

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

# Bergische Universität Wuppertal

## Schumpeter School of Business and Economics

**Bachelor of Science / Bachelor of Arts**

**WiSe 2024/2025**

---

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	05.02.2025
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.

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

# Aufgabenblock A:

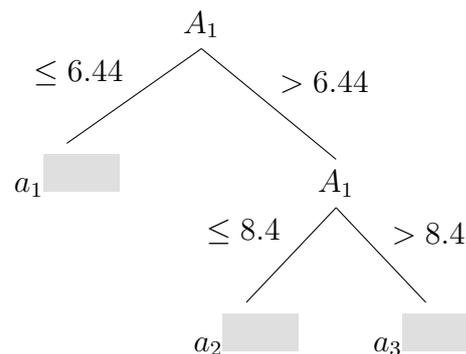
## Knowledge-based systems

### Aufgabe A.1: Entscheidungsbäume

[20 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	9	17	1	Y
2	5	14	1	X
3	9	11	1	Z
4	9	12	0	Y
5	7	19	0	Y
6	2	15	1	X
7	4	16	1	Y
8	8	11	1	Y
9	5	13	0	Z



- 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. (2 Punkte)
- Bestimmen Sie mittels des *Gini impurity*-Maßes den Test, der zur Verzweigung in Knoten  $a_1$  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, identifizieren Sie die dabei möglicherweise entstehenden Blattknoten und klassifizieren Sie diese. (9 Punkte)
- Begründen Sie, warum es unter den Bedingungen aus Teil b) nicht möglich ist, die Fälle in Knoten  $a_3$  mit der nächsten Verzweigung eindeutig zu klassifizieren. Wie könnte dagegen bei Aufhebung der Bedingungen aus Teil b) eine eindeutige Klassifizierung durch eine einzige Verzweigung im Knoten  $a_3$  gelingen? (5 Punkte)
- Betrachten Sie einen allgemeinen konsistenten Datensatz mit drei beliebigen Attributen und drei Klassen. Wie viele Einträge muss der Datensatz mindestens aufweisen, damit es einen Fall geben kann, bei dem der Weg zu jedem Blatt des dazugehörigen binären Entscheidungsbaumes mindestens die Länge 3 hat? Begründen Sie Ihre Antwort. (4 Punkte)

**Aufgabe A.2: Thesen****[6 Punkte]**

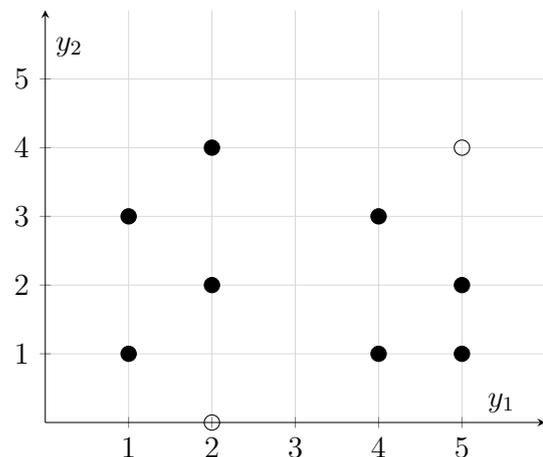
Nehmen Sie zu den folgenden Thesen begründet Stellung. Eine auf „richtig“ oder „falsch“ beschränkte Antwort erhält keine Punkte.

- a) Wir betrachten einen Datensatz und fragen uns, ob dieser linear separierbar ist. Um dies in endlicher Zeit zu entscheiden, können wir den aus der Vorlesung bekannten Algorithmus verwenden, der die Parameter des Perzeptrons bestimmt. Eine einmalige Ausführung dieses Verfahrens ist hierfür ausreichend. (3 Punkte)
- b) Wenn wir beim Density-Based Spatial Clustering zwischen zwei Ausführungen des Verfahrens für denselben Datensatz lediglich den Parameter  $\epsilon$  erhöhen, kann sich die Zahl der ermittelten Cluster im zweiten Durchlauf nicht gegenüber dem ersten Durchlauf erhöhen. (3 Punkte)

**Aufgabe A.3: Clustering, Separation und nächste Nachbarn****[19 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	3
3	2	2
4	2	4
5	4	1
6	4	3
7	5	1
8	5	2



Hinweis: Verwenden Sie in allen Teilaufgaben als Distanzmaß die *Minkowski-Distanz* mit  $\alpha = 1$ .

- a) Bestimmen Sie die aktuellen Cluster  $C_1$  und  $C_2$  für die in der Skizze eingezeichneten Zentren  $c_1 = (2, 0)$  und  $c_2 = (5, 4)$  wie im  $k$ -means Verfahren. (4 Punkte)
- b) Bestimmen Sie die im Rahmen des  $k$ -means Verfahrens aus den Clustern  $C_1$  und  $C_2$  resultierenden neuen Zentren  $c'_1$  und  $c'_2$ . (4 Punkte)
- c) Klassifizieren Sie den Punkt  $x_9 = (2, 1)$  mit Hilfe des einfachen *3-nearest-neighbor*-Verfahrens. Die Punkte  $x_1, x_2, x_3$  und  $x_4$  sind dabei der Klasse A und die Punkte  $x_5, x_6, x_7$  und  $x_8$  der Klasse B zuzuordnen. (3 Punkte)
- d) Bestimmen Sie die  $\epsilon$ -Umgebungen der Punkte  $x_2$  und  $x_8$  zum Wert  $\epsilon = 3$ . Zeigen Sie, dass die Punkt Mengen  $M_- := N_\epsilon(x_2)$  und  $M_+ := N_\epsilon(x_8)$  linear separierbar sind, indem Sie passende Gewichtsparameter und den Wert  $\theta$  für ein lineares Perzeptron angeben (es ist keine Durchführung eines Algorithmus notwendig!). Überprüfen Sie die Korrektheit des Perzeptrons nochmals anhand der Punkte  $x_2$  und  $x_8$ . (8 Punkte)

## Formelsammlung zum Aufgabenblock A

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

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

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

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

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

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

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

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

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

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

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

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

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

## **Aufgabenblock B:**

### **Computer Hardware und Systembetrieb**

#### **(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)**

##### **Aufgabe B.1: Pipeline**

**[5 Punkte]**

Beschreiben Sie das Prinzip einer Befehls-Pipeline.

##### **Aufgabe B.2: Von Neumann Konzept**

**[10 Punkte]**

Erläutern Sie kurz mindestens fünf (verschiedene) charakteristische Merkmale eines von Neumann Rechners.

##### **Aufgabe B.3: CISC vs. RISC**

**[10 Punkte]**

- a) Nennen Sie die technologischen Konzepte (Design-Prinzipien), die den zwei Architekturen RISC und CISC zugrunde liegen. (3+3 Punkte)
- b) Beschreiben Sie kurz die Vorteile und Nachteile der Architekturen. (4 Punkte)

##### **Aufgabe B.4: Cache-Kohärenz**

**[10 Punkte]**

- a) Beschreiben Sie das Verfahren, das bei einem Snoopy-Cache für die Sicherstellung der Konsistenz gleicher Cachezeilen in mehreren Caches Verwendung findet. (6 Punkte)
- b) Worin unterscheiden sich die Varianten write through und write back? (2+2 Punkte)

##### **Aufgabe B.5: HPC-Cluster**

**[10 Punkte]**

- a) Welche der Komponenten eines HPC-Clusters lassen sich bildlich mit Gehirn, Nerven, Herz und Seele vergleichen? (4 Punkte)
- b) Warum ist ein Netzwerk einzelner Rechner zunächst „seelenlos“? (3 Punkte)
- c) Welche „Zutat“ ist zusätzlich erforderlich, um aus einem Netzwerk einzelner Rechner einen HPC-Cluster zu bilden? (3 Punkte)

# Aufgabenblock C:

## Kommunikationssysteme (Internet-Technologien)

### Aufgabe C.1: Internet

[15 Punkte]

- a) Skizzieren Sie das Internet-Referenzmodell und erläutern Sie die Aufgaben und Eigenschaften der einzelnen Schichten. (4 Punkte)
- b) 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)
- c) Nennen Sie die wichtigsten Einträge im Header eines IP-Datagramms (IP-Header) und erläutern Sie jeweils deren Zweck. (4 Punkte)
- d) Nennen Sie die wichtigsten Einträge im Header eines TCP-Segments (TCP-Header) und erläutern Sie jeweils deren Zweck. (3 Punkte)

### Aufgabe C.2: Domain Name System

[15 Punkte]

- a) Welche Aufgabe hat das Domain Name System?  
Was versteht man unter einem FQDN?  
Welche Vorteile bieten symbolische Hostnamen im Vergleich zu IP-Adressen? (4 Punkte)
- b) Skizzieren Sie den Aufbau des Domain Name Space und zeichnen Sie folgende Host-Namen bzw. Repräsentationen ein:
- webmail.uni-wuppertal.de
  - www.linux.org
  - www.linux.com
  - 135.195.120.in-addr.arpa
  - 107.studs.math.uni-wuppertal.de
  - www.wiwi.uni-wuppertal.de
  - www.uni-wuppertal.de
- (7 Punkte)
- c) Welche gefälschten Informationen können einem Host bei den Auflösungsverfahren für Adressen (ARP) und Domännennamen (DNS) jeweils untergeschoben werden?  
In welchem netzwerktopographischen Bereich sind diese Fälschungen jeweils wirksam? (4 Punkte)

### **Aufgabe C.3: World Wide Web**

**[15 Punkte]**

- a) Erläutern Sie kurz die Funktion der drei Kernstandards HTTP, HTML und URL sowie der Standarderweiterungen HTTPS, CSS und JavaScript 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)

## 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: Transaktionsmanagement [15 Punkte]

- a) Beschreiben Sie drei typische Probleme, die beim Zugriff vieler Benutzer auf die Daten eines Datenbanksystems auftreten können. (6 Punkte)
- b) Erläutern Sie, wie mit Hilfe eines Sperrmechanismus die vorgenannten Probleme vermieden werden können. (6 Punkte)
- c) Welcher problematische Sperrzustand kann bei der Verwendung eines Sperrmechanismus auftreten und wie kann ein Datenbanksystem diesen Zustand auflösen? (3 Punkte)

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

- a) Benennen und definieren Sie die strukturellen Elemente des Relationenmodells in jeweils einem Satz. (6 Punkte)
- b) Definieren Sie ein Klassifikationsschema für *Status-Integritätsbedingungen* im Relationenmodell auf der Grundlage seiner strukturellen Elemente. Geben Sie für jede Klasse jeweils ein Beispiel. (9 Punkte)

#### Aufgabe D.3: Geodaten-Management [15 Punkte]

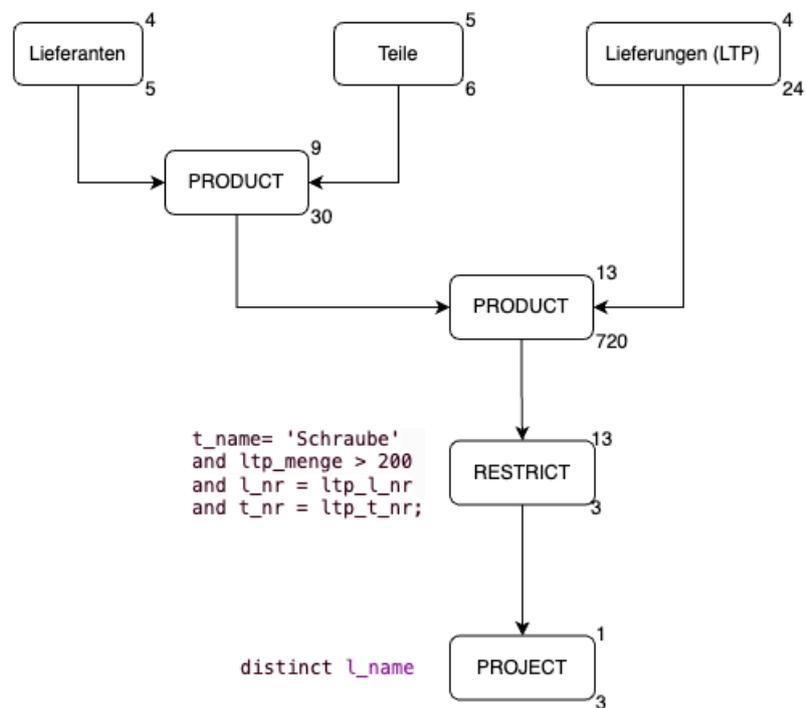
- a) Definieren Sie den Begriff des Geografischen Informationssystems (GIS). (4 Punkte)
- b) Welche Rolle spielt ein Geodatenbanksystem im Kontext eines GIS? (3 Punkte)
- c) Benennen und definieren Sie zwei Arten von Geodaten. Geben Sie für jede Datenart zwei Beispiele aus der unternehmerischen Praxis an. (8 Punkte)

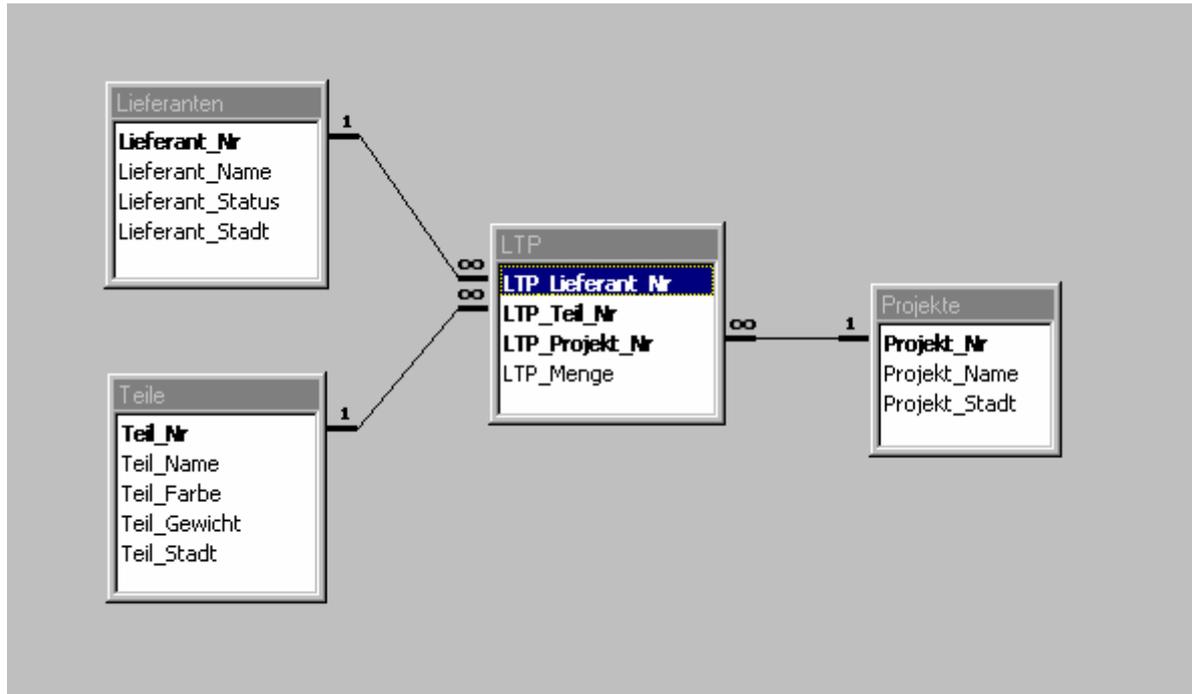
### Aufgabe D.4: Anfrageverarbeitung

[15 Punkte]

- a) In der folgenden Abbildung sehen Sie eine *SQL SELECT*-Anweisung sowie den vom Datenbanksystem generierten initialen logischen Anfrageplan (iLAP). Erstellen Sie einen verbesserten (präferierten) logischen Anfrageplan (pLAP). Benennen Sie zudem die von Ihnen angewendeten Optimierungsmethoden und/oder Steuerungsheuristiken, die Sie für die Erstellung des verbesserten LAP verwendet haben. (12 Punkte)
- b) Benennen Sie jeweils in einem Satz zwei Vorteile, die die Nutzung von im Systemkatalog vorhandenen Integritätsbedingungen im Rahmen der Semantischen Anfrageoptimierung bringen kann. (3 Punkte)

```
select distinct l_name
from lieferanten, teile, ltp
where t_name= 'Schraube'
and ltp_menge > 200
and l_nr = ltp_l_nr
and t_nr = ltp_t_nr;
```





### 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

---