

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

Nachname:	Vorname:
Matrikelnummer:	Studiengang & Semester:

Anmerkungen

- Legen Sie Ihren Studierendenausweis und ein gültiges Ausweisdokument gut sichtbar bereit.
- Tragen Sie Nachname, Vorname, Matrikelnummer, Studiengang & Semester 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 und
 - 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.
- 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 die Prüfer

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

**Aufgabe 1 – Induktives Lernen, Lerntheorie, Entscheidungsbäume
und Unüberwachtes Lernen****___/12P**

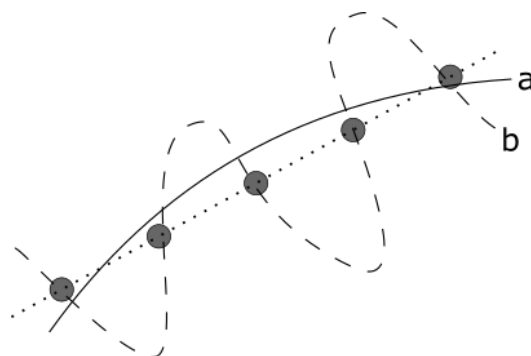
- a) Welche Ziele verfolgen Induktion und Deduktion? Wie erreichen sie diese Ziele?

(___/2P)

- b) Definieren Sie Konsistenz und Vollständigkeit im Hypothesenraum.

(___/2P)

- c) Welche Lernmaschine weist die höhere Kapazität auf? Die Lerndaten werden durch die Kreise dargestellt (Zielfunktion gepunktet). Welches Verhalten bringt eine höhere Kapazität bei Lernmaschinen mit sich?

(___/1P)

d) Wie verhalten sich Lern- und Testfehler beim „Overfitting“?

(___/1P)

e) Geben Sie die Formeln für Genauigkeit (Precision) und True-Positive-Rate (Recall) an.

(___/2P)

f) Welche Klassenverteilung wird bei Entscheidungsbäumen angestrebt? Wie wirkt sich dies auf die Entropie aus? Definieren Sie den Informationsgewinn.

(___/2P)

g) Welche Schritte sind für das „Reduced Error Pruning“ bei Entscheidungsbäumen nötig?

(___/2P)

Aufgabe 2 – Neuronale Netze**___/8P**

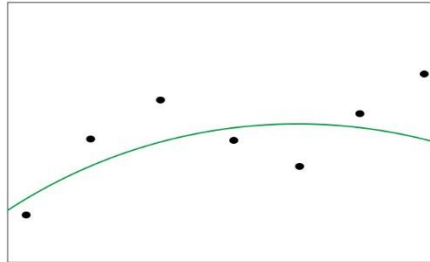
- a) Nennen Sie eine in der Vorlesung behandelte nichtlineare Aktivierungsfunktion und zeichnen Sie das zugehörige Schaubild der Aktivierungsfunktion. (___/1P)

- b) Geben Sie die quadratische Fehlerfunktion E des Gradientenabstiegs an sowie die Formel der iterativen Gewichtsoptimierung $\Delta \vec{w}$ in Abhängigkeit von E . Benennen Sie die verwendeten Variablen. (___/2P)

- c) Nennen Sie zwei Probleme, die bezüglich der Form der Fehlerflächen beim Gradientenabstieg auftreten können. Geben Sie zwei Methoden an, mit denen diese Probleme jeweils vermieden werden können? (___/2P)

- d) Was lässt sich über die VC-Dimension eines neuronalen Netzes sagen, das aus den untenstehenden Lerndaten (Punkte) die eingezeichnete Kurve approximiert? Wie muss die Topologie des Netzes angepasst werden um die Approximation zu verbessern.

(____/1P)



- e) „Je höher die VC-Dimension, umso besser kann das Netz aus einem bestehenden Datensatz lernen, d.h. generalisieren.“ Ist diese Aussage wahr oder falsch, begründen Sie Ihre Entscheidung.

(____/1P)

- f) Was versteht man unter „residual learning“ und wodurch wird damit das Training verbessert?

(____/1P)

Aufgabe 3 – Convolutional Neural Network**___/8P**

a) Für welche Arten von Daten werden CNNs typischerweise verwendet?

(___/1P)

b) Warum besitzen CNNs weniger Parameter/Gewichte als vollvernetzte Neuronale Netze?

(___/2P)

c) Wenden Sie die “max-pooling” Operation mit einer Filtergröße von 2x2, einem Padding von 0 und einem Stride von 1 an. Die Eingabe sieht wie folgt aus:

(___/2P)

1	1	2
0	1	0
3	1	4

d) Wie sehen die Filter von auf realen Bilddaten trainierten CNNs typischerweise aus? Gibt es Unterschiede je nach Position (Tiefe) des Filters im CNN? Beschreiben Sie diese.

(___/2P)

- e) Was ist ein Vorteil von Fully Convolutional Networks gegenüber CNNs mit Fully-Connected Schichten am Ende und wofür werden diese typischerweise verwendet?

(___/1P)

Aufgabe 4 – Support Vector Machines**___/8P**

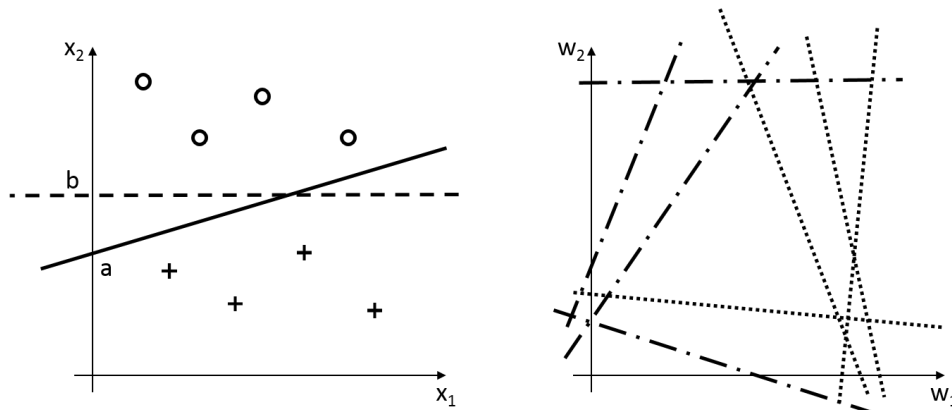
- a) Bezogen auf einen Support Vektor Klassifikator, beschreiben Sie das Problem welches gelöst wird, die Lösung die gefunden wird und Intuition für die Lösung. (___/1,5P)

- b) Formulieren Sie mathematisch das grundlegende Optimierungsproblem für einen linearen Support Vector Klassifikator. Geben Sie außerdem die Nebenbedingungen für die Optimierung an. (___/1,5P)

- c) Wie viele Stützvektoren werden für die eindeutige Lösung eines binären Klassifikationsproblems mindestens benötigt, wenn der Merkmalsraum $n > 0$ Dimensionen hat? (___/1P)

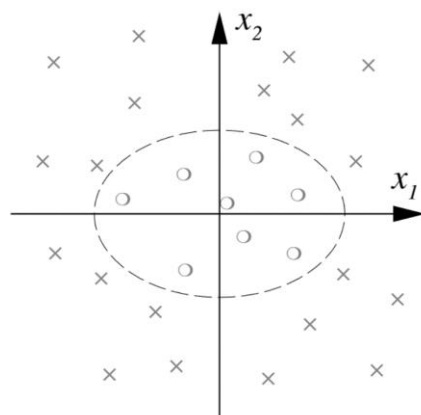
- d) Welche der beiden Hypothesen im linken Graph ist die optimale Hypothese? Markieren Sie die entsprechenden Stützvektoren (Support Vectors) im linken Graphen (Merkmalsraum) und zeichnen Sie die Hypothese (Parameter w) in den rechten Graphen (Hypothesenraum) ein.

(___/3P)



- e) Geben Sie eine gültige Transformationsregel an um die Daten in einem anderen Raum linear trennen zu können.

(___/1P)



Aufgabe 5 – Reinforcement Learning**___/12P**

- a) Was ist der wesentliche Unterschied zwischen Reinforcement Learning und überwachtem Lernen mit Bezug auf die Fehlerberechnung? (___/1P)

- b) Beschreiben Sie die Begriffe Zustand und Beobachtung und geben sie je ein Beispiel für einen Zustand und eine Beobachtung. (___/2P)

- c) Beschreiben Sie die Begriffe „Bootstrapping“ und „Sampling“ für die Wertaktualisierungen beim Reinforcement Learning. (___/3P)
Ordnen sie die Verfahren: Policy Iteration, SARSA und Monte Carlo Methoden den beiden Begriffen zu.

- d) Ein Agent wird mit Hilfe von Q-Learning trainiert. Wie lautet die Formel zur Berechnung des TD Fehlers? (___/2P)

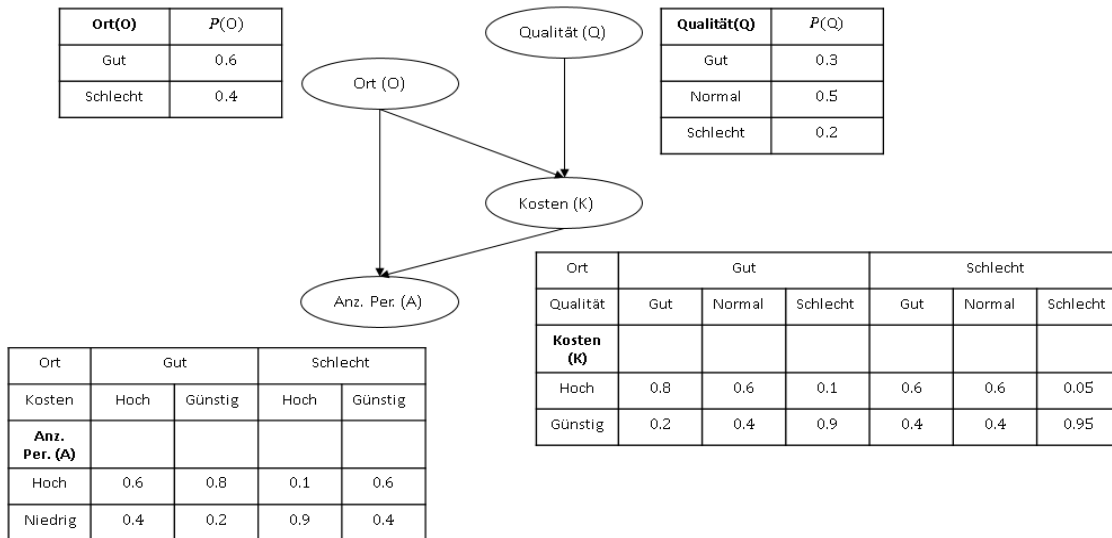
Der Agent führt eine Aktion a in s aus und erhält dabei eine Belohnung r in Höhe von 1 und wird in einen neuen Zustand s' überführt. Die approximierte Aktionswertfunktion hat in s' ein Maximalwert von 10. Die approximierte Aktionswertfunktion hat für (s, a) einen Wert von 8, der Diskontierungsfaktor beträgt 0,9. Bestimmen sie den TD-Fehler.

- e) Um die Varianz bei dem REINFORCE Algorithmus zu reduzieren kann eine „Baseline“ verwendet werden. Definieren sie eine geeignete „Baseline“ formal. Ist die Verwendung einer konstanten „Baseline“ zulässig? Begründen Sie Ihre Antwort. (___/2P)

- f) Warum bezeichnet man strategiebasierte Verfahren wie den REINFORCE Algorithmus als „on-policy“ Algorithmen? Warum kann die Strategieaktualisierung beim REINFORCE Algorithmus ohne „Baseline“ zu einer Verschlechterung führen? (___/2P)

Aufgabe 6 – Bayes HMM und SPN**___/12P**

Um ein gutes Restaurant auszuwählen, können vier Faktoren berücksichtigt werden: die Qualität (Q) und die Kosten (K) des Essens, der Ort des Restaurants (O) und die Anzahl der Personen (A), die das Restaurant besuchen. Das folgende Bayes-Netzwerk beschreibt die Beziehung zwischen den Faktoren



- a) Geben Sie die zugehörige Faktorisierung der Verbundwahrscheinlichkeitsdichte an:

(___/1P)

$$P(O, Q, K, A)$$

$$P(O, Q, K, A) =$$

- b) Wie hoch ist die Wahrscheinlichkeit, ein Restaurant zu besuchen, dessen Ort und Qualität des Essens gut sind, dessen Kosten niedrig sind und dessen Besucherzahl gering ist?
(Hinweis: Es ist hinreichend bei der Berechnung die korrekten Multiplikatoren aufzuschreiben)

(___/2P)

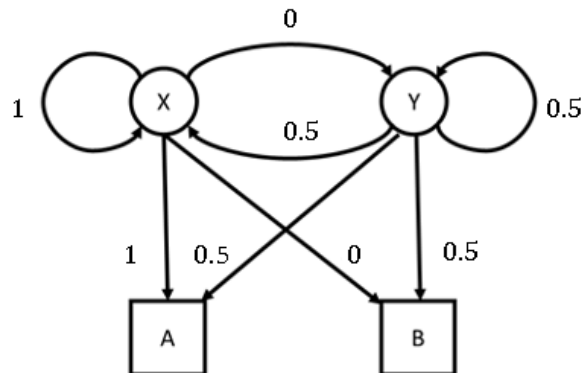
$$P(O = \text{Gut}, Q = \text{Gut}, K = \text{Günstig}, A = \text{niedrig}) =$$

- c) Angenommen, der Preis des Essens ist für die Entscheidung nicht wesentlich, wie hoch ist dann die Wahrscheinlichkeit, ein Restaurant zu besuchen, dessen Ort gut, die Qualität des Essens gut und die Anzahl der Personen gering sind? (Hinweis: Es ist hinreichend bei der Berechnung die korrekten Multiplikatoren aufzuschreiben)

(___/3P)

$$P(O = \text{Gut}, Q = \text{Gut}, A = \text{niedrig}) =$$

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 $S_0 = (0,1; 0,9)$.



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

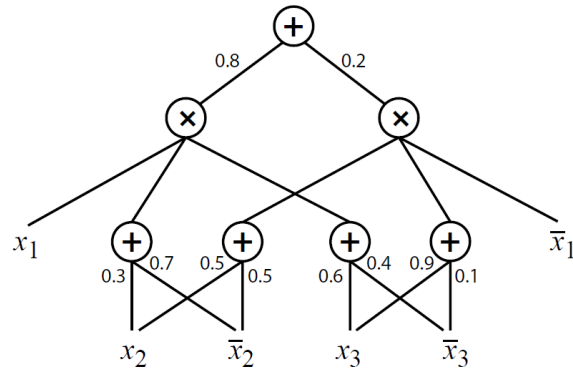
(___/2P)

- e) Sagen Sie den (System-)Zustand des Hidden-Markov-Modells zum Zeitpunkt $t = \infty$ voraus (Prädiktion).

(___/1P)

- f) Gegeben ist folgendes Sum-Product Netz (SPN), das eine Wahrscheinlichkeitsverteilung über die Zufallsvariablen X_1, X_2 und X_3 mit Hilfe der Indikatorvariablen $x_1, \bar{x}_1, x_2, \bar{x}_2, x_3, \bar{x}_3$ kodiert. Berechnen Sie die Wahrscheinlichkeiten der Belegung der folgenden Zufallsvariablen und tragen Sie diese in die Tabelle ein.

(___/3P)



X_1	X_2	X_3	$\Phi(X)$
1	1	1	
0	1	1	
0	0	0	