

Name: \_\_\_\_\_ Vorname: \_\_\_\_\_ Matrikel-Nr.: \_\_\_\_\_

# Bergische Universität Wuppertal

## Schumpeter School of Business and Economics

**Bachelor of Science / Bachelor of Arts**

**Wintersemester 2025/2026**

---

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	27.02.2026
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 10 (zehn) Seiten.

Unterschrift: \_\_\_\_\_

# Aufgabenblock A:

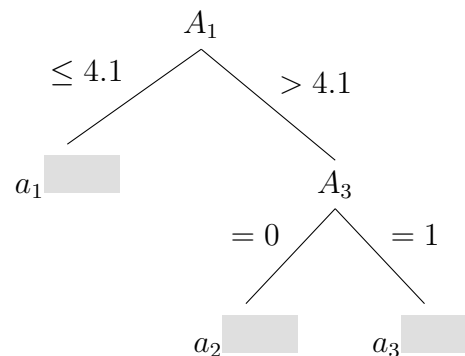
## Knowledge-based systems

### Aufgabe A.1: Entscheidungsbäume

[15 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	2	12	1	Y
2	2	13	0	Y
3	6	19	1	Z
4	4	13	0	X
5	7	14	1	Z
6	2	10	1	X
7	6	17	0	X
8	3	10	1	X
9	5	19	1	Z
10	4	16	1	Y



- Klassifizieren Sie die Blattknoten in  $a_2$  und  $a_3$ . (2 Punkte)
- Welche Eigenschaften muss ein Datensatz aufweisen, damit es möglich ist, einen (in der Zahl der Knoten gemessen) minimalen Entscheidungsbaum mit mehr als einem Knoten zum vollständigen und eindeutigen Klassifizieren dieses Datensatzes zu erstellen? (3 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 in Knoten  $a_1$  am geeignetsten 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  $A_j > M_j$ . (10 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.

- Wir betrachten einen Datensatz  $M$  mit  $m = |M|$  Elementen/Fällen,  $k$  möglichen Klassifizierungen und  $n$  Attributen. Dann ist die Entropie (Entropy) für diesen Datensatz genau dann größer Null wenn das Gini impurity-Maß ebenfalls größer Null ist. (5 Punkte)

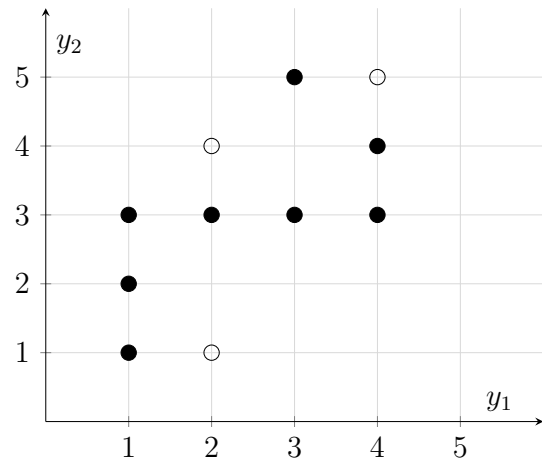
- b) Wir betrachten ein Regelbasiertes System (Rule-based system). Dann ist dieses System kommutativ unabhängig davon ob Negation zugelassen ist oder nicht. (5 Punkte)
- c) Die Minkowski Distanz (Minkowski distance)  $d^\alpha(x_1, x_2)$  für  $x_1, x_2 \in \mathbb{R}^d$  ist für alle Parameterwerte  $0 < \alpha < 1$  keine Metrik. (5 Punkte)

### Aufgabe A.3: Clustering und Separation

[15 Punkte]

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

i	$y_1$	$y_2$
1	1	1
2	1	2
3	1	3
4	2	3
5	3	3
6	3	5
7	4	3
8	4	4



Hinweis: Verwenden Sie in den Teilaufgaben a) und b) als Distanzmaß die *Manhattan-Distanz*.

- a) Bestimmen Sie die aktuellen Cluster  $C_1, C_2$  und  $C_3$  für die Zentren  $c_1 = (2, 1)$ ,  $c_2 = (2, 4)$  und  $c_3 = (4, 5)$  wie im  $k$ -means Verfahren. (4 Punkte)
- b) Bestimmen Sie die im Rahmen des  $k$ -means Verfahren aus den Clustern  $C_1, C_2$  und  $C_3$  resultierenden neuen Zentren  $c'_1, c'_2$  und  $c'_3$ . (4 Punkte)
- c) Zeigen Sie, dass die Punktmenge  $M_+ := \{x_1, x_2, x_3, x_4, x_6\}$  und  $M_- := \{x_5, x_7, x_8\}$  linear separabel sind, indem Sie passende Gewichtsparameter und den Wert  $\theta$  für ein lineares Perzeptron angeben (es ist keine Durchführung des Algorithmus notwendig). Überprüfen Sie die Korrektheit des ermittelten Perzeptrons anhand je eines Punktes aus  $M_+$  und  $M_-$ . (7 Punkte)

## Formelsammlung zum Aufgabenblock A

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

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

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

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

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

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

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

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

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

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

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

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

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

## Aufgabenblock B:

### Computer Hardware und Systembetrieb

#### (Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

##### Aufgabe B.1: Pipeline

[5 Punkte]

Erläutern Sie die Begriffe Stalling, Bubble und (Pipeline-)Flush.

##### Aufgabe B.2: Speculative Execution

[10 Punkte]

- a) Was versteht man unter Speculative Execution? (3 Punkte)
- b) Welche Aktionen sind erforderlich, wenn Befehle aufgrund falscher Sprungvorhersagen spekulativ fälschlich ausgeführt wurden? (3 Punkte)
- c) Welche Arten von Operationen sind dabei besonders problematisch? (2+2 Punkte)

##### Aufgabe B.3: 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.4: Copy-on-Write

[10 Punkte]

- a) Erläutern Sie kurz die Idee hinter Copy-on-Write. (6 Punkte)
- b) Wo wird dieses Konzept typischerweise eingesetzt? (2+2 Punkte)

##### Aufgabe B.5: HPC-Cluster/COTS-Produkte

[10 Punkte]

- a) Was versteht man unter COTS-Produkten? (4 Punkte)
- b) Wieso verwendet man heute oft Standardtechnologien anstelle spezieller Entwicklungen für den HPC-Bereich? (3 Punkte)
- c) Für welche Komponenten eines HPC-Clusters lassen sich COTS-Produkte nutzen? (3 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.  
(6 Punkte)
- b) Erläutern Sie kurz drei typische Anwendungen im WWW, die mit DHTML und nicht mit HTML realisiert werden können.  
(6 Punkte)
- c) Welche Aufgaben erfüllen die drei Standards SOAP, WSDL und UDDI beim Einsatz von Web-Services?  
(3 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 (drei) der folgenden 4 (vier) 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.

Als Ergänzung finden Sie im Anhang das in der Vorlesung eingeführte Anwendungsbeispiel.

#### Aufgabe D.1: Transaktionsmanagement [15 Punkte]

- a) Beschreiben Sie die grundlegenden Eigenschaften einer Transaktion. (6 Punkte)
- b) Erläutern Sie das Konzept des Write-Ahead-Log inklusive des Einsatzes von Checkpoints. (9 Punkte)

#### Aufgabe D.2: Relationenmodell [15 Punkte]

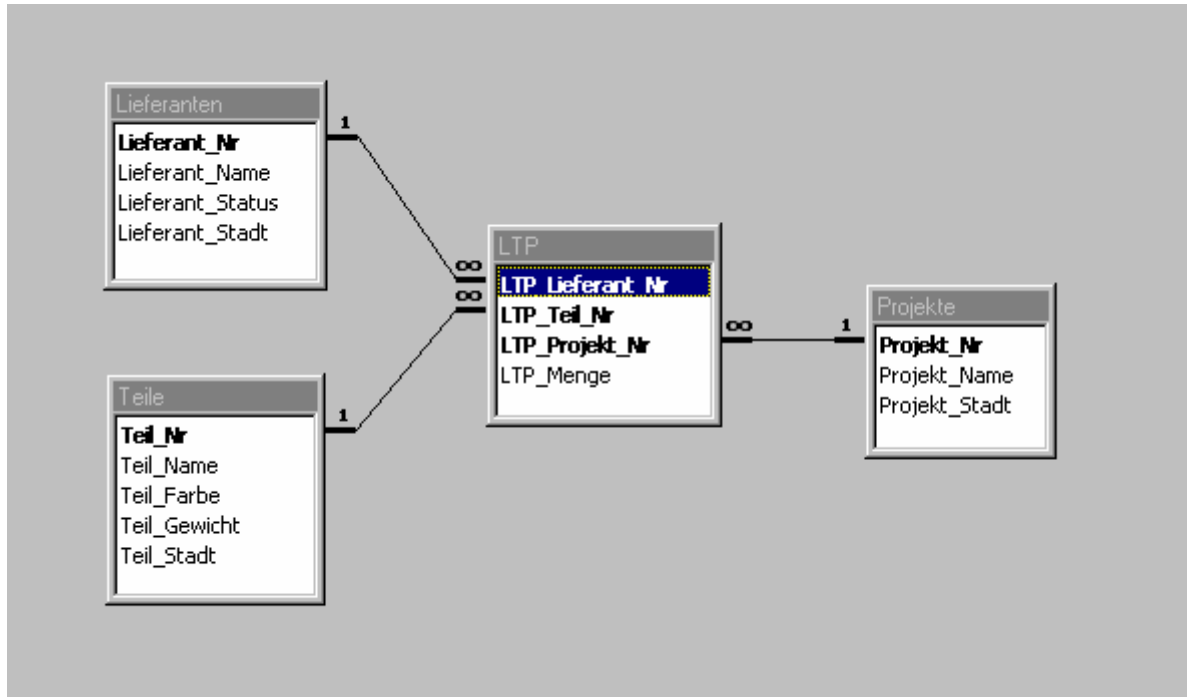
- a) Was besagt das *Geschlossenheitsprinzip* der relationalen Algebra? (3 Punkte)
- b) Welche Voraussetzung muß für die Mengenoperationen der Relationenalgebra *UNION*, *INTERSECT* und *DIFFERENCE/MINUS* gegeben sein? (3 Punkte)
- c) Erläutern Sie anhand eines selbst gewählten Beispiels die drei vorgenannten Mengenoperationen der Relationenalgebra. (9 Punkte)

#### Aufgabe D.3: Vektordaten [15 Punkte]

- a) Welches grundlegende Problem soll mit Hilfe von Vektordaten gelöst werden? (3 Punkte)
- b) Was haben *Chunking* und *Text embeddings* mit Vektordaten zu tun? (6 Punkte)
- c) Benennen Sie zwei Ihnen bekannte Indexverfahren für Vektordaten und erklären Sie kurz deren Funktionsweise. (6 Punkte)

#### Aufgabe D.4: Verschlüsselung [15 Punkte]

- a) Erläutern Sie die drei Verschlüsselungsvarianten für *Data at Rest*, *Data in Transit* und *Transparent Data Encryption*. (9 Punkte)
- b) Erläutern Sie im Kontext von Relationalen Datenbanksystemen die Begriffe *Authentifizierung* und *Authorisierung*. (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
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

---