

Name: _____ Vorname: _____ Matrikel-Nr.: _____

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

WiSe 2022/2023

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	15.02.2023
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar) Aufgabenblock A: beigegefügte Formelsammlung Aufgabenblock D: Anwendungsbeispiele 1 und 2.

Die Lösungen zu den Aufgaben sollen gegliedert und in vollständigen zusammenhängenden Sätzen dargestellt und Rechnungen mit ihren Zwischenschritten nachvollziehbar sein. Die Darstellungsform und die Systematik der Gedankenführung gehen in die Bewertung ebenfalls ein. In Klammern ist für jede Aufgabe die Anzahl der maximal möglichen Punkte angegeben, die bei einer richtigen und vollständigen Bearbeitung erreicht werden können. Sie entspricht in etwa dem erwarteten Zeitbedarf in Minuten.

Bearbeiten Sie Aufgabenblock A und einen weiteren Aufgabenblock!

Insgesamt können **90 Punkte** erreicht werden. Für eine erfolgreiche Bearbeitung müssen wenigstens **45 Punkte** erworben werden. Bei Bearbeitung von mehr als zwei Aufgabenblöcken wird die Bearbeitung des Aufgabenblocks A und des ersten weiteren bearbeiteten Blocks gewertet. Sie haben **90 Minuten** Bearbeitungszeit zur Verfügung.

Die Klausur besteht mit diesem Deckblatt aus insgesamt 10 (zehn) Seiten.

Unterschrift: _____

Aufgabenblock A:

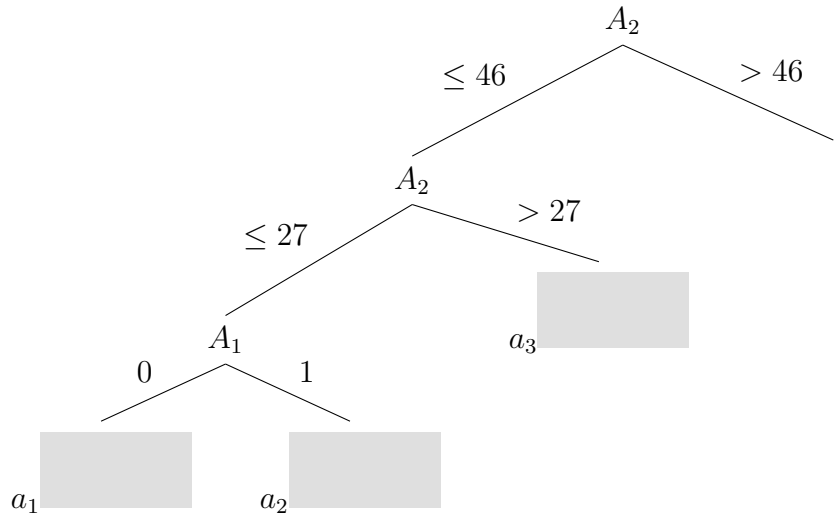
Knowledge-based systems

Aufgabe A.1: Entscheidungsbäume

[15 Punkte]

Sei der folgende Datensatz mit 9 Fällen, drei Attributen A_1, A_2, A_3 und der Klassenmenge $C = \{0, 1, 2\}$ gegeben. Attribut A_1 ist binär und die Attribute A_2 und A_3 sind kontinuierlich. Mithilfe des CART-Verfahrens wurde ein noch unvollständiger Entscheidungsbaum konstruiert.

i	A_1	A_2	A_3	C
1	0	55	100	1
2	0	53	150	1
3	1	33	163	2
4	1	30	98	2
5	1	61	120	1
6	0	25	170	2
7	1	20	137	1
8	0	64	143	0
9	0	73	134	0



- Vervollständigen Sie die linke Seite des Entscheidungsbaums, indem Sie für die drei Blattknoten a_1, a_2, a_3 die resultierende Klassifikation in den grauen Kästchen notieren. (3 Punkte)
- Bestimmen Sie mittels des *Gini impurity*-Maßes den Test, der zur nächsten Verzweigung der rechten Baumhälfte am geeignetsten ist. Testen Sie dabei jedes Attribut genau einmal. Bezüglich der Attribute A_j mit kontinuierlichen Werten wird der Mittelwert M_j der zugrundeliegenden Fälle des Knotens im Entscheidungsbaum herangezogen, also die Unterscheidung von $A_j \leq M_j$ und $A_j > M_j$. Erweitern Sie anschließend den obigen Baum. Wird nun der Datensatz vollständig und fehlerfrei klassifiziert? (9 Punkte)
- Begründen Sie, warum kein Schwellwert von A_3 existiert, sodass die Fälle mit $A_2 > 46$ sofort eindeutig klassifiziert werden. (3 Punkte)

Aufgabe A.2: Thesen**[15 Punkte]**

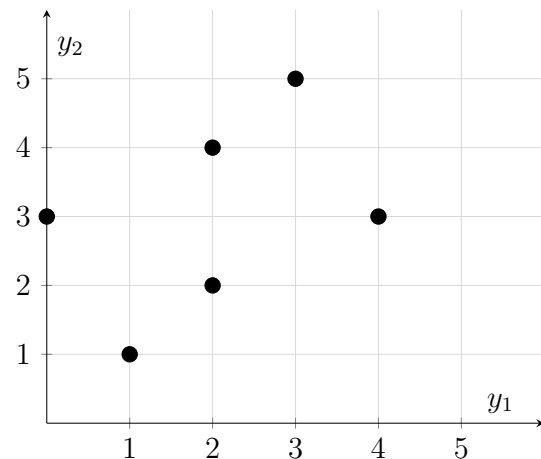
Nehmen Sie zu den folgenden Thesen begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort erhält keine Punkte.

- a) Angewendet wird das binäre AdaBoost-Verfahren. Sei nun ein Datensatz mit der folgenden Struktur gegeben:
 Es liegen $2l$ viele Fälle mit (x_i, y_i) vor, wobei $y_i \in \{-1, +1\}$ gilt. Außerdem ist $x_i = x_{i+l}$ und $y_i \neq y_{i+l}$ für alle $i \in \{1, \dots, l\}$. Dann bricht das AdaBoost-Verfahren in der ersten Iteration ab. (5 Punkte)
- b) Betrachtet wird die Iteration t im AdaBoost.M2-Verfahren. Für den schwachen Schätzer $h_t : X \times Y \rightarrow \{1\}$ erhält man einen Pseudo-loss von 0,5. (5 Punkte)
- c) Ist p direkt erreichbar (*directly density-reachable*) von q , so ist auch q direkt erreichbar von p . (5 Punkte)

Aufgabe A.3: Clustering, Separation und nächste Nachbarn**[15 Punkte]**

Wir betrachten die Datenpunkte x_1, \dots, x_6 .

i	y_1	y_2
1	0	3
2	1	1
3	2	2
4	3	5
5	4	3
6	2	4



Hinweis: Verwenden Sie in allen Teilaufgaben das *Manhattan-Distanzmaß* (*rectangular distance measure*).

- a) Bestimmen Sie die aktuellen Cluster C_1 und C_2 für die Zentren $c_1 = (1, 2)$ und $c_2 = (3, 4)$ wie im k -means Verfahren. (2 Punkte)
- b) Bestimmen Sie die im Rahmen des k -means Verfahren aus den Clustern C_1 und C_2 resultierenden neuen Zentren c'_1 und c'_2 . (3 Punkte)
- c) Zeigen Sie, dass die Punktmengen $M_+ := \{x_4, x_5, x_6\}$ und $M_- := \{x_1, x_2, x_3\}$ linear separabel sind, indem Sie passende Gewichtsparameter und den Wert θ für ein lineares Perzeptron angeben (es ist keine Durchführung eines Algorithmus notwendig!). Überprüfen Sie die Korrektheit des Perzeptrons nochmals anhand der Punkte in M_+ und M_- . (6 Punkte)
- d) Klassifizieren Sie den Punkt $x_7 = (3, 1)$ mit Hilfe des einfachen 3-nearest-neighbor Verfahrens. Die Punkte x_1, x_2 und x_3 sind dabei der Klasse A und die Punkte x_4, x_5 und x_6 der Klasse B zugeordnet. (4 Punkte)

Formelsammlung zum Aufgabenblock A

$$\text{gain}(M, A_j) = \max \{ \text{gain}(M, A_l) \mid l \in \{1, \dots, k\} \}$$

$$\text{gain}(M, A_j) = \text{info}(M) - \text{info}(M, A_j)$$

$$\text{info}(M) = - \sum_{i=1}^k \frac{\text{freq}(C_i, M)}{\text{freq}(M)} \cdot \log_2 \left(\frac{\text{freq}(C_i, M)}{\text{freq}(M)} \right)$$

$$\text{info}(M, A_j) = \sum_{i=1}^l \frac{\text{freq}(T_i)}{\text{freq}(M)} \cdot \text{info}(T_i)$$

$$\text{prob}(C_i, M) = \frac{\text{freq}(C_i, M)}{\text{freq}(M)} = \frac{\text{freq}(C_i, M)}{|M|}$$

$$G(b_l) = \frac{\#(b_l, l)}{m} \cdot \left(1 - \sum_{i=1}^k N(i, b_l, l)^2 \right) + \frac{\#(b_l, r)}{m} \cdot \left(1 - \sum_{i=1}^k N(i, b_l, r)^2 \right)$$

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)}$$

$$d^\alpha(\mathbb{X}_1, \mathbb{X}_2) = \left(\sum_{j=1}^d |x_{1,j} - x_{2,j}|^\alpha \right)^{1/\alpha}$$

$$h_t = \text{argmin} \{ \epsilon_j \mid \epsilon_j = \sum_{i=1}^m D_t(i) [y_i \neq h_j(x_i)] \wedge h_j \in \mathcal{X} \}$$

$$\epsilon_t = \frac{1}{2} \sum_{(i,y) \in B} D_t(i, y) \cdot (1 - h_t(x_i, y_i) + h_t(x_i, y))$$

$$w_{i,j} = \frac{1}{\sum_{C^k=1} \left(\frac{\|x_j - \mu_i\|}{\|x_j - \mu_k\|} \right)^{\frac{2}{m-1}}} = \frac{1}{\|x_j - \mu_i\|^{\frac{2}{m-1}} \cdot \sum_{C^k=1} \|x_j - \mu_k\|^{\frac{-2}{m-1}}}$$

$$\mu_i = \sum_{j=1}^n \frac{(w_{i,j})^m}{\sum_{j=1}^n (w_{i,j})^m} \cdot x_j = \frac{1}{\sum_{j=1}^n (w_{i,j})^m} \cdot \sum_{j=1}^n (w_{i,j})^m \cdot x_j$$

$$c_i = \frac{1}{|C_i|} \cdot \sum_{x \in C_i} x$$

$$p(x') = \frac{D(x')^2}{\sum_{x \in X} D(x)^2}$$

Aufgabenblock B:

Computer Hardware und Systembetrieb

(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

Aufgabe B.1: Von Neumann Flaschenhals

[5 Punkte]

Beschreiben Sie den üblicherweise als von Neumann Flaschenhals bezeichneten Sachverhalt in der von Neumann Architektur.

Aufgabe B.2: Superskalare CPUs

[10 Punkte]

- a) Was ist eine superskalare CPU? (3 Punkte)
- b) Mit welchen Architekturmerkmalen erreicht man Superskalarität? (3 Punkte)
- c) Welche neuen Probleme (Flaschenhälse) entstehen bei superskalaren CPUs? (4 Punkte)

Aufgabe B.3: Instruction Reordering

[10 Punkte]

- a) In welchen Fällen ist Instruction Reordering sinnvoll? (5 Punkte)
- b) Welche Voraussetzungen an die Prozessorhardware müssen dabei erfüllt sein? (3 Punkte)
- c) Was ist zu beachten, um keine falschen Resultate zu erhalten? (2 Punkte)

Aufgabe B.4: SMT vs. Multi-Core

[10 Punkte]

Erläutern Sie die Hauptunterschiede zwischen einem Prozessor mit 2 Kernen und einem Prozessor mit einem Kern und 2-Weg SMT.

Aufgabe B.5: Speicherhierarchien

[10 Punkte]

- a) Skizzieren Sie eine typische Speicherhierarchie vom Prozessor bis zu Archivierungsmedien. (6 Punkte)
- b) Welche Auswirkungen von Memory Wall und Power Wall könnten den Hauptspeicher zukünftig betreffen? (4 Punkte)

Aufgabenblock C:

Kommunikationssysteme (Internet-Technologien)

Aufgabe C.1: Internet [15 Punkte]

- a) Was versteht man unter dem Internet aus physischer Sicht bzw. aus logischer Sicht? (2 Punkte)
- b) Erläutern Sie die Aufgaben und Eigenschaften der einzelnen Schichten des Internet-Referenzmodells. (4 Punkte)
- c) Nennen Sie die wichtigsten Einträge im Header eines IP-Datagramms (IP-Header) und erläutern Sie jeweils deren Zweck. (4 Punkte)
- d) Nennen Sie mindestens fünf Vorteile von IPv6 gegenüber IPv4. (5 Punkte)

Aufgabe C.2: Netzzugang und Routing [15 Punkte]

- a) Welche Aufgabe hat ein DHCP-Server in einem Netzwerk? Welche Netzwerk-Konfigurationsdaten auf welcher Schicht des Internet-Protokollstapels eines Hosts sind unbedingt für die Kommunikation im Internet erforderlich? (4 Punkte)
- b) Erläutern Sie die im Internet für die Vermittlung der IP-Datagramme genutzten Methoden Speichervermittlung und Teilstreckenvermittlung. Was versteht man dabei unter Quellenunabhängigkeit? (3 Punkte)
- c) Welchem Zweck dient die IP-Adresshierarchie? Welche Informationen benötigt ein IP-Router für universelles und optimales Routing? (4 Punkte)
- d) Stellen Sie eine Skizze dar, wie ein Datenpaket aus einem privaten Netz modifiziert wird, wenn es einen NAT-Router hin zum Internet passiert (Masquerading), und wie die Antwortpakete beim Eintritt in das private Netz modifiziert werden. (4 Punkte)

Aufgabe C.3: E-Mail [15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren für die Übermittlung einer E-Mail vom E-Mail-Client des Absenders zum E-Mail-Client des Empfängers. (5 Punkte)
- b) Um welche Funktionalität werden E-Mails durch den MIME-Standard erweitert? (2 Punkte)
- c) Beschreiben Sie, wie sich E-Mail-Anhänge für sichere und vertrauenswürdige E-Mails nutzen lassen. Welche Daten werden in einer per S/MIME hybrid verschlüsselten und unterzeichneten E-Mail übertragen und welche Schlüssel sind an welcher Stelle erforderlich? (8 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

Bearbeiten Sie 3 der folgenden 4 Aufgaben! (insgesamt 45 Punkte)

Aufgabe D.1: Relationenmodell

[15 Punkte]

- a) Erläutern Sie die strukturellen Elemente des Relationenmodells Typ, Attribut, Tupel und Relation sowie die Zusammenhänge zwischen ihnen. (12 Punkte)
- b) Im Anhang sehen Sie das Bild der Relation für Lieferanten. Welche Mengen einer Relation können Sie identifizieren? Geben Sie diese als Stichworte an. (3 Punkte)

Aufgabe D.2: Architektur

[15 Punkte]

- a) Beschreiben Sie 8 Ziele verteilter Datenbanksysteme in jeweils einem Satz. (10 Punkte)
- b) Was ist der Systemkatalog eines Relationen Datenbanksystems? Welche Informationen legt ein RDBS im Systemkatalog ab? (5 Punkte)

Aufgabe D.3: Sicherheit

[15 Punkte]

- a) Beschreiben Sie den Prozess der Authentifizierung und Autorisierung eines Datenbanksystems. (5 Punkte)
- b) Diskutieren Sie die Sicherheitskonzepte *Discretionary Access Control (DAC)* und *Label-based Access Control (LBAC)* im Vergleich. (10 Punkte)

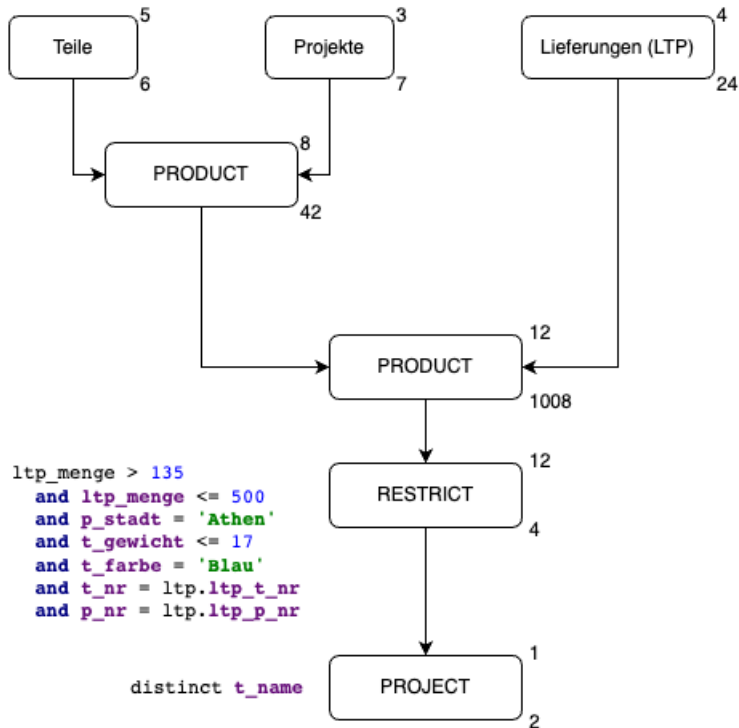
Bitte beachten Sie für Aufgabe D.4 die Folgeseite!

Aufgabe D.4: Anfrageverarbeitung

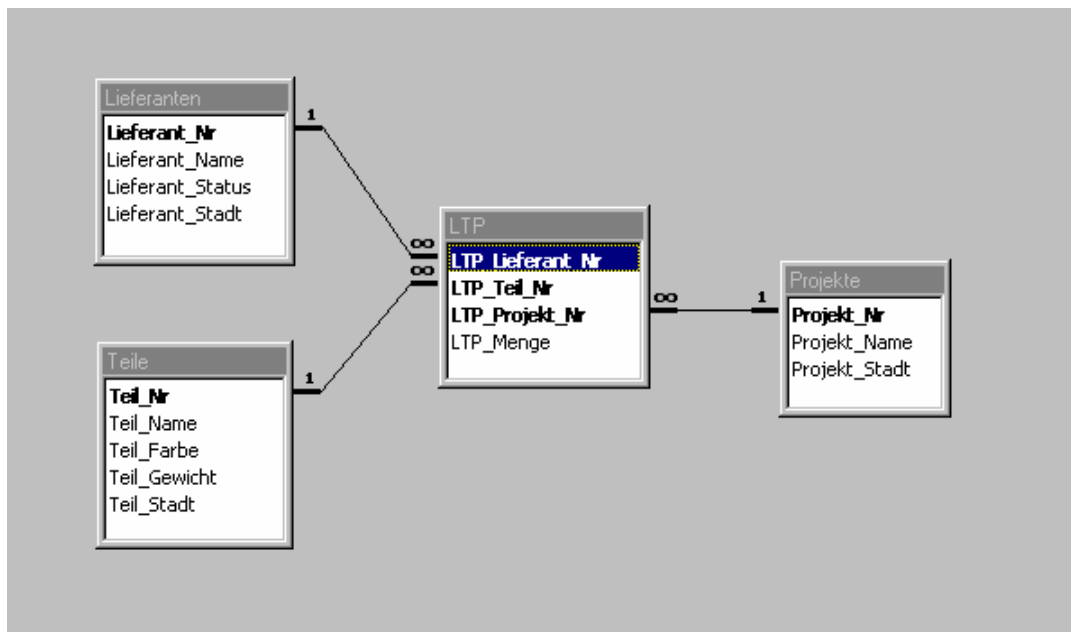
[15 Punkte]

a) In der folgenden Abbildung sehen Sie eine *SELECT*-Anweisung sowie den vom Datenbanksystem generierten initialen logischen Anfrageplan (iLAP). Erstellen Sie einen verbesserten logischen Anfrageplan (LAP). Ergänzen Sie (in Stichworten) die von Ihnen angewendeten Optimierungsmethoden und/oder -heuristiken, die Sie für die Erstellung des verbesserten LAP angewendet haben. (10 Punkte)

```
select distinct t_name
from ltp,
teile,
projekte
where ltp_menge > 135
and ltp_menge <= 500
and p_stadt = 'Athen'
and t_gewicht <= 17
and t_farbe = 'Blau'
and t_nr = ltp.ltp_t_nr
and p_nr = ltp.ltp_p_nr;
```



b) Erläutern Sie den Begriff der Semantischen Anfrageoptimierung und die mit dieser verbundenen Optimierungspotentiale. (5 Punkte)



Lieferanten

Lieferant_Nr	Lieferant_Name	Lieferant_Status	Lieferant_Stadt
L1	Smith	20	London
L2	Jones	10	Paris
L3	Blake	30	Paris
L4	Clark	20	London
L5	Adams	30	Athen

Teile

Teil_Nr	Teil_Name	Teil_Farbe	Teil_Gewicht	Teil_Stadt
T1	Mutter	Rot	12	London
T2	Bolzen	Grün	17	Paris
T3	Schraube	Blau	17	Rom
T4	Schraube	Rot	14	London
T5	Nocken	Blau	12	Paris
T6	Zahnrad	Rot	19	London

Projekte

Projekt_Nr	Projekt_Name	Projekt_Stadt
P1	Sortierer	Paris
P2	Bildschirm	Rom
P3	OCR	Athen
P4	Konsole	Athen
P5	RAID	London
P6	EDS	Oslo
P7	Bandlaufwerk	London

LTP – Lieferungen

LTP_Lieferant_Nr	LTP_Teil_Nr	LTP_Projekt_Nr	LTP_Menge
L1	T1	P1	200
L1	T1	P4	700
L2	T3	P1	400
L2	T3	P2	200
L2	T3	P3	200
L2	T3	P4	500
L2	T3	P5	600
L2	T3	P6	400
L2	T3	P7	800
L2	T5	P2	100
L3	T3	P1	200
L3	T4	P2	500
L4	T6	P3	300
L4	T6	P7	300
L5	T1	P4	100
L5	T2	P2	200
L5	T2	P4	100
L5	T3	P4	200
L5	T4	P4	800
L5	T5	P4	400
L5	T5	P5	500
L5	T5	P7	100
L5	T6	P2	200
L5	T6	P4	500