

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

Sommersemester 2025

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	09.09.2025
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar) Aufgabenblock A: beigelegte Formelsammlung Aufgabenblock D: Anhang mit Anwendungsbeispiel.

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 11 (elf) Seiten.

Aufgabenblock A:

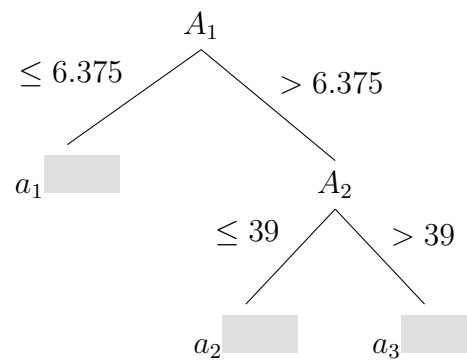
Knowledge-based systems

Aufgabe A.1: Entscheidungsbäume

[14 Punkte]

Gegeben ist der folgende Datensatz mit den drei Attributen A_1 , A_2 und A_3 sowie der Klassenmenge $C = \{X, Y, Z\}$. Die Attribute A_1 und A_2 sind kontinuierlich, das Attribut A_3 ist binär. Der noch nicht vollständige Entscheidungsbaum wurde mit dem CART-Verfahren erstellt.

i	A_1	A_2	A_3	C
1	5	25	1	X
2	3	38	0	X
3	8	34	0	Z
4	9	32	1	Z
5	2	33	0	X
6	9	49	0	X
7	6	35	1	Y
8	9	41	0	Y



- Klassifizieren Sie den Blattknoten in a_2 und geben Sie die Indizes der Daten an, die für den nächsten Test in Knoten a_3 relevant sind. (3 Punkte)
- Begründen Sie, warum im Knoten a_3 kein *Gini impurity*-Maß für die unterschiedlichen Attribute zu bestimmen ist. (2 Punkte)
- Vervollständigen Sie den Teil des Entscheidungsbaums, der in a_1 beginnt. Bestimmen Sie in jedem Schritt mittels des *Gini impurity*-Maßes den Test, der zur Verzweigung ab dem Knoten a_1 am besten geeignet ist. Testen Sie dabei, falls nötig, 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 $A_j > M_j$. (9 Punkte)

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

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

- a) Wir betrachten einen Datensatz (data set) mit m Elementen und 8 möglichen Klassifizierungen. Auf Basis der gegebenen Wahrscheinlichkeiten der Klassifizierungen ergibt sich eine Entropie von 3. Damit lässt sich feststellen, dass wir, ohne weitere Informationen zu haben, für ein zufällig gezogenes Element aus dem Datensatz jede mögliche Klassifizierung mit einer Wahrscheinlichkeit von $1/8$ ziehen können, d.h., keine Klassifizierung besitzt eine höhere Wahrscheinlichkeit als eine andere. (5 Punkte)
- b) Zwei endliche Datensätze sind genau dann linear separierbar (*linearly separable*) wenn sie auch absolut linear separierbar (*absolutely linearly separable*) sind. Für endliche Datensätze bedeutet dies, dass lineare Separierbarkeit aus absoluter linearer Separierbarkeit folgt und umgekehrt. (5 Punkte)
- c) Wir betrachten im Ada Boost M2 Verfahren die Berechnung des Pseudo-Loss ϵ_t in der Iteration t mit C möglichen Klassifikationen. Wenn dieser Wert größer als 0.5 ist muss das Verfahren in der Iteration t abgebrochen werden und das Ergebnis der vorherigen Iteration ($t - 1$) wird ausgegeben. Gilt $t = 1$, gibt es keine Ausgabe. (5 Punkte)

Aufgabe A.3: Bayes-Netzwerke**[10 Punkte]**

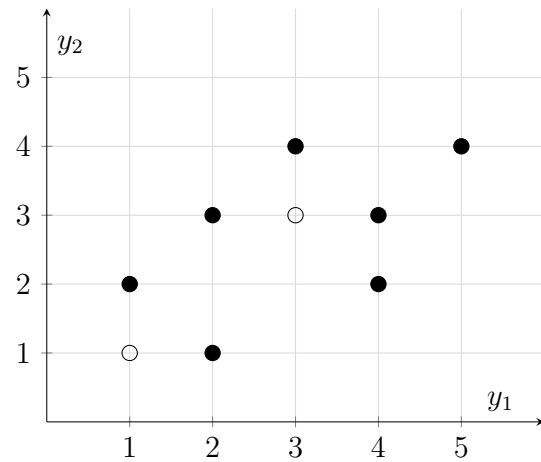
Wir betrachten fünf binäre Zufallsvariablen A, B, C, D und E in einem Bayes-Netzwerk. Die Tabellen bedingter Wahrscheinlichkeiten (CPTs) des Netzwerks werden durch die folgenden Wahrscheinlichkeiten bestimmt: $P(A) = 0,6$; $P(B) = 0,7$; $P(C|A) = 0,8$; $P(C|\neg A) = 0,3$; $P(D|B) = 0,6$; $P(D|A \wedge B) = 0,2$; $P(D|\neg A \wedge B) = 0,4$ und $P(E|C \wedge D) = 0,3$.

- a) Zeichnen Sie das Netzwerk inklusive der CPTs. Berechnen Sie, wenn möglich, weitere Wahrscheinlichkeiten in den einzelnen CPTs. Geben Sie auch an, welche Wahrscheinlichkeiten aus den gegebenen Daten nicht ermittelt werden können und deshalb noch angegeben werden müssten, um vollständige Tabellen zu erhalten. (6 Punkte)
- b) Berechnen Sie die Wahrscheinlichkeit $P(A|C)$. (4 Punkte)

Aufgabe A.4: Clustering**[6 Punkte]**

Im folgenden werden die Datenpunkte x_i mit Koordinaten y_1, y_2 betrachtet.

i	y_1	y_2
1	1	2
2	2	1
3	2	3
4	3	4
5	4	2
6	4	3
7	5	4



Verwenden Sie in allen Teilaufgaben als Distanzmaß die *Minkowski-Distanz* mit $\alpha \rightarrow \infty$.

- Bestimmen Sie die aktuellen Cluster C_1 und C_2 für die Zentren $c_1 = (1, 1)$ und $c_2 = (3, 3)$ wie im k -means Verfahren. (3 Punkte)
- 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)

Formelsammlung zum Aufgabenblock A

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

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

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

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

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

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

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

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

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

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

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

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

$$N_\varepsilon(p) := \{q \in D \mid \text{dist}(p, q) < \varepsilon\}$$

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? (2+2 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. (5+5 Punkte)

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, was aus logischer Sicht?
(4 Punkte)
- b) Erläutern Sie die Aufgaben und Eigenschaften der Schichten des Internet-Referenzmodells.
(4 Punkte)
- c) Wie ist eine Internet-Adresse (IPv4) aufgebaut?
Was ist eine Netzwerkmaske und wofür wird sie eingesetzt?
Welche Vor- und Nachteile hat das (klassenlose) CIDR-Schema im Gegensatz zum klassenbezogenen Adressierschema?
(4 Punkte)
- d) Nennen Sie die drei Hauptziele der Sicherheit im Internet und erläutern Sie diese in jeweils einem Satz.
(3 Punkte)

Aufgabe C.2: World Wide Web

[15 Punkte]

- a) Erläutern Sie kurz die Funktion der drei Kernstandards HTTP, HTML und URL sowie der Standarderweiterungen HTTPS, CSS, JavaScript und DOM der WWW-Architektur.
(7 Punkte)
- b) Stellen Sie in einer Skizze die HTTP-Interaktion zwischen einem Webclient (WWW Browser) und einem Webserver dar, wobei eine angeforderte DHTML-Ressource dynamisch per PHP-Script generiert werden soll und clientseitig JavaScript zum Einsatz kommt. Die Skizze soll dabei insbesondere die für die Erzeugung der dynamisch generierten Inhalte genutzten Funktionseinheiten enthalten. Beschreiben Sie den gesamten Vorgang mit allen notwendigen Arbeitsschritten von der Ressourcenanforderung bis zum Abschluss des Seitenaufbaus (Rendering).
(8 Punkte)

Aufgabe C.3: E-Mail

[15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren (Store-And-Forward) für die Übermittlung einer E-Mail vom E-Mail-Client (Mail User Agent) des Absenders zum E-Mail-Client des Empfängers. (8 Punkte)
- b) Welche Ziele verfolgt Phishing?
Erläutern Sie eine der Methoden, die für Phishing-Angriffe genutzt werden. (3 Punkte)
- c) Skizzieren Sie auf Anwendungsebene (Layer 4) die Nutzung eines Web-Mail-Dienstes per Webbrowser, wobei der Web-Mailer sowohl für den Abruf empfangener E-Mail von einem Posteingangsserver als auch für das Versenden von E-Mail über einen Postausgangsserver genutzt wird. Geben Sie in der Skizze die jeweils genutzten Anwendungsprotokolle zwischen den beteiligten Anwendungen und Diensten an. (4 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Sollten Sie alle vier Aufgaben bearbeiten, werden nur Ihre Lösungen zu den ersten drei Aufgaben des Aufgabenblattes bewertet.

Jede der Aufgaben kann entweder in deutsch oder englisch beantwortet werden. Im Anhang finden Sie das in der Vorlesung eingeführte Anwendungsbeispiel.

Aufgabe D.1: Architektur Relationaler Datenbanksysteme [15 Punkte]

- a) Erläutern Sie detailliert die 3-Ebenen-Architektur Relationaler Datenbanksysteme. (9 Punkte)
- b) Diskutieren Sie im Kontext der 3-Ebenen-Architektur das Konzept der Datenunabhängigkeit. (6 Punkte)

Aufgabe D.2: Relationenmodell [15 Punkte]

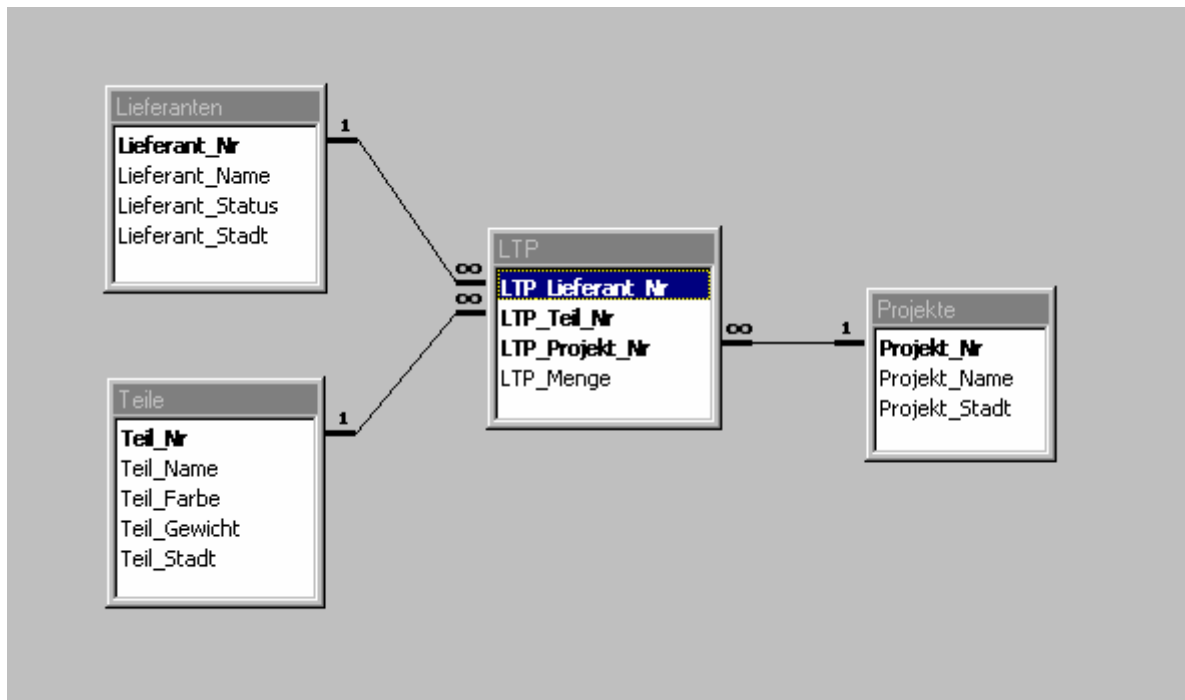
- a) Erläutern Sie das Geschlossenheitsprinzip der Relationalen Algebra. (5 Punkte)
- b) Die Operatoren der Relationalen Algebra INTERSECT (Bildung einer Schnittmenge), UNION (Bildung einer Vereinigungsmenge) und MINUS (Bildung einer Differenzmenge) sind dyadische Operatoren, die jeweils 2 Relationen als Eingabe erhalten. Welche Voraussetzungen müssen die Eingaberelationen erfüllen?
Erläutern Sie die 3 Operatoren anhand eines selbst gewählten Beispiels. (10 Punkte)

Aufgabe D.3: Sicherheit in Datenbanken [15 Punkte]

- a) Erläutern Sie den Prozess der Authentifizierung im Kontext Relationaler Datenbanksysteme. Verwenden Sie bei Ihrer Darstellung die jeweiligen Perspektiven und Aufgaben der Rollen Administrator und Datenbankbenutzer. (10 Punkte)
- b) Was bedeutet der Begriff der Authorisierung und wie wird dieses Konzept in Relationalen Datenbanksystemen umgesetzt? (5 Punkte)

Aufgabe D.4: Interne Ebene [15 Punkte]

- a) Beschreiben Sie, wie grundsätzlich Datensätze in Relationen Datenbanksysteme gespeichert werden. (5 Punkte)
- b) Definieren Sie den Begriff des Index. (4 Punkte)
- c) Was bedeuten die Begriffe des Dense Index und des Non-dense Index. (6 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
P 1	Sortierer	Paris
P 2	Bildschirm	Rom
P 3	OCR	Athen
P 4	Konsole	Athen
P 5	RAID	London
P 6	EDS	Oslo
P 7	Bandlaufwerk	London

LTP – Lieferungen

LTP_Lieferant_Nr	LTP_Teil_Nr	LTP_Projekt_Nr	LTP_Menge
L1	T1	P 1	200
L1	T1	P 4	700
L2	T3	P 1	400
L2	T3	P 2	200
L2	T3	P 3	200
L2	T3	P 4	500
L2	T3	P 5	600
L2	T3	P 6	400
L2	T3	P 7	800
L2	T5	P 2	100
L3	T3	P 1	200
L3	T4	P 2	500
L4	T6	P 3	300
L4	T6	P 7	300
L5	T1	P 4	100
L5	T2	P 2	200
L5	T2	P 4	100
L5	T3	P 4	200
L5	T4	P 4	800
L5	T5	P 4	400
L5	T5	P 5	500
L5	T5	P 7	100
L5	T6	P 2	200
L5	T6	P 4	500