

Übungsklausur: Künstliche Intelligenz 2

am 11.02.2022

Name, Vorname: _____

Matrikelnummer: _____

Studiengang: _____

Punkteübersicht:

Aufgabe Nr.:	1	2	3	4	5	6	Summe
Punktzahl:	11	8	14	12	8	7	60
Davon erreicht:							

Hinweise:

- Bitte zuerst Name, Matrikelnummer und Studiengang eintragen!
- Die Bearbeitungszeit beträgt 60 Minuten.
- Veranschlagen Sie für jede Aufgabe so viele Minuten, wie sie Punkte hat.

Aufgabe 1: Aufwärmaufgaben**11 Punkte**

- (a) (2 Punkte) Leiten Sie die Formel von Bayes für $P(B|A)$ ausgehend von der Definition einer bedingten Wahrscheinlichkeit für die Ereignisse A und B her.
- (b) (2 Punkte) Es sei $P(A) = 80\%$ und $P(\neg A|B) = 20\%$ gegeben. Prüfen Sie, ob A von B unabhängig ist. Geben Sie dafür zunächst allgemein an, welche Formel für Unabhängigkeit gegeben sein muss und setzen Sie dann zum Prüfen ein.
- (c) (3 Punkte) Bei einem Kniffel-Spiel haben Sie im ersten Wurf mit fünf Würfeln eine 2,2,3,4,5 gewürfelt. Da Sie sich eine große Straße erhoffen (Reihe von 1 bis 5 oder 2 bis 6), nehmen Sie einen der beiden Zweier und würfeln mit diesem weiter.
- Wie groß ist die Wahrscheinlichkeit bei maximal zwei weiteren Würfeln eine große Straße zu erhalten (Formel genügt)?

- (d) (2 Punkte) Was ist die grundlegende Idee von tiefen Neuronalen Netzen, d. h. Deep Learning, im Vergleich zu einfacheren linearen Klassifikatoren, z. B. SVM oder Perzeptron?

- (e) (2 Punkte) Berechnen Sie die Convolution-Operation auf folgenden Daten (Stride 1, ohne Padding):

3	1	2
1	2	1
-1	0	-1

mit der Filter-Matrix

1	-1
0	1

Aufgabe 2: Naive Bayes

8 Punkte

An Ihrem Campus gibt es zwei verschiedene Mensen. Die „Mensateria“ (a) liegt näher und hat die größere Essensauswahl, ist aber oft überfüllt. Die „Interimsmensa“ (b) ist zwar weiter weg und hat die geringere Auswahl, bietet dafür aber stets genug Platz.

Bei Ihren letzten tatsächlichen Mensabesuchen notierten Sie das Wetter, die Qualitäten der Essensauswahl in den beiden Mensen, ob die Mensateria überfüllt ist und natürlich Ihre Entscheidung.

Wetter	Auswahl (a)	Auswahl (b)	Überfüllt (a)	Mensa
regnerisch	gut	schlecht	ja	a
regnerisch	normal	normal	nein	a
sonnig	normal	schlecht	ja	b
regnerisch	gut	schlecht	nein	a
regnerisch	normal	normal	ja	b
sonnig	gut	schlecht	nein	a
sonnig	gut	normal	ja	b
sonnig	normal	normal	nein	b

- (a) (4 Punkte) Füllen Sie zunächst die folgenden Tabellen an bedingten Wahrscheinlichkeiten mit Hilfe der Informationen aus dem obigen Datensatz aus. Die Tabellen sind so zu lesen, dass jeder Wert stets in Abhängigkeit von der ausgewählten Mensa ist, also z.B. $P(\text{regnerisch}|a)$ (Angabe in Brüchen genügt):

Wetter		Auswahl (a)	
	Mensa a Mensa b		Mensa a Mensa b
regnerisch		gut	
sonnig		normal	

Auswahl(b)		Überfüllt (a)	
	Mensa a Mensa b		Mensa a Mensa b
normal		ja	
schlecht		nein	

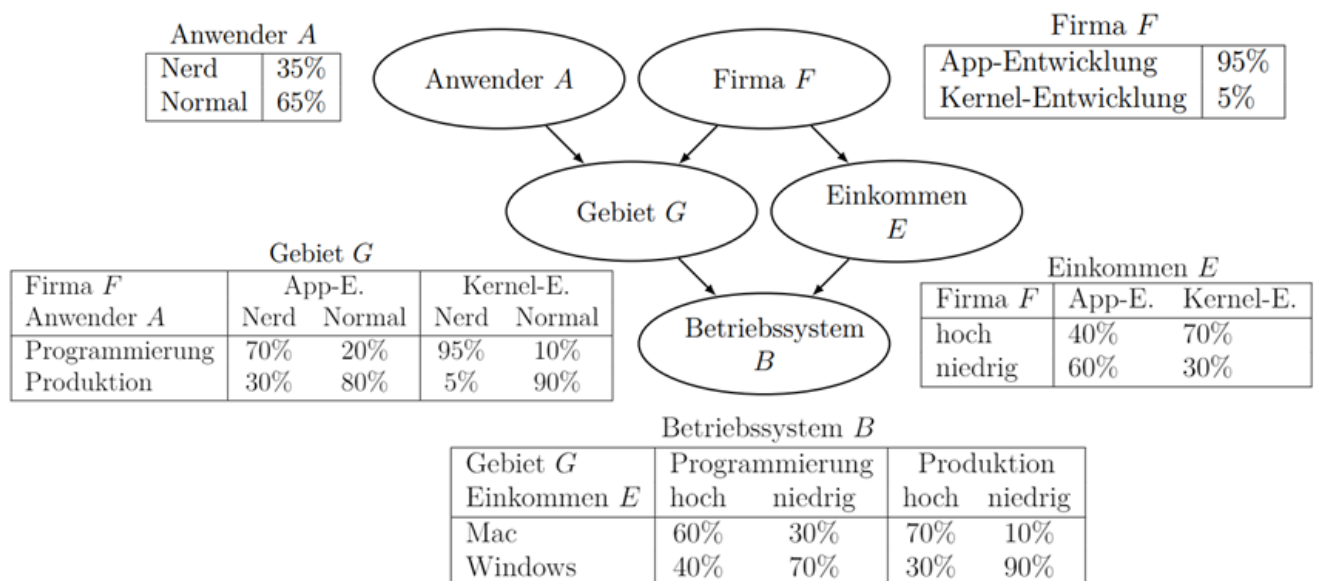
- (b) (4 Punkte) Geben Sie die Wahrscheinlichkeit für Mensa a und Mensa b unter den Beobachtungen „Wetter = sonnig, Auswahl (a) = gut, Auswahl (b) = schlecht, Überfüllt (a) = ja “ basierend auf einem Naive-Bayes Klassifikator an.

Aufgabe 3: Bayes'sche Netze

14 Punkte

Der Einsatz eines Betriebssystems ist von vielen Faktoren abhängig. In dieser Darstellung wird vereinfacht angenommen, dass er nur vom Einsatzgebiet und vom Einkommen einer Person abhängig sei. Weiterhin sei angenommen, dass sowohl der Anwender als auch die Firma Einfluss auf das Anwendungsgebiet haben. Das Einkommen ist von der Firma abhängig. Zusätzlich nehmen wir zur Vereinfachung nur je zwei Ausprägungen pro Attribut an.

Zu diesem Sachverhalt sei das folgende Bayes'sche Netz mit den zugehörigen Wahrscheinlichkeiten gegeben:



- (a) (2 Punkte) Es gibt eine Reihe an Sampling-Verfahren, wie das Likelihood-Sampling, die zum Abschätzen von Wahrscheinlichkeiten verwendet werden. Aus welchem Grund werden Sampling-Verfahren verwendet, um auch prinzipiell exakt berechenbare Wahrscheinlichkeiten abzuschätzen? Nennen Sie zusätzlich ein weiteres Sampling-Verfahren.

- (b) (6 Punkte) In der folgenden Tabelle sind einige Samples aufgelistet, die nach dem Likelihood-Weighting-Verfahren gezogen wurden. Schätzen Sie daraus die gesuchten Wahrscheinlichkeiten $P(A = \text{Nerd} | F = \text{Kernel}, G = \text{Produktion})$ und $P(A = \text{Normal}, B = \text{Windows} | F = \text{Kernel}, G = \text{Produktion})$ ab.

Ergänzen Sie dazu zunächst die fehlenden Gewichtungen der Samples und berechnen Sie anschließend die gesuchten Wahrscheinlichkeiten.

	A	F	G	E	B	Gewicht
Sample	Normal	Kernel	Produktion	niedrig	Mac	
Sample	Nerd	Kernel	Produktion	niedrig	Mac	
Sample	Normal	Kernel	Produktion	niedrig	Windows	
Sample	Nerd	Kernel	Produktion	hoch	Windows	
Sample	Normal	Kernel	Produktion	hoch	Mac	
Sample	Normal	Kernel	Produktion	hoch	Windows	

Hinweis: Die Angabe eines Bruches genügt.

- $P(A = \text{Nerd} | F = \text{Kernel}, G = \text{Produktion})$

- $P(A = \text{Normal}, B = \text{Windows} | F = \text{Kernel}, G = \text{Produktion})$

- (c) (6 Punkte) Berechnen Sie *exakt* die kombinierte Wahrscheinlichkeit der vier Variablen $P(A = \text{Normal}, F = \text{App}, G = \text{Produktion}, B = \text{Windows})$. *Hinweis:* Das Produkt, bzw. die Summe aus Wahrscheinlichkeiten muss eingesetzt, aber nicht ausgerechnet werden.

Aufgabe 4: Viterbi-Algorithmus

12 Punkte

- (a) (8 Punkte) Berechnen Sie mithilfe des Viterbi-Algorithmus die wahrscheinlichste Sequenz versteckter Zustände $S = A, B, C$ für die beobachtete Folge 1312 für das gegebene Hidden-Markov-Modell mit den folgenden Wahrscheinlichkeiten:

Emissionswahrscheinlichkeiten			
Zustand	A	B	C
1	0.1	0.5	0
2	1	0	0
3	0	0.5	0.5

Übergangswahrscheinlichkeiten				
	Aktuell			
↓ Neu	Start	A	B	C
A	0.2	0	0.3	0
B	0.4	1	0.2	1
C	0.4	0	0.5	0

Zeichnen Sie zur Visualisierung einen Graphen in welchem die nötigen Rechenschritte nachvollziehbar sind. Pfade mit Wahrscheinlichkeit Null können direkt ignoriert werden.

- (b) (2 Punkte) Warum kann ein Modell, bei dem die Übergangswahrscheinlichkeiten von einem in den nächsten versteckten Zustand nicht nur vom direkten Vorgänger, sondern von dem Vorgänger und dem Vorvorgänger abhängig ist, nicht mehr direkt mit dem Viterbi-Algorithmus behandelt werden?
- (c) (2 Punkte) Wie kann das Problem aus Aufgabenteil (b) umformt werden, sodass es dennoch mit dem Viterbi-Algorithmus berechenbar ist?

Aufgabe 5: Deep Learning**8 Punkte**

- (a) (2 Punkte) Wie lautet allgemein die Updateregeln des Stochastic Gradient Descent für die Gewichte W , der Loss-Funktion $L(x, y)$ eines Beispielpaares $\{x, y\}$ und der Lernrate α ?

- (b) (3 Punkte) Gegeben sei die folgende Loss-Funktion L eines neuronalen Netzes $f(x)$ im eindimensionalen Fall für eine Trainingsmenge $\{x, y\}$ der Größe N :

$$L = \frac{1}{4} \sum_i^N [(y_i - f(x_i))^4 + \lambda W]$$

mit

$$f(x_i) = W x_i ,$$

wobei W ein lernbarer Parameter und λ eine Konstante sind.

Geben Sie die Formel zum Update der Gewichte W gemäß des Gradient Descent an.

Hinweis: Wenn Sie die Update-Formel (Aufgabenteil (a)) nicht wussten, berechnen Sie zumindest die Ableitung als wichtigsten Teilschritt.

- (c) (3 Punkte) Eine besonders bei neuronalen Netzen vorhandene Schwierigkeit ist, dass je nach Aufgabenbereich oft nur wenige Trainingsdaten zur Verfügung stehen. Beschreiben Sie, welches Problem dadurch auftreten kann und nennen Sie zwei Möglichkeiten, um das Problem zu minimieren.

