle von einem Anfangszustand  $s_0$  ausgehen. Dabei wurde eine Policy  $\pi'$  verwendet.

- In der ersten Trajektorie hat der Agent folgende Belohnungen erhalten:  
 $\{r_0 = 0, r_1 = 3, r_2 = -2, r_3 = 4, r_4 = 1\}$ .
- In der zweiten Trajektorie hat der Agent folgende Belohnungen erhalten:  
 $\{r_0 = 0, r_1 = 2, r_2 = -2, r_3 = 3, r_4 = 3\}$

Berechnen Sie den diskontierten Gewinn für jede dieser Trajektorien  $\{G_1, G_2\}$  unter der Annahme, dass  $\gamma = 0,8$  ist. Evaluieren Sie anschließend die Policy  $\pi'$  im Zustand  $s_0$  mithilfe der Zustandswertfunktion  $V_{\pi'}(s_0)$  anhand der gesammelten Trajektorien.

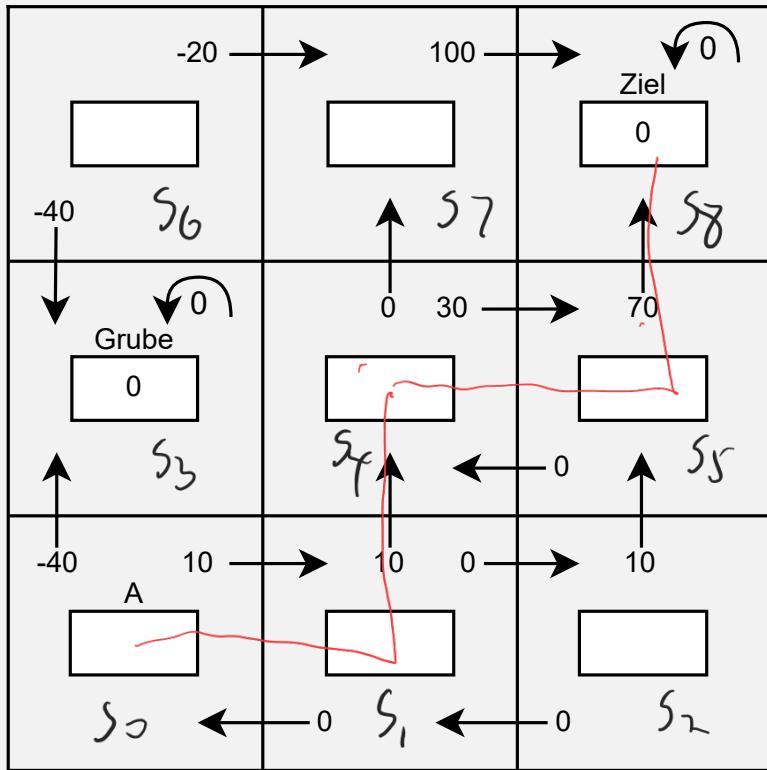
Hinweis: Richtige Gleichungen reichen für die volle Punktzahl.

$$\begin{aligned} G_1 &= r_0 + \gamma \cdot V_{\pi'}(r_1) + \gamma^2 \cdot r_2 + \gamma^3 \cdot r_3 + \gamma^4 \cdot r_4 \\ &= 0 + 0.64 \times 3 + 0.64^2 \times (-2) + 0.64^3 \times 4 + 0.64^4 \times 1 = 3.578 \end{aligned}$$

$$G_2 = 0 + 0.64 + 0.64^2 \times (-2) + 0.64^3 \times 3 + 0.64^4 \times 1 = 3.088$$

$$\hat{V}_{\pi'}(s_0) = \frac{G_1 + G_2}{2} = 3.331$$

- f) Betrachten Sie die untenstehende Welt. Ein Agent kann sich mit den angezeigten Aktionen (Pfeilen) von Zelle zu Zelle bewegen. Die Belohnung für eine Aktion entspricht der Zahl an dem entsprechenden Pfeil. Nehmen Sie an, dass die optimale Strategie gelernt wurde. (\_\_\_/4,5P)



f4

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.

$$V^*(s_3) = 0 \quad V^*(s_7) = 0$$

$$V^*(s_1) = 100 + 0.8 \times 0 = 100$$

$$V^*(s_0) = -20 + 0.8 \times 100 = 60$$

$$V^*(s_5) = 70 + 0.8 \times 0 = 70$$

$$V^*(s_4) = 30 + 0.8 \times 70 = 86$$

$$= 86$$

$$V^*(s_6) = 10 + 0.8 \times 86 = 79$$

$$V^*(s_2) = 10 + 0.8 \times 79 = 66$$

$$V^*(s_8) = 10 + 0.8 \times 66 = 74$$

$s_0 \quad s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \quad s_7 \quad s_8$   
right up up [ ] right up right right [ ]

## Aufgabe 6 HMM, Bayes, Entscheidungsbäume \_\_\_\_\_/12 Punkte

- a) Nennen Sie zwei Beispielsanwendungen aus der Vorlesung, die mithilfe eines Hidden Markov Models gelöst werden können. (\_\_\_\_/1P)

- b) Welche der folgenden Eigenschaften treffen auf HMMs zu? (\_\_\_\_/2P)

1. Vollständig beobachtbare Zustände
2. Nicht vollständig beobachtbare Zustände
3. Zeitinvariantes Modell
4. Zeitvariantes Modell
5. Beschränkter zeitlicher Horizont
6. Unendlicher zeitlicher Horizont
7. Doppelt-Stochastischer Prozess

- c) Zur Vorhersage von Verspätungen im Fernverkehr soll ein Naive Bayes Klassifikator eingesetzt werden. Die vorauszusagende Verbindung fährt bei gutem Wetter ( $W = \text{Gut}$ ) und am Tag ( $T = \text{Tag}$ ). Welche Klassifikation ist gemäß des Naive Bayes Ansatzes am wahrscheinlichsten? Geben Sie die Formel, den Rechenweg, sowie Ihre Schlussfolgerung an. (\_\_\_/3P)

Folgende Daten sind gegeben. Grau hinterlegte Einträge dienen zur Vereinfachung beim Ablesen und markieren die eingetretenen Verspätungen ( $V = \text{Ja}$ ).

Nr.	Wetter (W)	Tageszeit (T)	Verspätung (V)
1	Schlecht	Nacht	Nein
2	Schlecht	Nacht	Nein
3	Gut	Tag	Nein
4	Schlecht	Tag	Nein
5	Gut	Nacht	Nein
6	Gut	Tag	Ja
7	Schlecht	Tag	Ja
8	Schlecht	Nacht	Ja
9	Gut	Nacht	Ja
10	Schlecht	Tag	Ja

$$p(V=J_a) = \frac{1}{2} \quad p(V=N_{o,n}) = \frac{1}{2}$$

$$p(V=J_a) \cdot p(W=Gut | V=J_a) \cdot p(T=Tag | V=J_a)$$

$$= \frac{1}{2} \times \frac{3}{8} \times \frac{3}{5} = \frac{3}{25}$$

$$p(V=N) \cdot p(W=G | V=N) \cdot p(T=Tag | V=N)$$

$$= \frac{1}{2} \times \frac{5}{8} \times \frac{2}{5} = \frac{2}{25}$$

$$\frac{3}{25} > \frac{2}{25}$$

Classification:  $V=J_a$

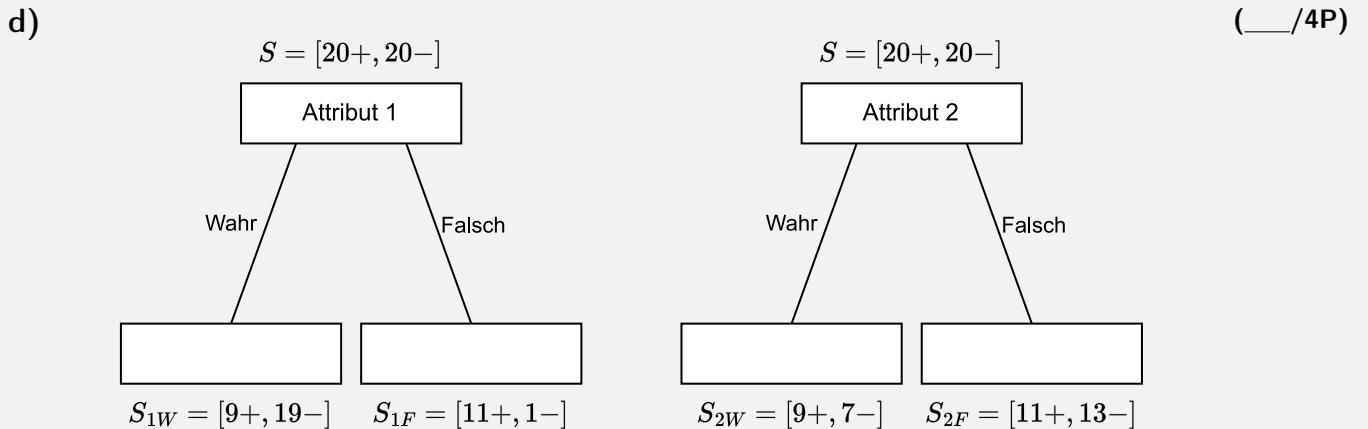


Abbildung 1: Schreibweise: [Anzahl positive Bsp. (+), Anzahl negative Bsp. (-)]

In der oben dargestellten Abbildung finden Sie den Vergleich zweier Attribute  $A_1$  (links) und  $A_2$  (rechts), die als Testattribute eines Entscheidungsbaumes ausgewählt werden können. Welches der Attribute eignet sich ausgehend vom Informationsgewinn/Informationgain besser als Entscheidungskriterium? Begründen Sie Ihre Antwort rechnerisch und runden Sie auf zwei Nachkommastellen.

Hilfestellung: Formel für den Informationsgewinn und die Entropie.

$$IG(S, A) = H(S) - \sum_{v \in V(A)} \frac{|S_v|}{|S|} H(S_v)$$

$$H(S) = - \sum_{i=1}^K p(y_i) \log_2 p(y_i)$$

Attribute 1:  $H(S) = -\frac{10}{45} \log_2 \frac{10}{45} - \frac{35}{45} \log_2 \frac{35}{45} = 1$

$H(S_{1W}) = -\frac{7}{28} \log_2 \frac{7}{28} - \frac{21}{28} \log_2 \frac{21}{28} = 0.91$

$H(S_{1F}) = -\frac{11}{12} \log_2 \frac{11}{12} - \frac{1}{12} \log_2 \frac{1}{12} = 0.41$

$IG(S, A_1) = 1 - \frac{28}{45} \times 0.91 - \frac{12}{45} \times 0.41 = 0.24$

Attribute 2:  $H(S) = 1$

$H(S_{2W}) = -\frac{9}{16} \log_2 \frac{9}{16} - \frac{7}{16} \log_2 \frac{7}{16} = 0.99$

$H(S_{2F}) = -\frac{11}{24} \log_2 \frac{11}{24} - \frac{13}{24} \log_2 \frac{13}{24} = 0.99$

$IG(S, A_2) = 1 - \frac{16}{45} \times 0.99 - \frac{24}{45} \times 0.99 = 0.01$

$$0.24 > 0.01$$

$A_1$  besser als <sup>18</sup> Entscheidungskriterium

- e) Nennen Sie **zwei** Vorteile von Entscheidungsbäumen gegenüber neuronalen Netzen. (\_\_\_/1P)

3 16

Interpretability      Non-parametric  
Simple to learn non-linear problems  
Generally low computing resource requirements

- f) Gegeben ist Bayes' Theorem. Geben Sie die Namen der Terme  $P(A)$  und  $P(A|B)$  an. (\_\_\_/1P)

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

10. 19

$P(A)$  : prior probability

$P(A|B)$  : posterior probability

