

Aufgabe	1	2	3	4	5	6	Gesamt	Note
Punkte	7	12	14	11	8	8	60	
Erreicht								

## Aufgabe 1

(7 Punkte)

- (a) Geben Sie 3 Einordnungskriterien von Lernverfahren, sowie deren Abstufungen bzw. Ausprägungen an. (\_\_\_\_/3P)

Supervised learning : Examples with input and desired output data

Unsupervised learning : Examples contain only input data

Reinforcement learning : Experience from interaction with an environment and resulting reward

- (b) Beschreiben Sie das Problem des **Overfitting**. Wie kann es verhindert werden? (\_\_\_\_/2P)

Initially both errors decrease. After some iterations, training error keeps decreasing while test error increase.

Early stopping disjoint

- (c) Was verbirgt sich hinter der Vapnik-Chervonenkis (VC) Dimension? Beschreiben Sie kurz den theoretischen Nutzen in der Anwendung. (\_\_\_\_/2P)

Measure of data complexity of learning

Measure of the capacity of learning systems.

huanb)

## Aufgabe 2: Lernen von probabilistischen Modellen

(12 Punkte)

- (a) Der folgende Datensatz beschreibt Beobachtungen des Status eines bestimmten Zuges an einem bestimmten Bahnhof gegeben der ebenfalls beobachteten Attribute:

Tag ( $T$ ) = {Wochentag, Wochenende}, Wind ( $W_i$ ) = {Kein, Leicht, Stark} und Wetter ( $We$ ) = {Sonne, Regen, Nebel}.

Tag ( $T$ )	Wind ( $W_i$ )	Wetter ( $We$ )	Status ( $S$ )
Wochentag	Kein	Sonne	Verspätet
Wochentag	Kein	Nebel	Verspätet
Wochentag	Stark	Regen	Pünktlich
Wochenende	Leicht	Regen	Verspätet
Wochentag	Stark	Sonne	Pünktlich
Wochenende	Stark	Nebel	Pünktlich
Wochenende	Kein	Sonne	Verspätet
Wochentag	Leicht	Regen	Pünktlich
Wochenende	Leicht	Nebel	Verspätet

Berechnen Sie folgenden (a-priori und bedingte) Wahrscheinlichkeiten: (\_\_\_/2P)

$$P(S = \text{Verspätet}) = \frac{5}{9}$$

$$P(S = \text{Pünktlich}) = \frac{4}{9}$$

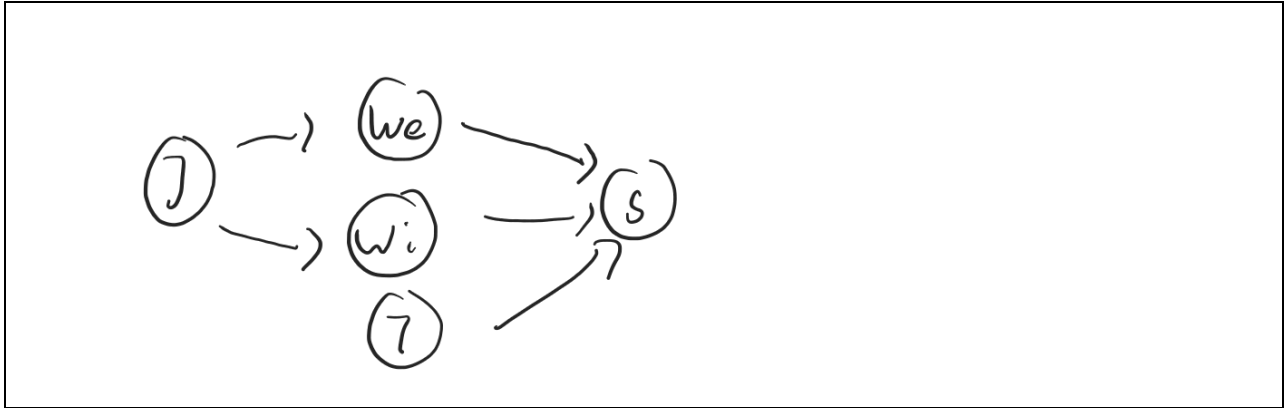
$$P(T = \text{Wochenende} | S = \text{Pünktlich}) = \frac{\frac{1}{9}}{\frac{4}{9}} = \frac{1}{4}$$

Heute ist ein Wochentag mit Nebel und Windstärke leicht. Ein naiver Bayes-Klassifikator soll genutzt werden, um den Status des Zuges vorherzusagen. Welches ist heute der wahrscheinlichere Status des Zuges? Begründen Sie Ihre Entscheidung formal.

(\_\_\_/2P)

$$\begin{aligned}
 &P(S=V) \cdot P(T=Wt | S=V) \cdot P(W_i=L | S=V) \cdot P(We=N | S=V) \\
 &= \frac{5}{9} \times \frac{2}{5} \times \frac{2}{8} \times \frac{2}{5} = \frac{8}{225} \\
 &P(S=P) \cdot P(T=Wt | S=P) \cdot P(W_i=L | S=P) \cdot P(We=N | S=P) \\
 &= \frac{4}{9} \times \frac{3}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{3}{96} = \frac{1}{48} < \frac{8}{225} \\
 &\quad S = \text{Verspätet}
 \end{aligned}$$

- (b) Das Szenario aus der vorherigen Teilaufgabe wird nun zusätzlich um das Attribut Jahreszeit ( $J$ ) erweitert. Folgende Abhängigkeiten zwischen den Attributen sind gegeben:
- Wetter ( $We$ ) und Wind ( $Wi$ ) sind abhängig von der Jahreszeit ( $J$ )
  - Der Status ( $S$ ) ist abhängig von Wetter ( $We$ ), Wind ( $Wi$ ) und Tag ( $T$ )
- Zeichnen Sie ein Bayes'sches Netz, welches das Szenario beschreibt. (\_\_\_/2P)



- (c) Was kann bei einem Bayes'schen Netz gelernt werden?  
Mit welcher Methode erfolgt dies, wenn die Struktur bekannt ist und Variablen nur teilweise beobachtbar sind? (\_\_\_/2P)

Describes conditional dependencies with respect to subset of variables.  
Gradient ascent, expectation-maximization algorithm

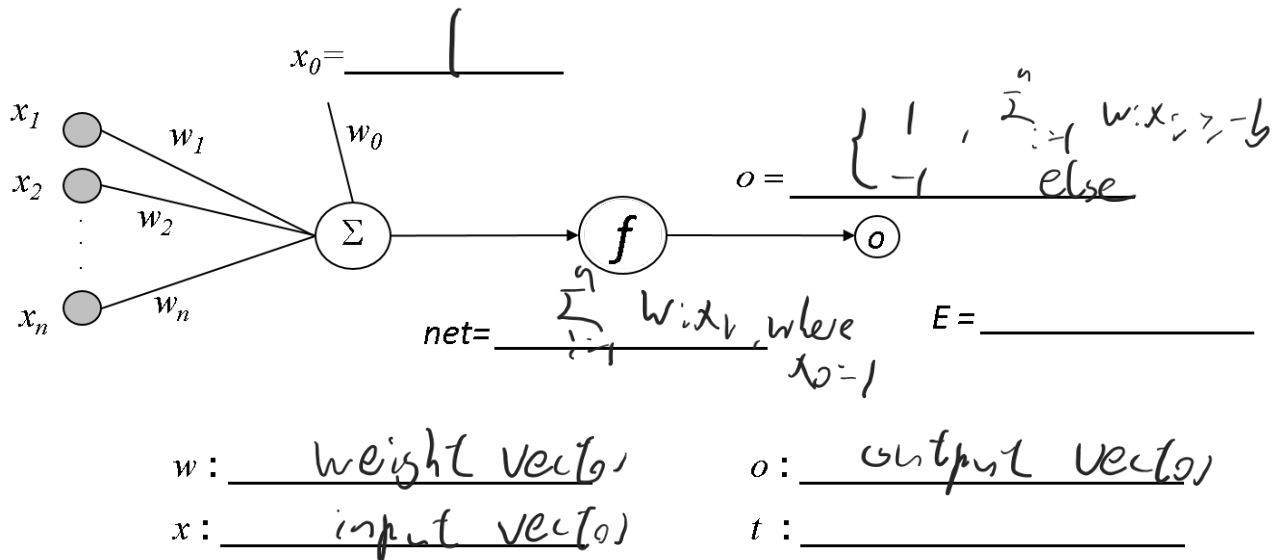
- (d) Ein HMM (Hidden Markov Modell) ist definiert als:

$\lambda = \{S - \text{Zustände}, V - \text{Ausgabezeichen}, A - \text{Übergangswahrscheinlichkeiten}, B - \text{Emmisionswahrscheinlichkeiten}, \Pi - \text{Verteilung Anfangswahrscheinlichkeiten}\}.$

Beschreiben sie das Lernproblem (was ist gegeben, was gesucht)? Welcher Lernansatz kann dafür verwendet werden und welche Parameter werden dabei gelernt? (\_\_\_/4P)

**Aufgabe 3: Neuronale Netze****(14 Punkte)**

- (a) Ergänzen Sie fehlende Begriffe und Formeln in der untenstehenden Abbildung eines Neurons. (\_\_\_\_/4P)



- (b) Nennen Sie zwei typische nichtlineare Aktivierungsfunktionen, die jeweils dazugehörige Formel und die jeweiligen Ableitungen. (\_\_\_\_/3P)

ReLU	$f(x) = \max(0, x)$	$f'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$
Leaky ReLU	$f(x) = \begin{cases} x & x > 0 \\ \alpha x & x \leq 0 \end{cases}$	$f'(x) = \begin{cases} 1 & x > 0 \\ \alpha & \text{else} \end{cases}$

- (c) Welche Bedingung muss die Aktivierungsfunktion eines Neurons für das Lernen unter Verwendung des Gradientenabstiegs erfüllen? (\_\_\_\_/1P)

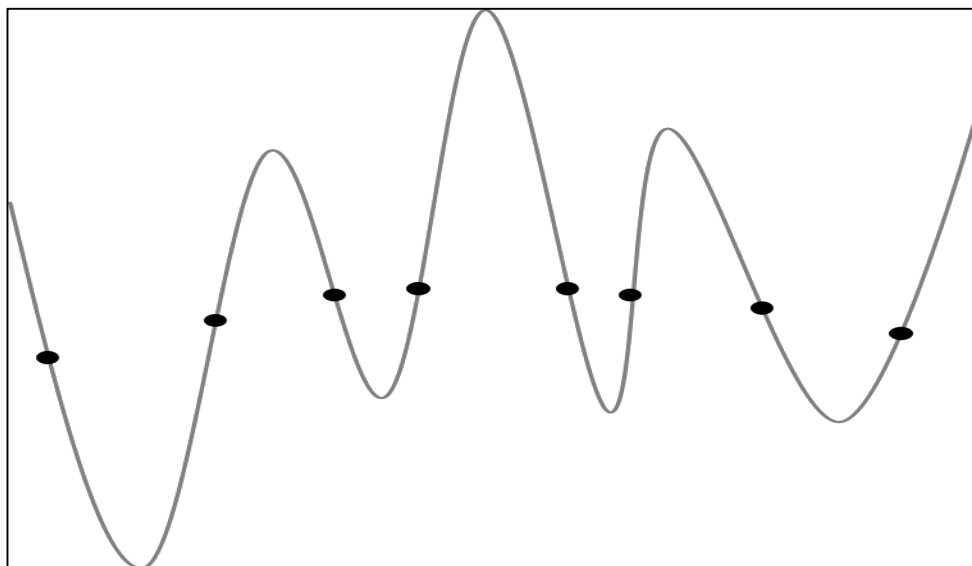
Derivative of function must exist.

- (d) Nennen Sie ein Lernverfahren, um vorwärts gerichtete, mehrschichtige neuronale Netze zu trainieren und eine Herausforderung, die beim involvierten Gradientenabstieg auftreten kann. Mit welcher Methode kann dabei das Lernen verbessert werden? (\_\_\_\_/2P)

Backpropagation algorithm

SGD

- (e) Was lässt sich über die VC-Dimension des neuronalen Netzes sagen, das aus den untenstehenden Lerndaten die eingezeichnete Kurve approximiert? Nennen Sie das resultierende Phänomen und beschreiben Sie kurz, wie sich die Approximation verbessern lässt. (\_\_\_/3P)



System with high capacity. Hypothesis space spanned by many parameters can sometimes be calculated with VC dimension.  
Overfitting. Greedy choice and search for optimal hypothesis

- (f) Nennen Sie ein konstruktives Verfahren zur schrittweisen Optimierung der Netzwerktopologie. (\_\_\_/1P)

Growing Neural Gas

## Aufgabe 4: Reinforcement Learning

(11 Punkte)

- (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) Beschreiben sie kurz die Markov-Bedingung (\_\_\_/1P)

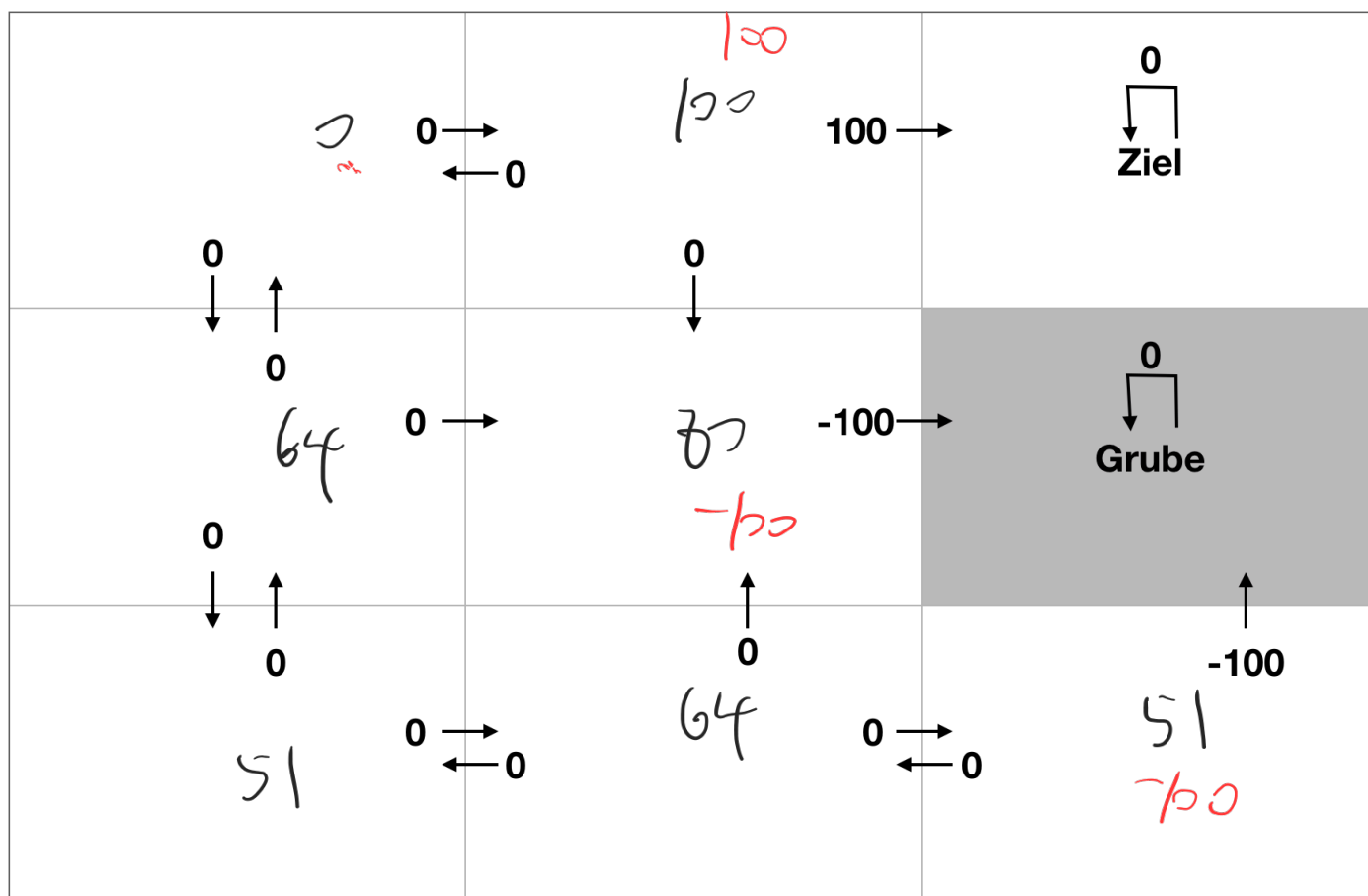
$$P(S_{t+1} | S_t, a_t) = P(S_{t+1} | s_0, a_0, \dots, S_t, a_t)$$

- (c) Wie lautet die rekursive Definition der Q-Funktion (Bellmann-Gleichung)? (\_\_\_/2P)

$$q_{\pi}(s, a) = \mathbb{E}_{\pi} [R_{t+1} + \gamma q_{\pi}(S_{t+1}, A_{t+1}) | S_t = s, A_t = a]$$

- (d) Nennen Sie die zwei wesentlichen Unterschiede zwischen den Suchstrategien *Exploration* und *Exploitation*. (\_\_\_/1P)

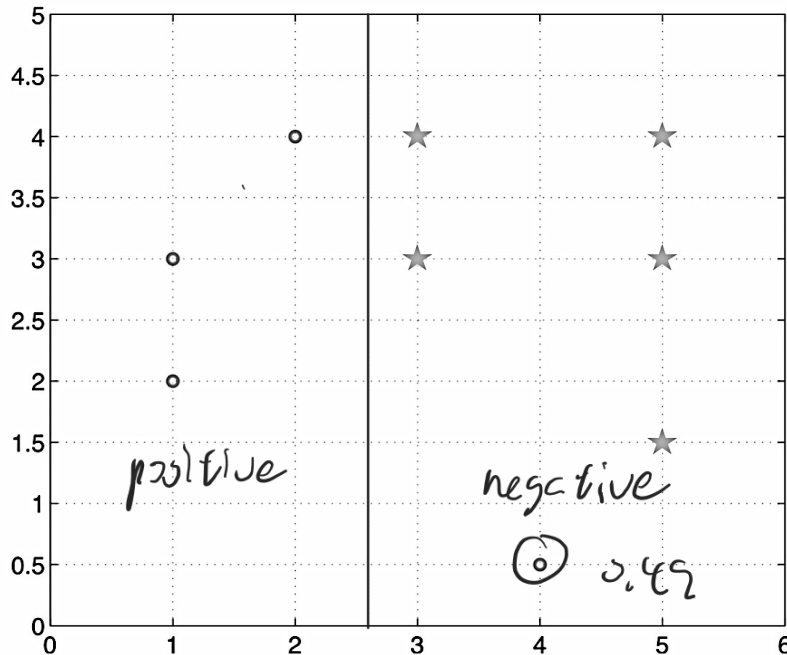
- (e) Betrachten sie das untenstehende Labyrinth. Ein Agent kann sich mit den angezeigten Zustandsübergängen von Raum zu Raum bewegen. Der Reward für einen Übergang ist jeweils in der Zeichnung abgebildet. Zu Beginn des Trainings gilt:  $\forall s, a \ Q(s, a) = 0$ . Führen sie den Q-Lernen Trainingsalgorithmus durch und tragen Sie die Schätzung  $\hat{Q}(s, a)$  für alle  $(s, a)$  nach konvergieren des Lernalgorithmus ein (Diskontierungsfaktor  $\gamma = 0,8$ ). Runden Sie die Ergebnisse auf natürliche Zahlen. (\_\_\_\_/4P)





**Aufgabe 5: Boosting****(8 Punkte)**

Mit Hilfe von AdaBoost soll eine Klassifizierung durchgeführt werden. Hierzu soll ein Entscheidungsstumpf (1-Merkmal Entscheidungsbaum mit lediglich einer Wurzel und zwei Blättern) verwendet werden. In jeder Iteration wählen Sie den Stumpf, der den gewichteten Trainingsfehler minimiert. Nutzen Sie hierzu die untenstehende Zeichnung.



- (a) Zeichnen Sie die Entscheidungsgerade in die obere Abbildung ein. Markieren Sie die positive und negative Seite der Klassifikation. (\_\_\_/1 P)
- (b) Berechnen Sie die Gewichtung jedes Datenpunktes nach der ersten Iteration und markieren Sie den Datenpunkt, der die höchste Gewichtung nach der ersten Iteration besitzt. (\_\_\_/4 P)

$$\begin{aligned}
 w_i &= \frac{1}{n} = \frac{1}{8} & \bar{E}_1 &= \frac{1}{8} & \alpha_1 &= \frac{1}{2} \ln \left( \frac{1 - \bar{E}_1}{\bar{E}_1} \right) = 1.04 \\
 w_i &\leftarrow \begin{cases} \frac{1}{8} \cdot e^{-1.04} = 0.04 & \text{if } h_1(\bar{x}_i) = y_i \\ \frac{1}{8} \cdot e^{+1.04} = 0.31 & \text{if } h_1(\bar{x}_i) \neq y_i \end{cases} \\
 \text{normalization:} & \\
 w_i &\leftarrow \begin{cases} \frac{0.04}{8 \times 0.04 + 1 \times 0.31} = 0.06 & \text{if } h_1(\bar{x}_i) = y_i \\ \frac{0.31}{8 \times 0.04 + 1 \times 0.31} = 0.49 & \text{if } h_1(\bar{x}_i) \neq y_i \end{cases}
 \end{aligned}$$

- (c) Erklären Sie anhand von AdaBoost die Anwendung der strukturellen Risikominimierung. Gehen Sie dabei auf die wesentlichen Aspekte der Hypothesenraumstrukturierung und der Fehlerminimierung ein. (\_\_\_/3P)

$$\min_{h \in \mathcal{H}} \left( \hat{L}(h) + \sqrt{\frac{V(h)}{n}} \right)$$

## Aufgabe 6: Support Vector Machines

(8 Punkte)

- (a) Beschreiben Sie kurz die Grundidee, die der Methode der Support Vektor Klassifikation zugrunde liegt und wie das Verfahren einzuzuordnen ist. (\_\_\_/2P)

Size of the margin determines the generalization capacity. Find the best separating line/hyperplane with maximum margin to the classes.

- (b) Geben Sie die Formeln für das Optimierungskriterium der **optimalen** Hyperebene und für die Randbedingung einer korrekten Klassifikation an (gegeben Trainingsbeispiele der Form  $(\vec{x}, y)$ ) (\_\_\_/2P)

Optimierungskriterium: minimize  $\|\vec{w}\|^2$

Randbedingung:  $y_i (\vec{w} \vec{x}_i + b) \geq 1$

- (c) Erklären Sie die Dualität zwischen Hypothesenraum und Merkmalsraum im Kontext des SVM Verfahrens (Version Space Duality). (\_\_\_/2P)

points in feature space correspond to hyperplanes in hypothesis space and vice versa.

- (d) Welche Vorteile ergeben sich durch den sogenannten „Kerneltrick“? Welche Beobachtung liegt dem Kerneltrick zugrunde? (\_\_\_/2P)

Saving a lot of computational operations  
kernel function

