

## Prüfung "Künstliche Intelligenz in der Fahrzeugtechnik" 10.03.2020

**Bearbeitungszeit: 90 Minuten**

Die Prüfung besteht aus 12 Aufgaben auf 30 Blättern

**Hilfsmittel: keine Unterlagen,**  
nur einfacher, nichtprogrammierbarer Taschenrechner,  
nichtelektronisches Wörterbuch (wird während der Prüfung überprüft)

**Bitte verwenden Sie ausschließlich DOKUMENTENSICHERE Schreibgeräte und  
NICHT die Farbe ROT.  
Der Rechenweg ist anzugeben und wird mitbewertet.**

Name	Vorname	Matrikelnummer

### MUSTERLÖSUNG

Aufgabe	1	2	3	4	5	6	7	8	9	10	11	12
Punkte												
<b>Punkte (maximal)</b>	<b>15</b>	<b>15</b>	<b>14</b>	<b>16</b>	<b>14</b>	<b>15</b>	<b>15</b>	<b>13</b>	<b>14</b>	<b>14</b>	<b>14</b>	<b>16</b>
										<b>Summe</b>		
										<b>175</b>		

Hiermit bestätige ich, dass ich vor Prüfungsbeginn darüber in Kenntnis gesetzt wurde, dass ich im Falle einer plötzlich während der Prüfung auftretenden Erkrankung das Aufsichtspersonal umgehend informieren muss. Dies wird im Prüfungsprotokoll vermerkt. Danach muss unverzüglich ein Rücktritt von der Prüfung beim zuständigen Prüfungsausschuss beantragt werden. Ein vertrauensärztliches Attest – ausgestellt am Prüfungstag – kann gegebenenfalls innerhalb der nächsten Tage nachgereicht werden. Wird die Prüfung hingegen in Kenntnis der gesundheitlichen Beeinträchtigung dennoch regulär beendet, kann im Nachhinein kein Prüfungsrücktritt aufgrund von Krankheit beantragt werden.

Studiengang: \_\_\_\_\_

Garching, den .....

\_\_\_\_\_  
(Unterschrift)

## 1 Aufgabe: Einführung in die künstliche Intelligenz

a: 4

b: 4

c: 4

d: 3

/ 15 Punkte

- a) Nennen Sie eine mögliche Einteilung von künstlicher Intelligenz, die uns vier mögliche Ziele aufzeigt, die wir mit künstlicher Intelligenz verfolgen können. Geben Sie für jede diese Kategorie ein Beispiel, wie sich die Intelligenz in dieser Kategorie darstellen kann. (*English translation: Name a possible classification of artificial intelligence, which shows us four possible goals that we can pursue with artificial intelligence. For each of these categories, give an example of how the intelligence in that category may be presented.*)

Vorlesung 01: Introduction to AI – Folie 22 → Max 4 Punkte

1. Thinking Rationally: z. B. Correct Reasoning, Logic,...
2. Acting Rationally: z. B. Pursue Goals, Create, Adapt...
3. Thinking Humanly: z. B. Human Performance, ...
4. Acting Humanly: z. B. Perception, Knowledge Storage,...

Jeweils 0,5 Punkte für richtige Kategorie

Jeweils 0,5 Punkte für richtiges Beispiel

- b) Erklären Sie, was der Unterschied zwischen Artificial Intelligence, Machine Learning und Deep Learning ist. Erläutern Sie in Stichpunkten, was die Vorteile von Deep Learning ggü. Machine Learning ist (*English translation: Explain what the difference is between Artificial Intelligence, Machine Learning and Deep Learning. Explain in key points what are the advantages of Deep Learning compared to Machine Learning ?*)

Vorlesung 01: Introduction to AI – Folie 46 – Maximal 4 Punkte

Artificial Intelligence ist der große Überbegriff, unter welchen viele Sub-Probleme fallen (1P)

Machine Learning ist ein Teilgebiet von AI (0,5) und beschäftigt sich mit Mustererkennung/ dem lernen aus Daten (0,5 P)

Deep Learning ist ein Teilgebiet von Machine Learning und beschäftigt sich mit neuronalen netzen (0,5). Das „Deep“ bezeichnet die große Anzahl von layern und Neuronen für bessere Performance (0,5P)

Vorteil: (Maximal 1P)

- Mehr Daten verarbeiten (0,5 P), bessere performance bei gleicher Datenazahl (0,5P), keine manuelle feature Eingabe: Representation Learning (0,5P)

- c) In der Vorlesung haben wir das Sub-Problem der Argumentation und Problemlösung kennen gelernt. Erläutern Sie kurz, was unter diese Problemdefinition fällt und nennen Sie zwei Methoden und Algorithmen, mit welcher wir dieses AI-Problem lösen können? *(English translation: c) In the lecture we got to know the sub-problem of argumentation and problem solving. Do you briefly explain what falls under this problem definition and name two methods and algorithms with which we can solve this AI problem?)*

**Vorlesung 01: Introduction to AI – Folie 33 - Maximal 4 Punkte**

**Problembeschreibung:**

- A given problem or task should be solved (0,5P)
- A machine can use step-by-step argumentation/reasoning for solving this task (0,5P)
- A machine can use formal logic for solving this task (0,5P)
- Integration of uncertainty and probability necessary (0,5P)

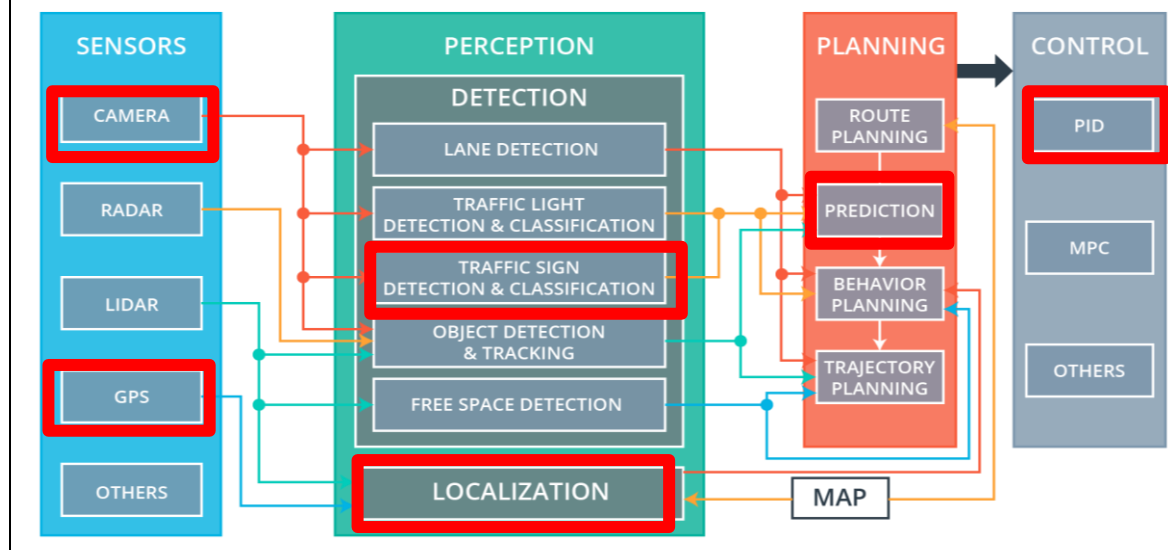
**Methoden & Algorithmen (Max. 2P):**

- Searching (0,5P): Intelligently searching through many possible solutions, e.g., tree search, Dijkstra, Kruskal, nearest neighbour, A\*-search (0,5P)
- Optimization (0,5P): Minimize/maximize a given problem with boundaries, e.g., lineare programming, quadratic programming, heuristics, ... (0,5P)
- Evolutionary computation (0,5 P): Optimization search based on evolutions, e.g., genetic algorithms, particle swarm optimization, ant colony optimization,... (0,5P)

Jeweils 0,5 Punkte für Problembeschreibung, Jeweils 0,5 Punkte für richtige Methode, Jeweils 0,5 Punkte für richtigen Algorithmus

- Jeweils 0,5 Punkte für richtiges Beispiel

- d) Im nachfolgenden Diagramm finden Sie die Software Pipeline, die für das autonome Fahren nötig ist. Bitte ergänzen sie die fehlenden Softwarebausteine mit dem entsprechenden Fachausdrücken. *(English translation: In the following diagram you will find the software Pipeline, which is necessary for autonomous driving. Please add the appropriate technical terms to the missing software modules?)*



## 2 Aufgabe: Computer Vision

a: 2      b: 6      c: 2      d: 5      /15 Punkte

- a) Nennen Sie vier Methoden, ein digital aufgenommenes Bild zu verarbeiten.  
(English translation: Name four methods of processing a digitally captured image.)

Vorlesung 02: Computer Vision – Seite 25 – Maximal 2 Punkte

- Change of Color Spaces (0.5 P)
- Filtering (0.5 P)
- Contrast Enhancement (0.5 P)
- Affine Transformation (0.5)
- Resampling/Compression (0.5 P)

Gegeben ist folgendes digitales Bild (English translation: The following digital image is given)

0	0	1	6
2	3	6	0
9	4	2	5
8	2	0	5

- b) Wenden Sie einen Mean-Filter auf das Bild an und fügen Sie die neuen Pixelwerte in die freien Zellen ein. (English translation: Apply a mean filter to the image and insert the new pixel values in free cells.)

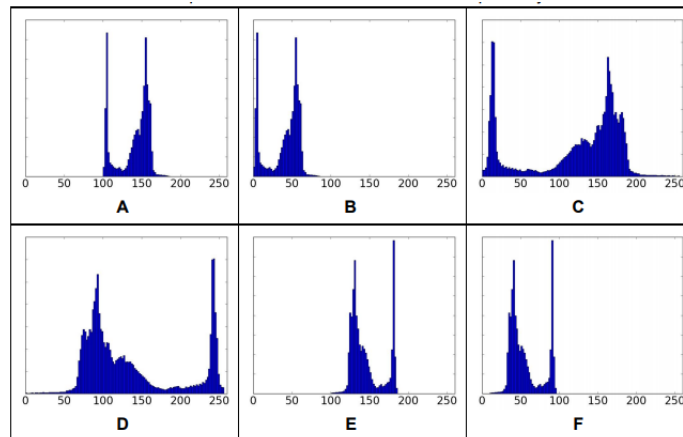
- Richtige Auswahl von Kernel: Mean-Kernel

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} (1 \text{ P})$$

- Äussere Bereiche besitzt keine Werte (1P)
- Pro richtigem Wert (1 P)

X	X	X	X
X	3	3	X
X	4	3	X
X	X	X	X

Sehen Sie sich die folgenden Histogramme für die Bilder A, B, C, D, E und F an, Null und 255 stehen für die Schwarz-Weiß-Intensität. (*English Translation: Refer to the following histograms for images A, B, C, D, E, and F; zero and 255 represent black and white intensity.*)



- c) Welches der Bilder B, D und E hat einen höchsten Kontrast und Warum? (*English Translation: Which of the images B, D and E has the highest contrast and Why?*)

D hat den höchsten Kontrast (1 P)

Da das Bild D ein breiteres Histogramm hat, das einen größeren Bereich von Intensitäten hat (1 P)

- d) Welche der oben genannten Histogramme entsprechen den folgenden Bild? Erklären Sie Ihre Antwort. (*English Translation: Which of the above histograms correspond to the following Picture? Explain your answer.*)



Maximal 5 P

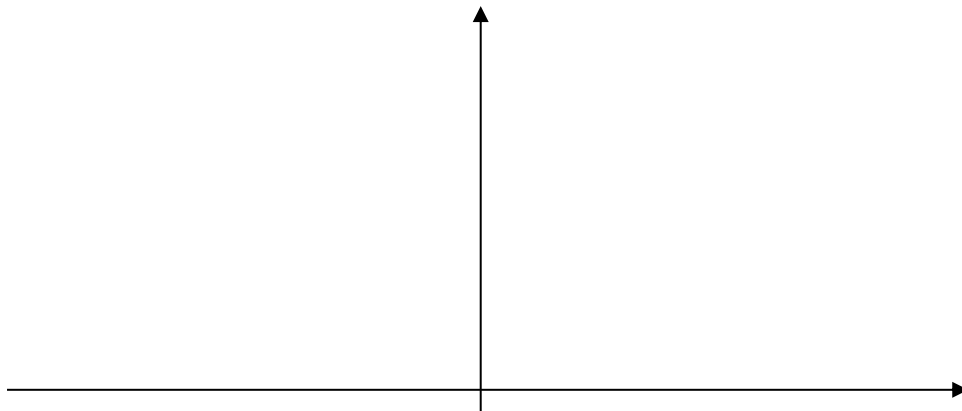
- Das Histogramm muss zwei Peaks haben (1 P).
- Ein relativ schmaler Peak bei dunklere Intensitäten (Mantel des Kameramanns und Beine des Stativs) und eine breitere Spitze, die meist aus helleren Intensitäten besteht (das Gras, der Himmel, die Gebäude im Hintergrund usw.) (2P)
- Dies liegt daran, dass es weniger dunkle Pixel im Vergleich zu den hellen gibt. (1 P)

Die Antwort lautet Bild **C** (1 P)

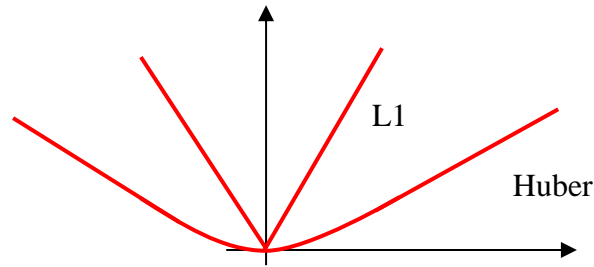
### 3 Aufgabe: Regression

a: 2	b: 2	c: 6	d: 4	/ 14 Punkte
------	------	------	------	-------------

- a) Skizzieren Sie qualitativ den Verlauf einer L1 und Huber Verlustfunktion und beschriften Sie diese entsprechend. (*English translation: Please sketch the qualitative behavior of a L1 and Huber loss function and label them accordingly.*)



Gesamtpunktzahl: 2 Punkte (einen Punkt für die L1 Funktion, einen Punkt für die Huber Funktion)



Nennen Sie einen Vor- und einen Nachteil dafür, dass eine Huber anstatt einer L1 Verlustfunktion für das Regressionsproblem verwendet werden soll. (English translation: Name one advantage and one disadvantage of using a Huber instead of an L1 loss function for a regression problem.)

Gesamtanzahl: 2 Punkte (einen Punkt für einen Vorteil, einen Punkt für einen Nachteil)

Mögliche Vorteile:

- Huber Verlust ist überall differenzierbar
- Huber Verlust verbindet die Vorteile von L1 und L2 Verlustfunktionen

Mögliche Nachteile:

- Mehr Hyperparameter

- c) Sie erhalten die Aufgabe, einen Algorithmus zur Identifikation der Kalibrierparameter  $c_1$  und  $c_0$  des Modells  $y = c_1x + c_0$  auf einem Mikrocontroller zu implementieren. Auf Grund der beschränkten Rechenkapazität kann die L2-optimale Lösung  $\vec{w} = (X^T X)^{-1} X^T \vec{y}$  nicht zur Laufzeit berechnet werden. Dabei bezeichnet  $X$  die Design-Matrix,  $\vec{w}$  den Gewichtsvektor und  $\vec{y}$  den Vektor mit den gemessenen Kalibrierwerten. Die Kalibrierwerte werden stets an folgenden Punkten gemessen:

$$x_1 = -2 \quad x_2 = 0 \quad x_3 = 2$$

Bestimmen Sie den L2-optimalen Prädiktor  $P$ , so dass  $\vec{w} = P\vec{y}$  gilt.

English translation: You have been given the task to design an algorithm for the identification of the calibration parameters  $c_1$  and  $c_0$  of the model  $y = c_1x + c_0$  for the implementation on a microcontroller. Due to limited calculation-power, you cannot calculate the L2-optimal solution  $\vec{w} = (X^T X)^{-1} X^T \vec{y}$  at runtime.  $X$  is the design matrix,  $\vec{w}$  the weight vector and  $\vec{y}$  the calibration value vector. The calibration values are always measured at the following points:

$$x_1 = -2 \quad x_2 = 0 \quad x_3 = 2$$

Calculate the L2-optimal predictor  $P$  such that  $\vec{w} = P\vec{y}$  holds.

Gesamtpunktzahl: 6 Punkte

Erkennen dass der Prädiktor wie folgt gegeben ist:  $P = (X^T X)^{-1} X^T$  (1 Punkt)

Berechnen der Design-Matrix  $X = \begin{pmatrix} -2 & 1 \\ 0 & 1 \\ 2 & 1 \end{pmatrix}$  (2 Punkte)

Berechnen der Matrix  $X^T X = \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix}$  (1 Punkt)

Berechnen der Matrix  $(X^T X)^{-1} = \begin{pmatrix} 0.125 & 0 \\ 0 & 0.333 \end{pmatrix}$  (1 Punkt)

Berechnen des Prädiktors  $P = \begin{pmatrix} -0.25 & 0 & 0.25 \\ 0.333 & 0.333 & 0.333 \end{pmatrix}$  (1 Punkt)

d) Für ein anderes Kalibrierproblem wurde der folgende Prädiktor bestimmt:

$$P = \begin{pmatrix} -0.5 & 0 & 0.5 \\ 0.333 & 0.333 & 0.333 \end{pmatrix}$$

Dabei werden die Kalibrierwerte an den Stellen  $x_1 = -1$ ,  $x_2 = 0$   $x_3 = 1$  aufgenommen und der Messvektor lautet wie folgt:  $y = (y_1 \ y_2 \ y_3)^T$ . Bestimmen Sie die Kalibrierparameter  $c_1$  und  $c_0$  für jede der folgenden Messreihen:

Messreihe A mit Datenpunkten der Form  $(x, y)$ : (1, 3); (0, 0); (-1, -3)

Messreihe B mit Datenpunkten der Form  $(x, y)$ : (1, 4); (0, 0); (-1, -1)

*English translation: For another calibration problem the following predictor was determined:*

$$P = \begin{pmatrix} -0.5 & 0 & 0.5 \\ 0.333 & 0.333 & 0.333 \end{pmatrix}$$

*The calibration values are located at  $x_1 = -1$ ,  $x_2 = 0$   $x_3 = 1$  and the calibration value vector is given as  $y = (y_1 \ y_2 \ y_3)^T$ . Determine the calibration parameters  $c_1$  and  $c_0$  for each of the following measurement datasets:*

*Dataset A with data points of the form  $(x, y)$ : (1, 3); (0, 0); (-1, -3)*

*Dataset B with data points of the form  $(x, y)$ : (1, 4); (0, 0); (-1, -1)*



Gesamtpunktzahl: 4 Punkte (jeder richtige Parameter ein Punkt)

Messreihe A:  $\begin{pmatrix} c_1 \\ c_0 \end{pmatrix} = P \begin{pmatrix} -3 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3.0 \\ 0 \end{pmatrix}$

Messreihe B:  $\begin{pmatrix} c_1 \\ c_0 \end{pmatrix} = P \begin{pmatrix} -1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 2.5 \\ 1 \end{pmatrix}$

#### 4 Aufgabe: Classification

a: 2	b: 1	c: 2	d: 3	e: 4	f: 4	/ 16 Punkte
------	------	------	------	------	------	-------------

- a) Ordnen Sie die folgenden Begriffe, soweit möglich, den Methoden „Classification“ und „Clustering“ richtig zu (*English translation: Correctly assign the following terms to the "Classification" and "clustering" methods as far as possible*): „discrete output“, „continuous output“, „supervised“, „unsupervised“

Classification: discrete output, supervised  
Clustering: discrete output, unsupervised  
Je 0,5 → 2 P

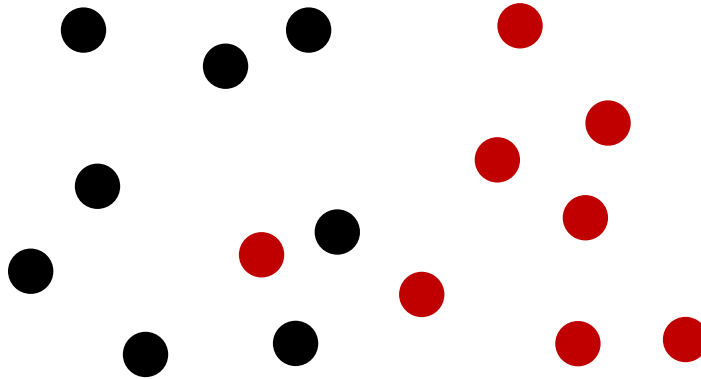
- b) Sie wollen k-NN mit k=7 anwenden. Welche Besonderheit ergibt sich für das „Training“ dadurch das k-NN eine Methode mit „lazy evaluation“ ist? (*English translation: You want to apply k-NN with k=7. What is special about the "training" because k-NN is a method with "lazy evaluation"?*)

Es gibt kein Training, da bei k-NN kein Model aufgebaut wird, sondern sämtliche verfügbaren Daten bei jedem Klassifizierungsschritt durchgelaufen werden.  
1P für „es gibt kein Training“

- c) Wie können bei k-NN „Overfitting“ oder zu starke „Generalisation“ auftreten? (*English translation: How can "overfitting" or excessive "generalisation" occur with k-NN?*)

Overfitting wenn k zu klein gewählt wird  
Zu starke Generalisierung, wenn k zu groß gewählt wird.  
Je 1 P → 2 P

- d) Welche zwei Methoden zur Gewichtung von k-NN (weighted k-NN) kennen Sie. Welche der beiden Methoden macht im folgenden Beispiel keinen Unterschied zu k-NN (ohne Gewichte) und warum? (*English translation: Which two methods are known for weighting k-NN (weighted k-NN). Which of the two methods in the following example makes no difference to k-NN (without weights) and why?*)



- Methode 1: Gewichtet nach Häufigkeit der Klassenelemente.
  - Macht in diesem Beispiel keinen Unterschied zu k-NN da beide Klassen gleich viele Elemente haben.
- Methode 2: Gewichtung nach Distanz

Je 1 P → 3 P

- e) Sie wollen eine SVM für die weitere Klassifizierung nutzen. Nennen und beschreiben Sie kurz zwei Möglichkeiten wie Sie mit den nicht linear trennbaren Daten von Aufgabe d) umgehen können. (*English translation: You want to use an SVM for further classification. Briefly name and describe two ways in which you can handle the non-linearly separable data from task d)*)

- Soft Margin 1P
  - Kleine Klassifizierungsfehler in den Trainingsdaten bei der Konstruktion der MMH in Kauf nehmen → Den roten Ausreißer ignorieren 1P
- Kernel Trick 1P
  - Transformation in einen höheren Raum (am besten mit einer Kernel Funktion), in welchem die Daten linear trennbar sind 1P
- Max 4P

- f) Berechnen Sie aus der folgenden Konfusions-Matrix die Precision (Genauigkeit) und die Specificity (Spezifität) für „Klasse 3“. (*English translation: Calculate the precision and specificity for “Klasse 3” from the following Confusions Matrix.*)

		Classified as		
Correct Label		Klasse 1	Klasse 2	Klasse 3
	Klasse 1	23	8	4
	Klasse 2	2	36	0
	Klasse 3	3	2	78

Precision =

Specificity =

Precision:  $TP/(TP+FP) = 78/(78+4+0) = 95,12 \%$

Specificity =  $TN/(TN+FP) = (23+8+2+36)/(23+8+2+36 + 4+0) = 94,5 \%$

Je 1 P für Formel, je 1 P für richtiges Ergebnis, je 0.5 P Abzug für Rechenfehler  
→ 4P

## 5 Aufgabe: Clustering

a: 3

b: 5

c: 4

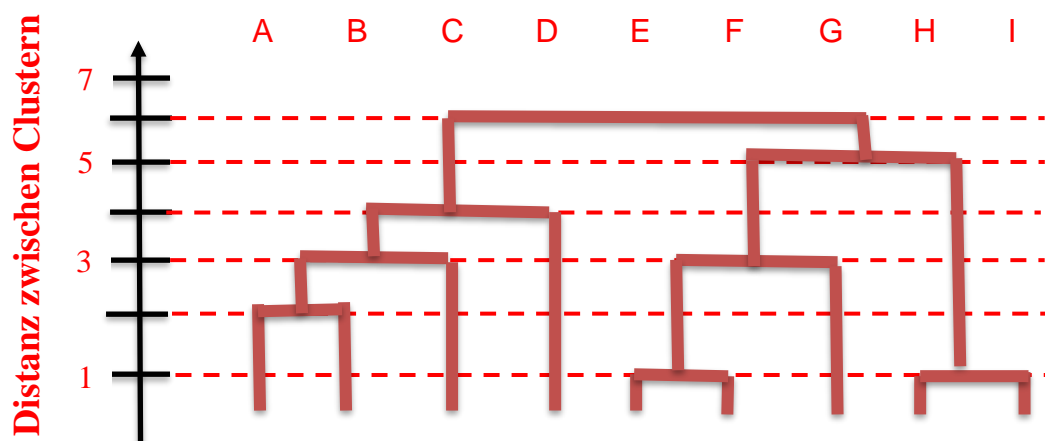
d: 2

/ 14 Punkte

- a) Die 9 Elemente A bis I liegen auf einem Raster. Zeichnen Sie das Dendrogramm für ein Hierarchisches Clustering unter Verwendung von Single Link und der Manhattan Distanz. Beschriften Sie die Hochachse vollständig. (*English translation: The 9 elements A to I lie on a grid. Draw the dendrogram for hierarchical clustering using Single Link and the Manhattan Distance. Label the vertical axis completely.*)

				D					
	B							H	I
A									
		C							
						G			
							E	F	

3 P, pro Fehler (auch Länge der Kanten oder fehlenden Beschriftung) 0,5 Abzug



- b) Nutzen Sie das Dendrogramm um die beste Aufteilung der 9 Elemente in 2 Cluster zu bestimmen. Berechnen Sie dann den Silhouetten Koeffizienten für das Element D und interpretieren Sie das Ergebnis. (*English translation: Use the dendrogram to determine the best division of the 9 elements into 2 clusters. Then calculate and interpret the silhouette coefficient for element D.*)

$$sim(o) = \frac{1}{|a|} \sum_{e \in a} distance(o, e), \quad sim(D) = \frac{4 + 6 + 5 + 0}{4} = \frac{15}{4} = 3,75$$

1,5 P, auch 15/4 ist als richtig zu werten

$$dsim(o) = \min_{c \neq a} \left( \frac{1}{|c|} \sum_{e \in c} distance(o, e) \right), \quad dsim(D) = \frac{6+7+7+10+11}{5} = \frac{41}{5} = 8,2$$

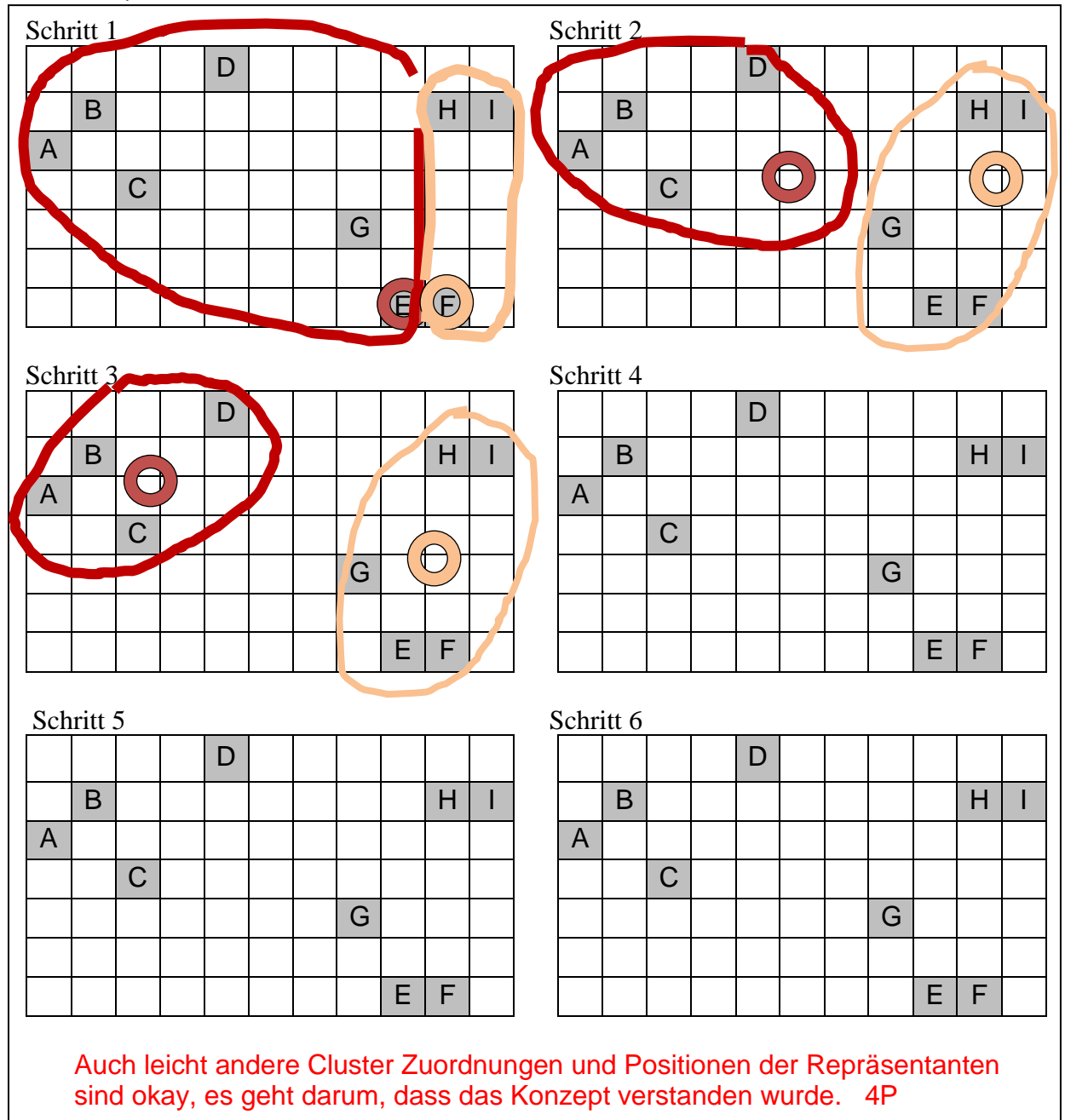
1,5 P

$$s(o) = \frac{dsim(o) - sim(o)}{\max\{sim(o), dsim(o)\}}, \quad s(D) = \frac{8,2 - 3,75}{8,2} = 0,54 \quad (1P)$$

s(D) zeigt, dass das Element einem passenden Cluster zugeordnet wurde.  
(1P)

Jeder Rechenfehler ist 0,5 P Abzug → Max 5 P

- c) Nutzen Sie nun k-means mit  $k=2$  um die gleichen 9 Elemente in dem Raster zu clustern. Als Initiale Repräsentanten wählen Sie Element E und F, in den drauffolgenden Iterationsschritten können künstliche Repräsentanten erstellt werden. Schätzen Sie die Euklidischen Distanzen zwischen den Elementen visuell ab und zeichnen Sie zu jedem Iterationsschritt die Repräsentanten und die Clusterzugehörigkeit der 9 Elemente ein. (English translation: Now use k-means with  $k=2$  to cluster the same 9 elements in the grid. As initial representatives choose element E and F, in the following iteration steps artificial representatives can be created. Estimate the Euclidean distances between the elements visually and draw for each iteration step the representatives and the cluster association of the 9 elements.)



- d) Welche Probleme können durch die Initiale Wahl der Repräsentanten auftreten? Nennen Sie eine Methode um dem entgegenzuwirken. (*English translation: What problems can arise from the initial choice of representatives? Name at least one method to deal with this.*)

Durch eine ungünstige Wahl kann k-means in ein lokales Optimum laufen und das globale Optimum verfehlen.

Möglichkeiten um das zu verhindern:

- Wiederholt zufällige Repräsentanten auswählen und bestes Resultat behalten
- Durch hierarchisches Clustern einer Teilmenge bereits eine durchdachte Auswahl der initialen Repräsentanten treffen
- Künstliche Repräsentanten gleichverteilt über den gesamten Raum legen

Je 1 P für Problem und eine Möglichkeit → 2 P



## 6. Aufgabe: Pathfinding

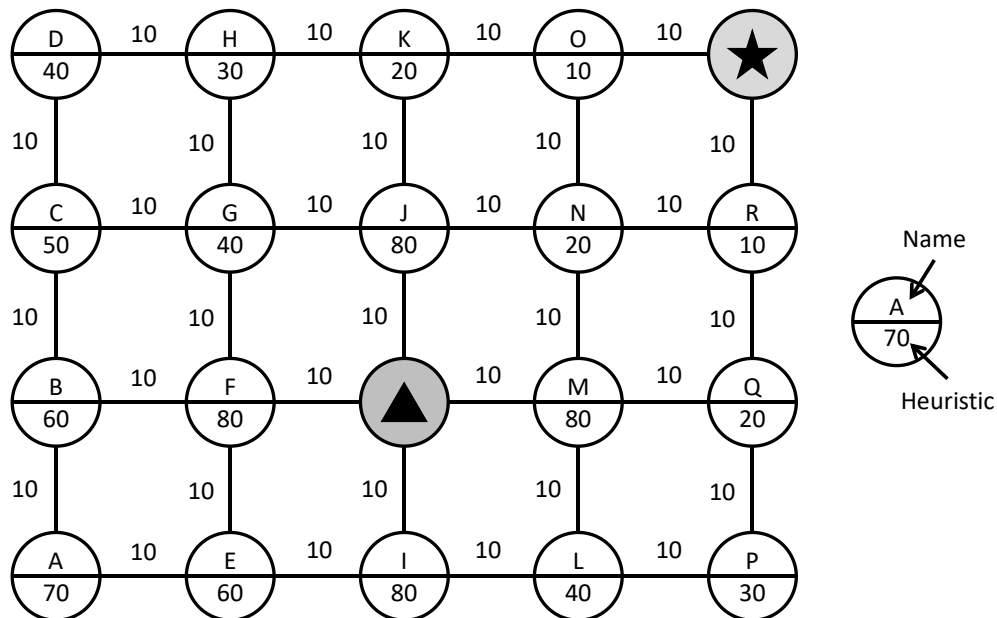
a: 4

b: 5

c: 6

/ 15 Punkte

Für die Aufgabenteile a) und b) liegt Ihnen folgender, routingfähiger Graph vor:  
(For tasks a) and b), consider the following routable Graph:)



Die Kantenlänge zwischen zwei benachbarten Knoten beträgt stets 10 Einheiten. Nehmen Sie an, dass „Dreieck“ der Startpunkt und „Stern“ der Zielpunkt ist!  
(The edge length between two adjacent nodes is always 10 units. Let “triangle” be the origin and “star” the destination!)

- a) Wenden Sie den A\*-Algorithmus an! Verwenden Sie an Stelle der Luftliniendistanz den im jeweiligen Knoten gegebenen Wert als Heuristik. Nennen sie den sich ergebenden Pfad! Ist dieser Pfad optimal bezüglich seiner Länge?  
(Apply the A\*-algorithm! Instead of the airline distance, use the value given in the respective node as a heuristic. Name the resulting path! Is this path optimal regarding its length?)

Triangle, F, G, H, K, O, Star (3 P)

No (1 P)

- b) Welche Werte dürfen die Heuristiken der Knoten J und M maximal einnehmen, wenn die Heuristik für den Knoten F auf 10 Einheiten und die Heuristik für den Knoten I auf 20 Einheiten gesetzt wird und ein optimaler Pfad mit dem A\*-Algorithmus gefunden werden soll?

*(What are the maximum values for the heuristics of nodes J and M if the heuristic for node F is set to 10 units and the heuristic for node I to 20 units and an optimal path is to be found using the A\*-algorithm?)*

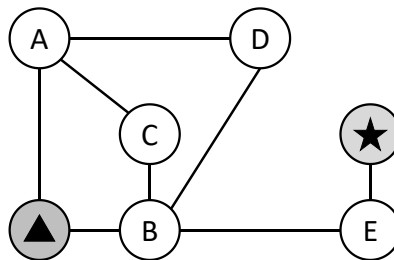
Maximum Heuristic J: 30 (2.5 P)

Maximum Heuristic M: 30 (2.5 P) -> Accordance with upper estimate

Also 2.5P each for 50, because the result would be correct as well, also the resulting estimate is not a valid upper estimate.

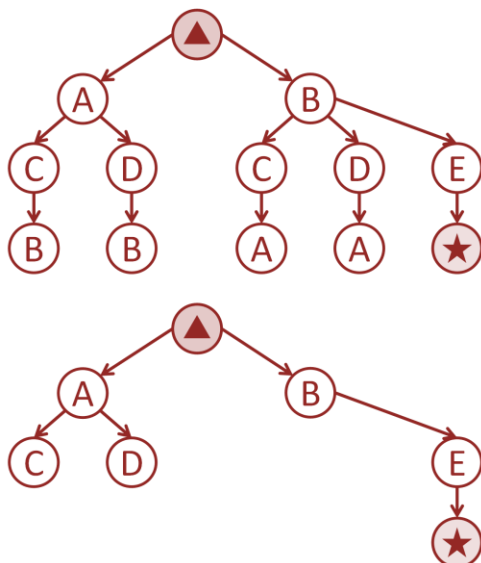
- c) Gegeben ist der untenstehende Graph. Nehmen Sie an, dass „Dreieck“ der Startpunkt und „Stern“ der Zielpunkt ist. Zeichnen Sie je einen Suchbaum für die Breitensuche mit Wiederbesuch („revisitation“) und die Breitensuche ohne Wiederbesuch.

*(Consider the graph below. Assume that "triangle" is the origin and "star" is the destination. Draw one search tree for the Breadth-First-Search with revisitation and the Breadth-First-Search without revisitation)*



Search Tree with revisitation: 3 P

Search Tree without revisitation: 3 P



## 7. Aufgabe: Introduction to Artificial Neural Networks

a: 2

b: 2

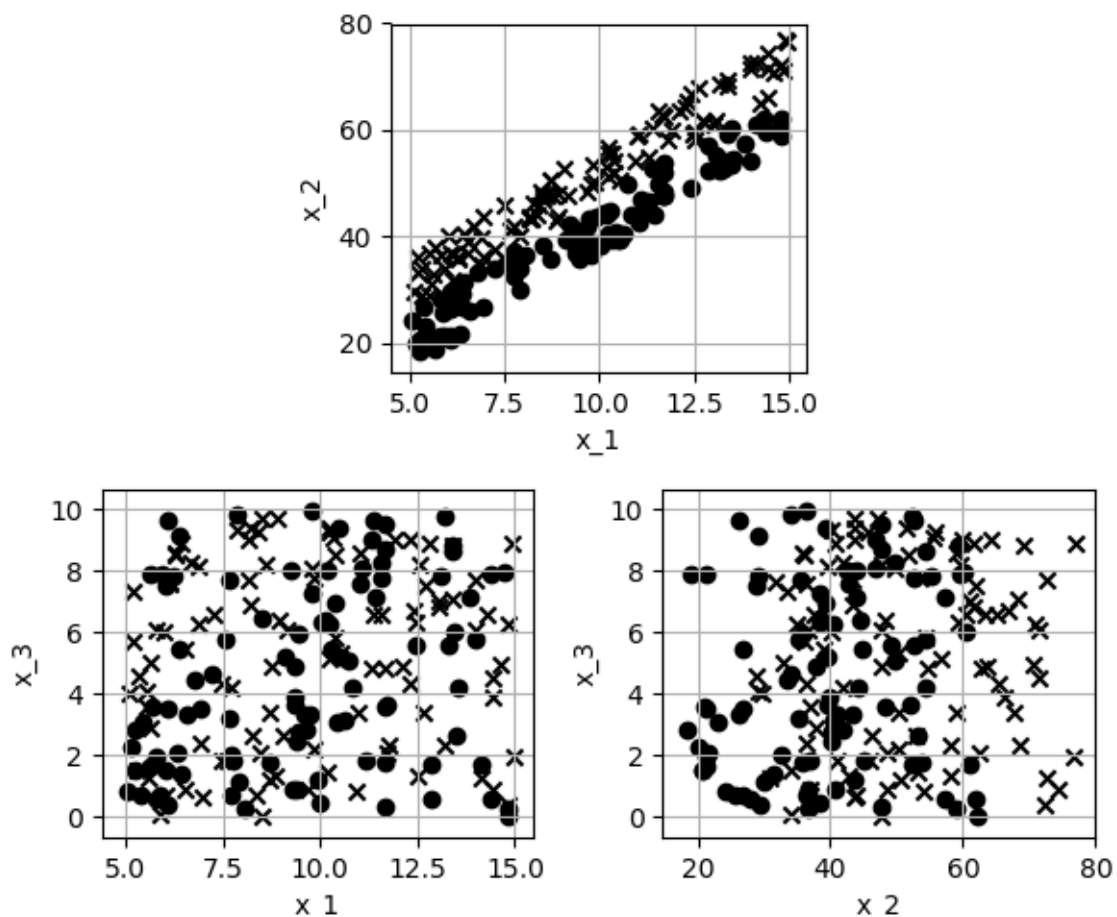
c: 8

d: 3

/ 15 Punkte

Folgender, dreidimensionaler Datensatz mit zwei unterschiedlichen, gekennzeichneten Klassen (Kreise und Kreuze) liegt Ihnen vor:

*(Consider the following three dimensional data set with two differently labeled classes (circles and crosses):)*



- a) Sie wollen für unbekannte Punkte bestimmen, ob diese Kreuze oder Kreise sind. Wie nennt sich diese Problemkategorie genau?  
*(You want to determine for unknown points whether they are crosses or circles. What is this problem category called exactly?)*

Binary Classification (2 P)

Missing "Binary": - 1P

- b) Sie möchten ein einzelnes künstliches Neuron trainieren, um zwischen Kreuzen und Kreisen zu unterscheiden. Welche Aktivierungsfunktion wählen Sie?  
(*You want to train a single artificial neuron to distinguish between crosses and circles. Which activation function do you choose?*)

Sigmoid OR Tangens Hyperbolicus (2 P)

- c) Skizzieren Sie den Berechnungsgraphen eines künstlichen Neurons mit maximal zwei Eingängen, das dazu trainiert werden kann, zwischen Kreuzen und Kreisen zu unterscheiden! Begründen Sie die Wahl Ihrer Eingänge.  
(*Sketch the computational graph of an artificial neuron with a maximum of two inputs, which can be trained to distinguish between crosses and circles! Justify your choice of inputs.*)

Inputs:  $x_1$  und  $x_2$  (1.5 P each)

Correct Neuron sketch including inputs, weights, bias, summation and activation function (2 P)

Reason for  $x_1/x_2$ :

Circles and crosses are separable (2P) by a linear function in the  $x_1, x_2$  plane only (1P for linear/plane)

Total: 8 P

- d) Wäre ein künstliches Neuron mit drei Eingängen besser für die Lösung des Problems geeignet? Begründen Sie Ihre Antwort.  
(*Would an artificial neuron with three inputs be better suited to solve the problem? Justify your answer.*)

No (1 P)

There seems to be no correlation between the  $x_3$  value of the data points and the assigned classes. (1P for no advantage of  $x_3$ )

Using another input to the artificial neuron would potentially hinder the training process, because the corresponding weight would need to be learned to 0, before convergence is observed. (1P for disadvantage of  $x_3$ )

## 8 Aufgabe: Deep Neural Networks

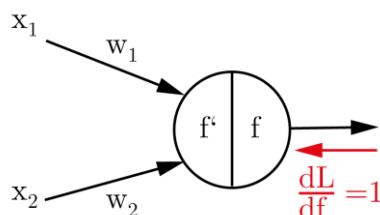
a: 8

b: 2

c: 3

/ 13 Punkte

- a) Sie haben ein neues künstliches Neuron entwickelt und möchten die Änderung der Gewichte berechnen. Vervollständigen hierzu den angegebenen Computational Graph und berechnen Sie mithilfe der Backpropagation die Änderung des Gewichtes  $w_1$  nach einer Trainingsiteration. Der eingehende downstream Gradient ist gegeben. (English translation: You developed a new artificial neuron and want to calculate the weight update. Complete the given computational graph and calculate the weight update after one training iteration with help of the backpropagation algorithm. The incoming downstream gradient is given)



$$f = \frac{1}{e^{x_1 w_1} + e^{x_2 w_2}}$$

$$x_1 = 1$$

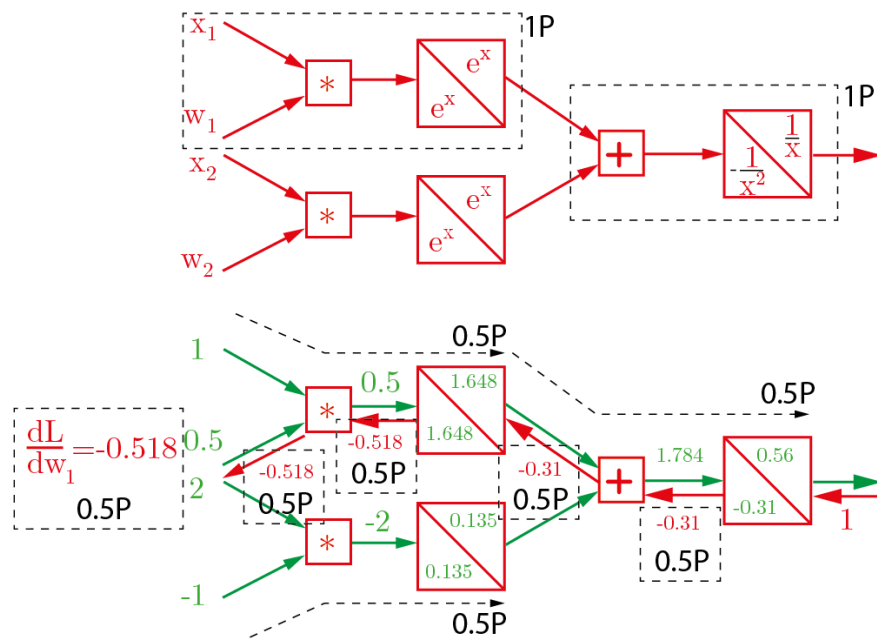
$$x_2 = 2$$

$$w_1 = 0.5$$

$$w_2 = -1$$

$$\alpha = 0.01$$

Mulö, Vorlesung 8 Folie 44:



$$\Delta w_1 = -\alpha \frac{dL}{dw_1} = -0.01 * -0.518 = 0.00518 \quad (1P \text{ Formel, } 1P \text{ Ergebnis})$$

Computational Graph nicht zwingend erforderlich. Dann gleiche Bewertung:

- Bis zu 2P für Ableitung
- Bis zu 4P für Ergebnis von  $dL/dw_1$ . (Wenn Teilergebnisse vorhanden, dann Teilpunkte möglich. 4P für korrektes Ergebnis.)
- 2P für Gewichtänderung oder neues Gewicht: 1 Formel + 1 Ergebnis

- b) Sie wollen ein Netzwerk zur Klassifizierung von Objekten entwickeln. Welche Aktivierungsfunktion nehmen Sie für das letzte Layer? Warum haben Sie diese Aktivierungsfunktion gewählt? *(English translation: You are developing a new artificial neural network for image classification. Which activation function are you choosing for the last layer and why?)*

Mulö, Vorlesung 08 Folie 97

Antwort: Sigmoid / Softmax (1P)

Begründung: Ausgabe zwischen 0-1 -> Interpretierbar als Vorhersagewahrscheinlichkeit (1P)

- c) In der Regel wird das Batch Gradient Descent Verfahren zum Training künstlicher Neuroner Netze angewandt. Warum wird dieses Verfahren verwendet? Nennen Sie einen Vor- und Nachteil des Verfahrens z.B. gegenüber dem Stochastic Gradient Descent. *(English translation: Usually the batch gradient descent algorithm is used for training artificial neural networks. Why is this algorithm used? State one advantage and one disadvantage compared to e.g. stochastic gradient descent!)*

Mulö, Vorlesung 08 Folie 114-117

Antwort: In der Regel ist es möglich auf dem gesamten Datenset auf einmal zu trainieren. (1P)

Vorteil: Konvergiert schneller als das Stochastic Gradient Decent (1P)

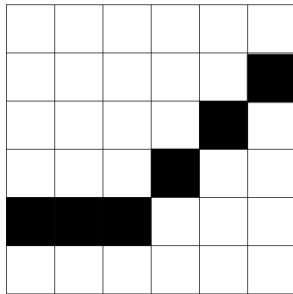
Nachteil: Höhere Rechenleistung notwendig (1P)

## 9 Aufgabe: Convolutional Neural Networks

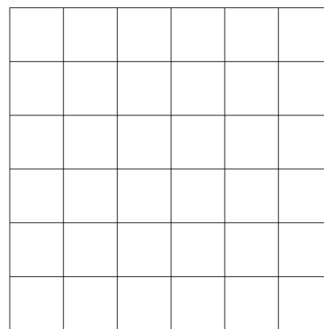
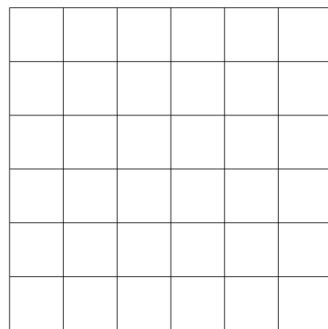
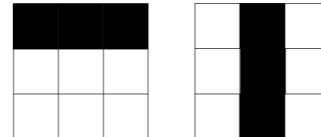
a: 4	b: 5	c: 1	d: 2	e: 2	/ 14 Punkte
------	------	------	------	------	-------------

- a) Gegeben ist ein Input Layer und zwei Filter eines Convolutional Layers. Zeichnen Sie die qualitativ die dazugehörigen Activation Maps! (*English translation: Given is one input layer and two filters of a convolutional layer. Draw qualitatively the according activation maps!*)

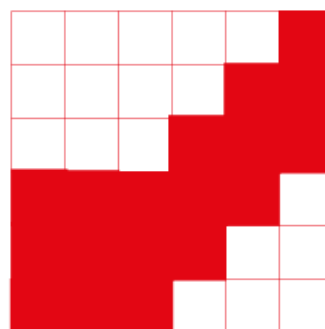
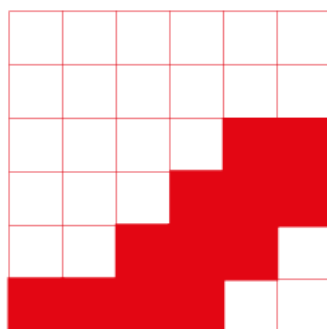
Input Layer:



Filter:



Mulö, Vorlesung 9:



Jeweils 2P

- b) Gegeben ist der Input und Output eines Convolutional Layers mit Relu Aktivierungsfunktion. Berechnen Sie den fehlenden Wert des dazugehörigen Filters (kein Bias)! Markieren Sie zudem den falschen Wert im Output Layer und geben Sie eine Begründung an warum dieser falsch ist. (English translation: Given is the input and output of a convolutional layer with Relu activation function. Calculate the missing value of the corresponding filter (no bias). In addition mark the wrong value in the output layer and state why this value is wrong!)

Input Layer						Filter			Output Layer					
-2	8	2	5	1	-6	2	-1	-1	0	43	103	21	0	0
5	1	-4	2	-4	8	7	0	5	0	0	54	55	144	40
-3	7	6	-6	-3	-8		-9	-9	56	61	53	60	0	14
3	-6	3	-9	6	-3				20	82	45	152	0	24
-4	-2	-7	-7	-5	0				11	0	0	0	0	-12
-2	0	3	8	2	0				6	2	50	29	47	4

Mulö:

Irgendeinen Wert aus dem Output Layer nehmen: 82, Berechnungsformel aufstellen und auflösen:

$82 = -3 \cdot 2 + 7 \cdot -1 + -1 \cdot 6 + 7 \cdot 3 + 0 \cdot -6 + 3 \cdot 5 + x \cdot -4 + -9 \cdot -2 + -7 \cdot -9$  // Test Berechnung von Convolution verstanden, ReLu verstanden.

$82 = 96 - 4 \cdot X$

$X = 4$  (3P) //

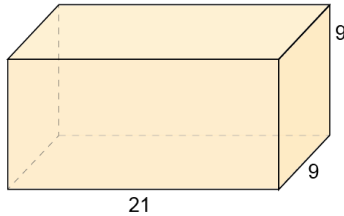
(1 P) Markierung

Output Layer					
0	43	103	21	0	0
0	0	54	55	144	40
56	61	53	60	0	14
20	82	45	152	0	24
11	0	0	0	0	-12
6	2	50	29	47	4

Begründung: Relu kann nicht negativ sein (1P)



- c) Auf den unten gezeigten Ausgang eines Convolutional Layers wird ein 3x3 MaxPool Layer mit Schrittweite 3 angewandt. Welche Dimension hat der Ausgang des MaxPool Layer? (*English translation: A 3x3 MaxPool layer with stepsize 3 is applied to the output of a convolutional layer shown below. What are the dimensions of the MaxPool layer's output?*)



Mulö:

$21 \times 9 \times 9 \rightarrow 21 \times 3 \times 3$

Or  $(9-3)/3 + 1 = 3$  (1P)

- d) Berechnen Sie die Ausgangsmatrix eines 2x2 AveragePool Layers für die gegebene Eingangsmatrix!. (*English translation: Calculate the output of a 2x2 Average Pool layer for the given input!*)

Eingangsmatrix:

0.9	1.0	-17.0	-21.0
1.0	1.1	-11.0	-15.0
-5.0	0.9	10.0	0.0
0.5	-0.3	-3.0	1.0

Ausgangsmatrix:


Mulö:

1	-16
-0.975	2

(0.5P pro Wert)

- e) Bevor Sie mit dem Training von neuronalen Netzen beginnen können, sollten Sie Ihre Trainingsdaten vorbereiten. Welcher Schritt ist hierzu in der Regel vorher notwendig und warum sollten Sie diese durchführen? *(English translation: Before training artificial neural networks, you should usually prepare the training data. Which step is here usually required and why should you prepare the training data?)*

Mulö:

1. Normalize trainings data. (1P)
2. Improves numerical stability of the model. (1P)

## 10. Aufgabe: Recurrent Neural Networks

a: 3

b: 2

c: 9

/ 14 Punkte

- a) Nennen Sie drei sinnvolle Ansätze um den versteckten Zustand  $h_0$  zu initialisieren! (*English: Enumerate three reasonable approaches to initialize the hidden state  $h_0$  !*)

3 of the following 4:

1. Initialize as zero, 2. Noisy zero mean, 3. Treat as parameters to learn, 4. Use a second NN

3p

- b) Bei der Initialisierung der Gewichte von rekurrenten neuronalen Netzen ist die Xavier Initialisierung weit verbreitet. Welche möglichen Probleme kann man durch die Benutzung dieser Methode reduzieren? Was ist der Mittelwert von Gewichten, welche mittels dieser Methode initialisiert wurden?

(*English: When initializing the weights of recurrent neural networks, a common approach is to use the Xavier initialization. Which possible problems can be reduced using this approach? What is the mean of the weights initialized using Xavier initialization?*)

1) Exploding and vanishing gradients can be avoided or at least reduced.

2) The mean of the weights is zero.

1p

1p

c) Die Berechnungsformel für ein rekurrentes neuronales Netz lauten:

$$h_t = \tanh(a \cdot h_{t-1} + x_t)$$

$$k_t = \tanh(b \cdot k_{t+1} + x_t)$$

$$y_t = c \cdot (h_t + k_t)$$

c.1) Um welche Art von rekurrentem neuronalem Netz handelt es sich?

c.2) Berechnen Sie  $L = y_1^2 + y_2^2$  für die gegebenen Eingangswerte und Startwerte!

$$h_0 = 0; \quad k_3 = 0; \quad a = 2; \quad b = 2; \quad c = 1$$

$$x_1 = 1; \quad x_2 = -1$$

c.3) Würde sich  $L$  vergrößern oder verkleinern für ein neues  $c_{neu} = c - \alpha \frac{\partial L}{\partial c}$  mit sehr kleinem  $\alpha$ ?

Runden Sie auf 2 Nachkommastellen!

(English: The formula of a recurrent neural network is given above.

c.1) What type of recurrent neural network is it?

c.2) Compute  $L = y_1^2 + y_2^2$  for the input and starting values given above!

c.3) Would  $L$  increase or decrease if we would use a new  $c_{new} = c - \alpha \frac{\partial L}{\partial c}$  with a sufficiently small  $\alpha$ ?

Round all values to 2 decimal places!

c.1) Es handelt sich um ein bidirektionales RNN

1p

c.2)

$$h_1 = \tanh(2 \cdot 0 + 1) = 0.76$$

$$h_2 = \tanh(2 \cdot 0.76 - 1) = 0.48$$

3p

$$k_2 = \tanh(2 \cdot 0 - 1) = -0.76$$

$$k_1 = \tanh(-2 \cdot 0.76 + 1) = -0.48$$

2p

$$y_1 = 0.76 - 0.48 = 0.28$$

$$y_2 = 0.48 - 0.76 = -0.28$$

$$L = 0.16$$

2p

c.3)  $L$  würde sich **verkleinern**, da es sich um einen Gradientenabstieg handelt. Für  $\alpha = \frac{c}{\frac{\partial L}{\partial c}}$  ergibt sich  $c_{neu} = 0$  und somit das Minimum  $L=0$ .

1p

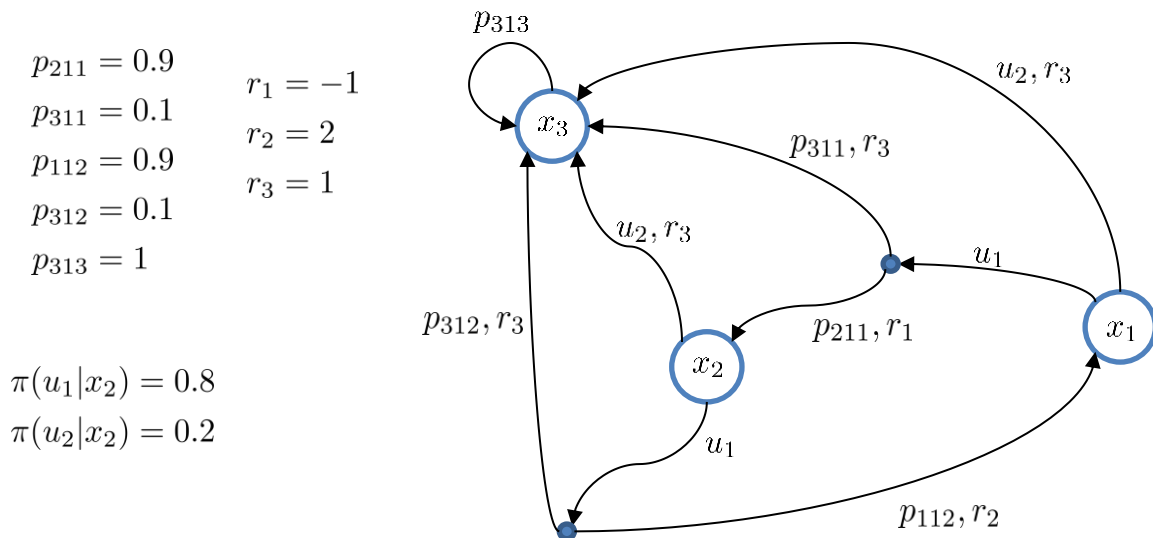
## 11. Aufgabe: Reinforcement Learning

a: 10

b: 2

c: 2

/ 14 Punkte



- a) Der gegebene Markov-Prozess besteht aus 3 Zuständen, mit dem Endzustand  $x_3$  und somit  $V(x_3) = 0$ . Es gilt  $\gamma = 0.9$ . Bestimmen Sie im Zustand  $x_1$  eine neue  $\epsilon$ -greedy Policy mit  $\epsilon = 0.1$ . Gehen Sie dazu wie folgt vor:
- Es ist gegeben  $V(x_1) = 1,08$ . Bestimmen Sie  $V(x_2)$ !
  - Berechnen Sie  $Q(x, u)$  für alle Aktionen in  $x_1$ .
  - Geben Sie die neue  $\epsilon$ -greedy Strategie in  $x_1$  an, mit  $\epsilon = 0.1$ .

(English:

The Markov-Process given above consists of three states, with an absorbing state  $x_3$  and thus  $V(x_3) = 0$ . We have  $\gamma = 0.9$ .

Find an improved  $\epsilon$ -greedy policy in the state  $x_1$  with  $\epsilon = 0.1$ , by following the steps:

- Determine  $V(x_2)$ , given that  $V(x_1) = 1,08$ !
  - Determine  $Q(x, u)$  for all actions in  $x_1$ .
  - Write down the new  $\epsilon$ -greedy strategy in  $x_1$  with  $\epsilon = 0.1$ .
- )

a.1)

$$\begin{aligned} V(x_2) &= \pi(u_1|x_2)(p_{112}(r_2 + \gamma V(x_1)) + p_{312}(r_3 + \gamma V(x_3)) + \pi(u_2|x_2)r_3) \\ &= 0.8(0.9 \cdot (2 + 0.9V(x_1)) + 0.1 \cdot (1 + 0)) + 0.2 \cdot 1 \\ &= 1.44 + 0.65V(x_1) + 0.08 + 0.2 \\ &= 0.65V(x_1) + 1.72 \\ &= 2.42 \end{aligned}$$

4p

a.2)

$$\begin{aligned} Q(x_1, u_1) &= p_{311}(r_3 + \gamma V(x_3)) + p_{211}(r_1 + \gamma V(x_2)) \\ &= 0.1 \cdot (1 + 0.9 \cdot 0) + 0.9 \cdot (-1 + 0.9 \cdot 2.42) \\ &= 1.16 \end{aligned}$$

$$Q(x_1, u_2) = r_3 + \gamma V(x_3) = 1.0$$

4p

a.3)

$$\pi_\epsilon(u_1|x_1) = 1 - \epsilon = 0.9$$

$$\pi_\epsilon(u_2|x_1) = 0.1$$

2p

- b) Wieso kann der einfache Q-learning Algorithmus, so wie er in der Vorlesung gezeigt wurde, nicht für das Training einer KI für das Spiel "Go" benutzt werden?

(English: What is a reason that simple Q-learning, as presented in the lecture, cannot be applied to train an AI for the boardgame "Go"?)

The game has **too many actions and states** for it to be solved using this simple approach. The computer would **not have enough memory** and training would take long because of no generalization over states. (2P)

- c) Wieso wird eine Aktions-Wertefunktion (action value function) anstatt nur einer Wertefunktion (value function) für die meisten Algorithmen des Reinforcement-Learning benötigt?

(*English*: Why do we learn an action value function rather than just a value function in most Reinforcement Learning algorithms?)

The value function only contains information about how good a state is under the current policy, but **no information about what action might be good or bad**. That information, needed for policy improvement, is contained in the action-value function. (2P)

## 12. Aufgabe: AI-Development

a: 4

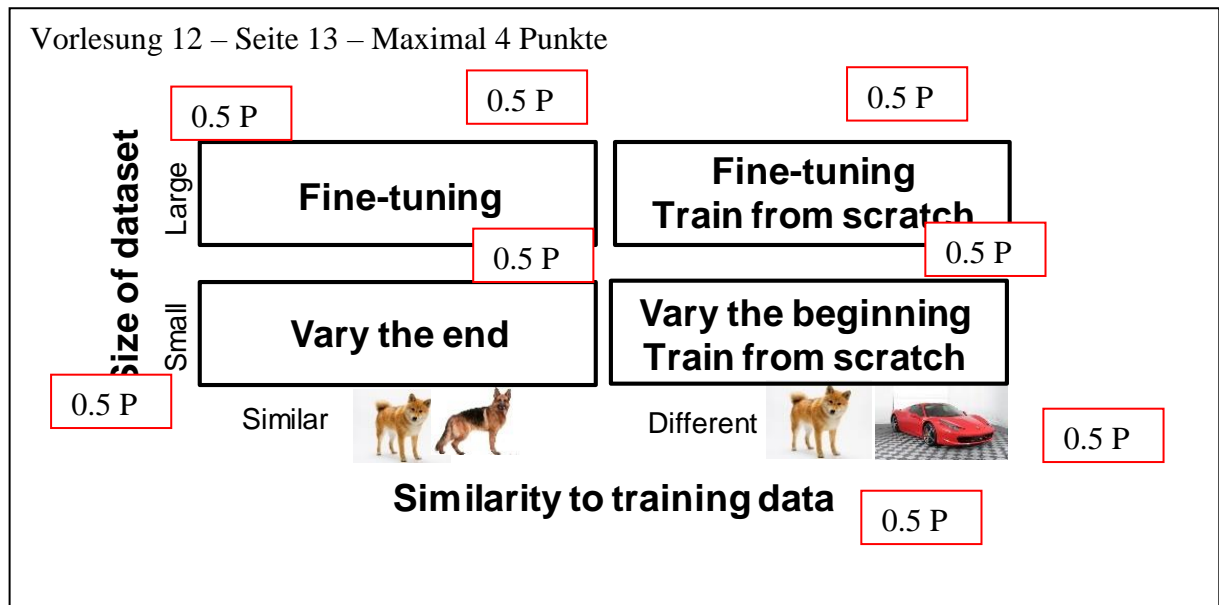
b: 2

c: 6

d: 4

/ 16 Punkte

- a) Skizzieren Sie, welche Möglichkeiten es im Transfer Learning gibt, das Neuronale Netze anzupassen und wodurch die Entscheidung für die finale Auswahl getroffen wird. (*English Translation: Outline which possibilities exist in Transfer Learning to adapt the neuronal network and how the decision for the final selection is made.*)



- b) Nachdem Sie Transfer Learning mit einem kleinen Datenset für ein komplett anderes Problem angewandt haben: Kann es passieren, dass Ihr Netz weiterhin ein Overfitting zeigt? Falls ja, wie würden Sie dies verhindern? (*English Translation: After you have applied Transfer Learning with a small data set for a completely different problem: Can it happen that your network still shows an overfitting? If so, how would you prevent this?*)

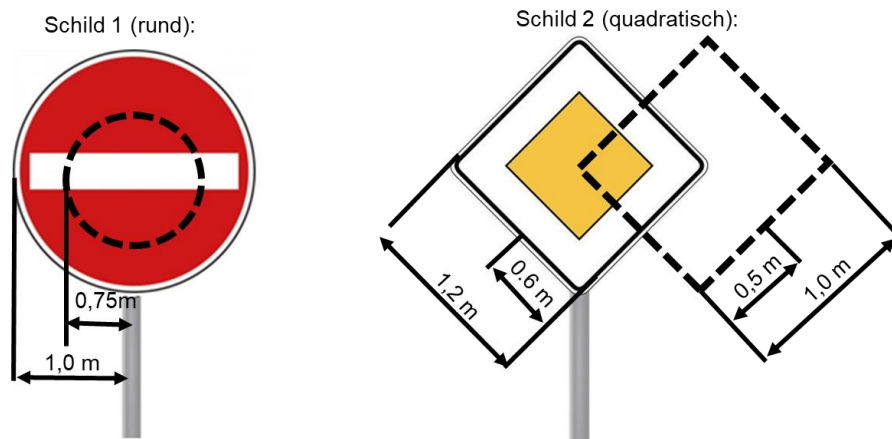
Maximalpunktzahl 2P

Fokus dieser Aufgabe: TRANSFER LEARNING (TL) mit Small Dataset und Different Data

- Ja, Overfitting ist immer noch ein Problem (1P)
- Um Overfitting zu verhindern, müssen im Transfer Learning mit kleinen Datensätzen und unterschiedlichen Daten die Gewichte in den ursprünglichen Layern des Neuronalen Netzes konstant gehalten werden. (1Punkt)  
→ Variation der Daten gilt auch, wenn eine Erklärung dabei steht wie es gemacht wurde



- c) Als Entwickler für die Detektion von Straßenschildern wollen Sie die Methoden der künstlichen Intelligenz nutzen. Sie haben dafür das ihnen bekannte „AlexNet“ eingesetzt und mit einem ihnen vorliegenden Datensatz für Straßenschilder trainiert. Als Ergebnis liefert Ihnen die Detektion für folgende ausgewählte Schilder die Bounding Boxes in gestrichelten Formen. Wählen Sie eine geeignete Metrik zur Beurteilung der Qualität der Detektion und berechnen Sie diesen Wert für beide Schilder. *(English Translation: As a developer for the detection of road signs, you want to use the methods of artificial intelligence. You are using the well known "AlexNet" and you trained this ANN for street signs with a data set available to you. As a result, the detection for the following selected signs provides bounding boxes with the dotted forms. Select a suitable metric to assess the quality of the detection and calculate this value for both signs.)*



Schild 1:

- Area of Union:  $A_1 = r_1^2 \cdot \pi = \pi \cdot 1,0^2 \text{m} = 3.14 \text{ m}^2$  (1 Punkt)
- Area of Overlap:  $A_2 = r_2^2 \cdot \pi = \pi \cdot 0.75^2 \text{m} = 1.767 \text{ m}^2$  (1 Punkt)
- Intersection over Union:  $\text{IoU} = A_2 / A_1 = 1.767 \text{ m}^2 / 3.14 \text{ m}^2 = 0.56 / 56\%$  (1 Punkt)

Schild 2:

- Area of Union:  $A_1 = (1,2\text{m})^2 + (1,0\text{m})^2 - (0,5\text{m} \cdot 0,6\text{m}) = 2,14 \text{ m}^2$  (1 Punkt)
- Fläche Area of Overlap:  $A_2 = (0,5\text{m} \cdot 0,6\text{m}) = 0.3 \text{ m}^2$  (1 Punkt)
- IoU: Intersection over Union:  $\text{IoU} = A_2 / A_1 = 0.3 \text{ m}^2 / 2,14 \text{ m}^2 = 0.14 / 14\%$  (1 Punkt)

- d) Sie möchten das folgende Problem lösen: Entwickeln Sie einen Algorithmus, der genau dann auslöst, wenn er das Wort „Kardinal“ in einem 10 Sekunden langen Audioclip erkennt. Erklären Sie in einem kurzen Absatz, was die beste Vorgehensweise für den Aufbau des Datensatzes ist. *(English Translation: You want to solve the following problem: build a trigger word detection algorithm to spot the word cardinal in a 10 second long audio clip. Explain in a short paragraph what is the best practice for building the dataset.)*

- Data collection can be done by collecting clips of positive ("cardinal") and negative (other words) words, as well as background noise clips. (1P)
- label by hand (1P)
- Data synthesis/Data augmentation can be then performed by overlaying words from the positive/negative clips on the background clips. The labels can be placed simultaneously because the insertion index of the positive word is known (chosen). (1P)
- Building the training/dev/test by splitting all the data 60% / 20% / 20% (1P)