

Aufgabe 1 – Induktives Lernen, Lerntheorie, Entscheidungsbäume und Unüberwachtes Lernen

____ / 11P

- a) Definieren Sie Konsistenz und Vollständigkeit im Hypothesenraum. (____/2P)

Consistent: No negative examples are classified positive
 Complete: All positive examples are classified positive.

- b) Was ist das Optimierungsziel beim Lernen von Entscheidungsbäumen bzgl. der Klassenverteilung? Welches Maß wird hierfür herangezogen? Wie wirkt sich dies auf die Tiefe des gelernten Baums aus? (____/1,5P)

Objective: Select attributes that minimize entropy and maximize information gain
 Model: ID3 - algorithm
 Increase the depth of decision tree

- c) Nennen Sie zwei Maßnahmen gegen Overfitting von Entscheidungsbäumen? (____/1P)

Early stopping pruning

- d) Geben Sie die Formeln für Genauigkeit (Precision) und True-Positive-Rate (Recall) an. (____/2P)

Precision: $P = \frac{TP}{TP + FP}$ TP: True positive FP: False positive

Recall: $R = \frac{TP}{TP + FN}$ FN: False negative

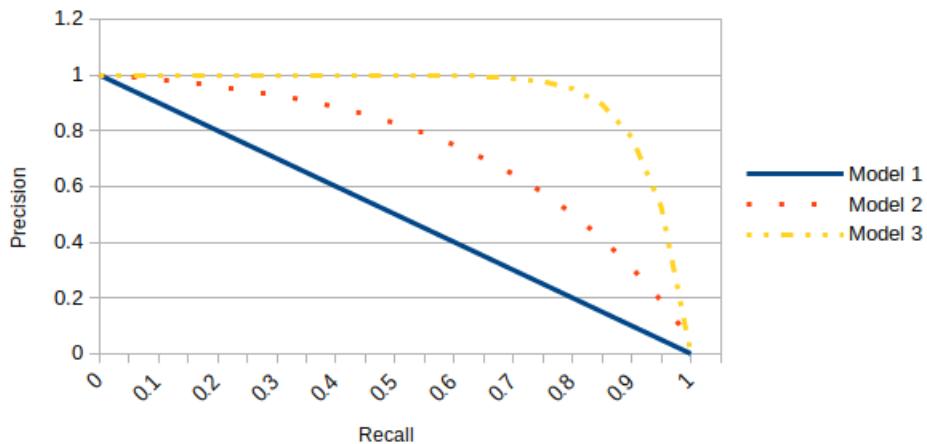
- e) Nennen Sie vier positive Eigenschaften eines Random Forests (verglichen mit einem einfachen Entscheidungsbaum). (3 P)

1. Generally, much better results than single decision tree
 2. Reduced Overfitting
 3. Outlier Handling 4. Automatic Feature Selection

- f) Welches Modell ist das Beste? Geben sie eine kurze Begründung an? (1,5P)

02. P28

Precision–Recall–Graph



Model 3. precision and recall are both as high as possible.

- g) Was ist für den k-Means Algorithmus mindestens festzulegen? (Hinweis: Parameter, Metriken) (1P)

08. P13

Number of clusters k
 Initial Cluster Centers c_j

Aufgabe 2 – Neuronale Netze___ / 7P

- a) Geben Sie die Formel und Ableitung für die Aktivierungsfunktion ReLU an.
Welchen entscheidenden Nachteil hat ReLU? Welche alternative
Aktivierungsfunktion versucht dies zu beheben?

04. P41 (2P)

$$f(x) = \max(0, x) \quad f'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

Draw: Not centered around 0
perivative vanishes at $x=0$ "Dying ReLU"

Leaky ReLU: $f(x) = \begin{cases} x & x > 0 \\ \alpha x & x \leq 0 \end{cases} \quad f'(x) = \begin{cases} 1 & x > 0 \\ \alpha & x \leq 0 \end{cases}$

- b) Was muss bei der Initialisierung der Gewichte eines Neuronalen Netzes beachtet werden?
Nennen und erläutern Sie kurz eine gängige Initialisierungsmethode.

gpt (2P)

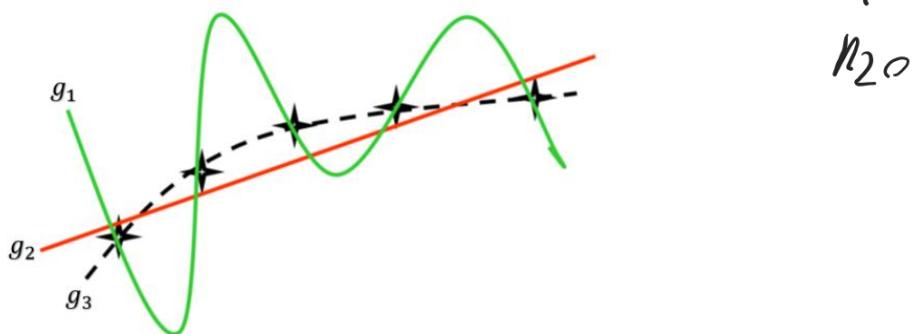
Avoiding too small or too large weights; Breaking symmetry

Method: Xavier, Optimized initialization of parameters that prevents vanishing and exploding gradients. Keep variance of activation distribution constant for each layer.

- c) Die folgende Abbildung zeigt die Kurven g_1 und g_2 , die von zwei verschiedenen neuronalen Netzen an die gleichen Trainingsdaten (Sterne) angepasst wurden. Die Kurve g_3 entspricht der zu erlernenden Funktion (Grundwahrheit). (___ /3P)

1. Nennen Sie die Fachbegriffe für das Fehlverhalten (zugrundeliegende Verhalten) von g_1 und g_2 und beschreiben Sie den Fehler in der Topologie f_3, f_h des neuronalen Netzes, der zu diesem Verhalten führen kann.

2. Skizzieren Sie den typischen Verlauf des Trainings- und Testfehlers in Abhängigkeit der Trainingszyklen für g_1 und g_2 . $O.L \neq 17$



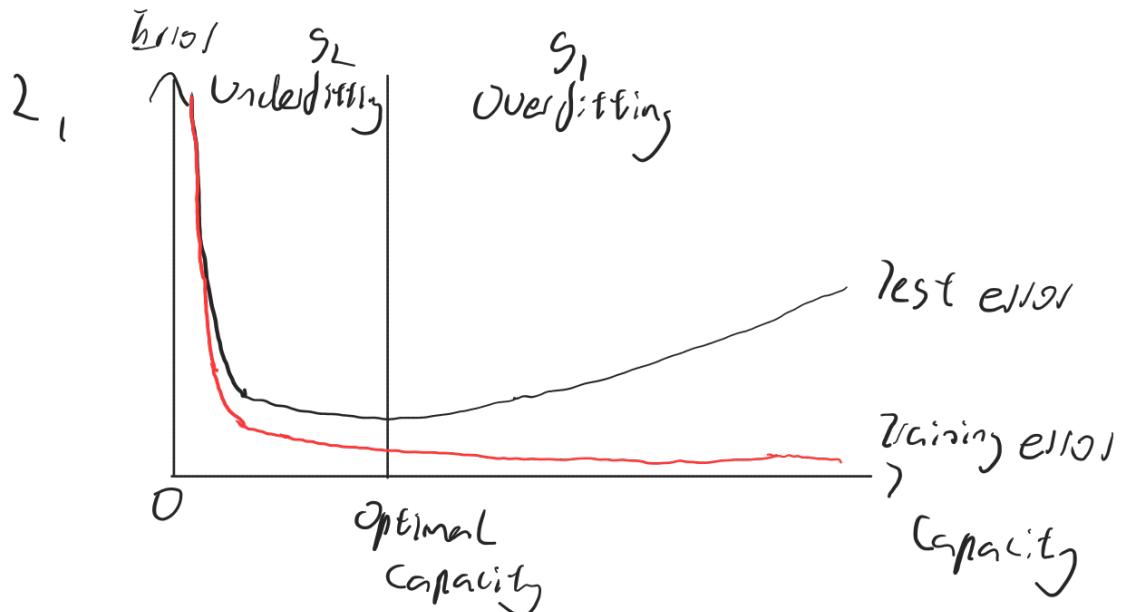
1, g_1 : Overfitting

Model capacity is too large

Model is trained for too many

g_2 : underfitting

Model capacity is iteration too small



Aufgabe 3 – CNN

____ / 9P

- a) Erklären Sie "Weight Sharing". Welche Vorteile ergeben sich hierdurch gegenüber klassischen Neuronalen Netzen? (1,5P)

"Weight Sharing": One weight of convolutional kernel is used multiple times for different inputs.

Advantages: Reduction in the Number of Parameters
Efficient Modeling; Better Generalization

Transfer Learning

Translation Invariance

- b) Nennen Sie zwei verschiedene Methoden zur Initialisierung von Gewichten. (1P)

Xavier

Kaiming He - Initialization Transfer Learning

- c) Wie funktioniert Dropout, welche Vorteile bringt es? (2P)

Dropout: Randomly deactivate neurons (and their connections) during training with probability p.

Advantages: Regularization against Overfitting
Improved Generalization

Reduced Dependency on Specific Weights

- d) Wie wird das Vanishing-Gradient Problem bei ResNet gelöst? Erläutern Sie dies kurz. (1P)

skip connection: Identity parallel to each residual block
add identity to output of block
learn the change of existing feature map

- e) Welche Nachteile hat die Verwendung der Sigmoid Aktivierungsfunktion für tiefe neuronale Netze wie Resnet ohne „skip connections“?

69, PS

(___ /1P)

Not centered around 0

Saturation behavior

Perceptive vanishes in saturation areas

Computationally intensive

- f) Berechnen Sie die Ausgaben der Faltungs-, sowie der Poolingsoperation, gegeben die nachfolgenden Eingangsdaten und die folgenden Operationen in sequentieller Reihenfolge.

Operation 1: Faltung mit 3×3 , Stride 1, Padding 0

Operation 2: Max Pooling mit 3×3 , Stride 1, Padding 0

66, 133

Eingangsdaten:

8	0	1	1	8
8	0	1	1	8
8	0	1	1	8
8	0	1	1	8
8	0	1	1	8

Faltungsfilter:

1	0	0
0	1	0
0	0	1

1:

9	2	10
9	2	10
9	2	10

2:

10

Aufgabe 4 – SVM / 9P

- a) Bezogen auf einen Support Vektor Klassifikator, beschreiben Sie das Problem welches gelöst wird, die Lösung die gefunden wird und Intuition für die Lösung.

(____/1,5P)

07. pg

Problem: Classification

Solution: Find the best separating line/ hyperplane with maximum margin to the classes

Intuition: Size of the margin determines the generalization capability.

- b) Formulieren Sie das grundlegende Optimierungsproblem für einen Support Vector Klassifikator formal. Geben Sie außerdem die Nebenbedingungen für die Optimierung an.

(____/1,5P)

07. P10 P11

Find the hyperplane $\{\bar{x}' \in X | \bar{w}\bar{x} + b = 0, (\bar{w}, b) \in \mathcal{X} \times \mathcal{Y}\}$, with maximum margin.

$$\text{minimize } \|\bar{w}\|^2, \quad y_i(\bar{w}\bar{x}_i + b) \geq 1, \quad i=1, \dots, n$$

- c) Nennen Sie zwei Methoden, mit denen Support Vector Klassifikatoren auch nicht linear separierbare Daten trennen können.

(____/1P)

07. P22 P23

Soft Margin SVM

Kernel-Trick

- d) Wie viele Stützvektoren werden mindestens benötigt, wenn der Merkmalsraum n Dimensionen hat?

(____/1P)

CPT

nt

- e) Gegeben sind Daten mit fünf unterschiedlichen Klassenzugehörigkeiten. Wie können diese mithilfe von SVMs klassifiziert werden? Wie viele SVMs werden dafür mindestens benötigt? (___/1P)

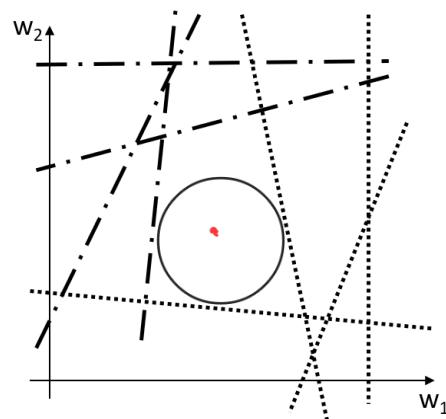
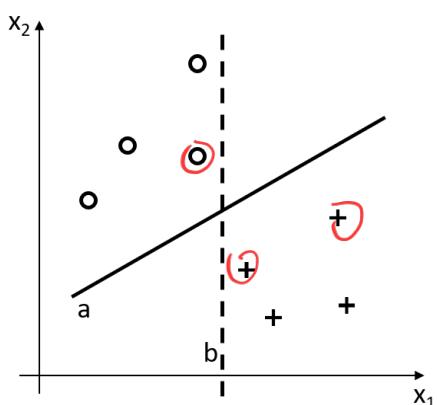
07.36

One-versus-all: 5

One-versus-one: $5 \times 4 / 2 = 10$

Multi-class classification: 5

- f) Welche der beiden Hypothesen im linken Graph ist die optimale Hypothese? Markieren Sie die entsprechenden Stützvektoren (Support Vectors) im linken Graphen (Merkmalsraum) und zeichnen Sie die Hypothese in den rechten Graphen (Hypothesenraum) ein. (___/3P)



c

Aufgabe 5 – Reinforcement Learning / 12P

- a) Welches Ziel verfolgt Reinforcement Learning? (____/1P)

Maximize rewards

- b) Durch welches Entscheidungsmodell lässt sich die Problemstellung beim Reinforcement Learning formal darstellen? (____/3P)

Welche vier Bestandteile werden für die Modellierung benötigt?

Markov Decision Process
State Space (S) Action Space (A)
Reward function (R) Transition function (P)

- c) Was besagt die Markov-Eigenschaft? (____/1P)

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

- d) Was ist ein wesentlicher Unterschied zwischen den Spielen „Schach“ und „Poker“ mit Bezug auf die Beschreibung der Zustände? Wodurch entsteht dieser Unterschied? (____/1P)

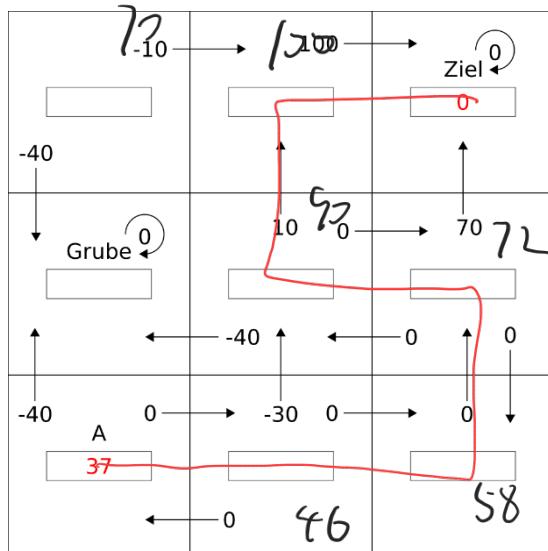
States and Observations

- e) Gegeben ist $V^\pi(s)$: Bestimmen Sie die Optimale Strategie formal. Außerdem, definieren Sie die rekursive Form der Bellmann Optimalitätsgleichung in Abhängigkeit von der optimalen Wertfunktion. (___/2P) 11, P38

$$\pi^* = \arg\max_a V^\pi(s)$$

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

- f) Betrachten Sie die untenstehende Welt. Ein Agent kann sich mit den angezeigten Aktionen (Pfeile) von Zelle zu Zelle bewegen. Die Belohnung für eine Aktion 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, HMM und SPN / 12P

Gegeben sei das folgende Hidden Markov Model (HMM) (grafische Darstellung siehe unten) mit Zustandsraum $S = \{S_1, S_2, S_3\}$, Beobachtungsraum $V = \{V_1, V_2, V_3\}$. Die Übergangswahrscheinlichkeitsmatrix A ist gegeben durch:

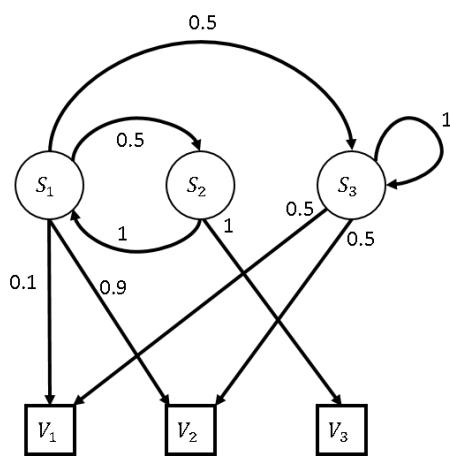
$$A = a_{(ij)} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Die Emissions-/Beobachtungs-Wahrscheinlichkeiten $B = (b_{ik})$ sind wie folgt definiert:

$$B = b_{(ik)} = \begin{bmatrix} 0.1 & 0.9 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

Hinweis: q_t beschreibt den Zustand zum Zeitpunkt t, o_t die Beobachtung im Zeitpunkt t.

- a) Zu Beginn befindet sich das System in Startzustand $q_1 = S_1$.
 Sie erhalten folgende Sequenz von Messungen: $o_2 = V_3, o_3 = V_1, o_4 = V_3$
 Geben Sie die a posteriori Verteilung $P(q_4 | o_2 = V_3, o_3 = V_1, o_4 = V_3, q_1 = S_1)$ an.
Hinweis: Decodierungsproblem



$$\begin{aligned} P(q_4 = S_1 | o_2 = V_3, o_3 = V_1, o_4 = V_3, q_1 = S_1) &= \\ P(q_4 = S_2 | o_2 = V_3, o_3 = V_1, o_4 = V_3, q_1 = S_1) &= \\ P(q_4 = S_3 | o_2 = V_3, o_3 = V_1, o_4 = V_3, q_1 = S_1) &= \end{aligned}$$

- b) Das in Aufgabe a) gegebene HMM befindet sich im Zeitpunkt t in Zustand $q_t = S_2$. Geben Sie folgende Wahrscheinlichkeiten für zukünftig mögliche Beobachtungen zum Zeitpunkt $t + 2$ bzw. $t + 3$ an:

(___ /3P)

$$\begin{aligned} P(o_{t+2} = V_3 | q_t = S_2) &= \\ P(o_{t+2} = V_1 | q_t = S_2) &= \\ P(o_{t+3} = V_3 | q_t = S_2) &= \end{aligned}$$

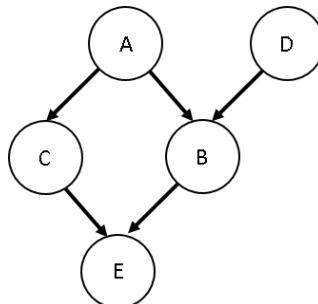
- c) Gegeben sei folgendes Bayes'sche Netz. Geben Sie die zugehörige Faktorisierung der Verbundwahrscheinlichkeitsdichte an:

(___ /2P)

14.11.19

$$P(A, B, C, D, E)$$

17.11.19



$$\begin{aligned} p(A, B, C, D, E) &= p(A) \cdot p(B) \cdot p(C|A) \cdot p(D|A, B) \\ &\quad \cdot p(E|B, C) \end{aligned}$$

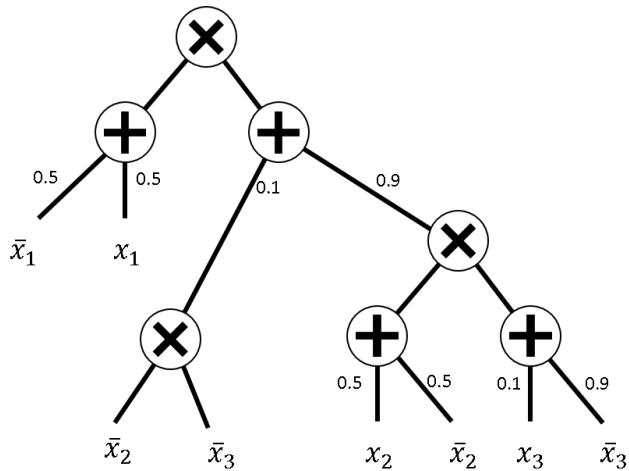
- d) Nennen Sie zwei Arten/Typen der Inferenz, die mit SPN ausgeführt werden können.

(___ /2P)

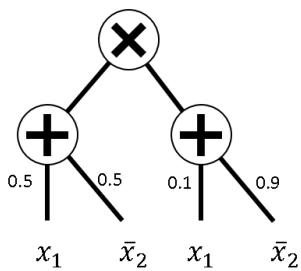
e) Geben Sie für die folgenden Sum Product Netze (SPN) jeweils an, ob diese valide sind oder nicht.

(__/3P)

A



B



C

