

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

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

SoSe 2022

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	15.09.2022
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar) Aufgabenblock A: beigelegte 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

[11 Punkte]

Ein Weinfreund pflegt eine Weintabelle mit den vier Attributen **F**arbe, **J**ahrgang, **A**lkoholgehalt und Anzahl der **S**terne in einem Weinmagazin sowie seiner Klassifikation in die **G**eschmacksklassen **E**klig, **N**eutral und **L**ecker.

i	F	J	A	S	G
1	Rot	2016	14	1	E
2	Rot	2015	16	2	E
3	Weiß	2016	12	3	N
4	Rot	2017	15	4	L
5	Weiß	2005	13	3	N
6	Rot	2007	14	2	E
7	Weiß	2013	11	1	L
8	Rot	2015	12	4	N
9	Weiß	2010	12	2	E

- a) Bestimmen Sie das *Gini impurity*-Maß für die Verzweigung über das Attribut J im Wurzelknoten eines Entscheidungsbaums, wenn das binäre Kriterium $J \leq 2010 \vee J > 2010$ Anwendung finden soll. (4 Punkte)
- b) Bestimmen Sie die resultierende Reduktion oder Erhöhung des *Gini impurity*-Maßes im Vergleich zu Teilaufgabe a) bei einer weiteren Verzweigung aller jungen Weine ($J > 2010$) über das Attribut Farbe. (4 Punkte)
- c) Kann für den Teildatensatz, der alle Weißweine umfasst, eine vollständige Klassifikation über ein einzelnes (ggf. nichtbinäres) Attribut erreicht werden? Geben Sie den entsprechenden Entscheidungsbaum an oder begründen Sie Ihre Entscheidung. (3 Punkte)

Aufgabe A.2: Bayes-Netzwerke

[9 Punkte]

Wir betrachten vier binäre Zufallsvariablen A, B, C und D in einem Bayes-Netzwerk. Die Tabellen bedingter Wahrscheinlichkeiten (CPTs) des Netzwerks enthalten unter anderem die Wahrscheinlichkeiten $P(A) = 0,5$, $P(B|A) = 0,3$, $P(C|\neg A) = 0,1$, $P(C|A) = 0,2$, $P(D|C) = 0,3$ und $P(D|\neg C) = 0,1$.

- a) Zeichnen Sie das Netzwerk inklusive der CPTs. Geben Sie auch an, welche weiteren Wahrscheinlichkeiten in den einzelnen CPTs noch angegeben werden müssten, um vollständige Tabellen zu erhalten. (4 Punkte)
- b) Berechnen Sie die Wahrscheinlichkeit $P(C|D)$. (5 Punkte)

Aufgabe A.3: Thesen**[10 Punkte]**

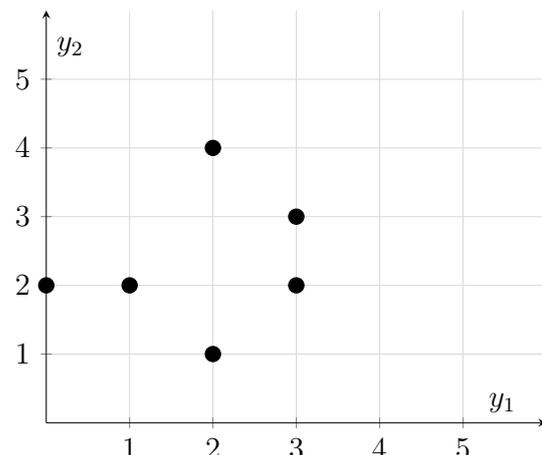
Nehmen Sie zu den folgenden Thesen begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort erhält keine Punkte.

- Ein in der Hauptschleife des Algorithmus DB-Scan untersuchter Randpunkt (border point) kann niemals ein neues Cluster definieren. (5 Punkte)
- Die Kombination von verschiedenen fehlerbehafteten schwachen Schätzern kann nicht zu einem erfolgreichen Lernverfahren führen. (5 Punkte)

Aufgabe A.4: Clustering, Separation und nächste Nachbarn**[15 Punkte]**

Wir betrachten die Datenpunkte x_1, \dots, x_6 .

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



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

- Bestimmen Sie die aktuellen Cluster C_1 und C_2 für die Zentren $c_1 = (2, 2)$ und $c_2 = (3, 4)$ wie im k -means Verfahren. (2 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)
- Zeigen Sie, dass die Punktmenge $M_+ := \{x_4, x_5\}$ und $M_- := \{x_1, x_2\}$ 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_- und klassifizieren Sie dann die Punkte x_3 und x_6 . (6 Punkte)
- 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

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

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

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

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

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

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

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

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

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

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

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

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

Aufgabenblock B:

Computer Hardware und Systembetrieb

(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

Aufgabe B.1: 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. (6 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? (4 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) Was versteht man unter dem Internet aus physischer Sicht und was aus logischer Sicht? (4 Punkte)
- b) Skizzieren Sie das Internet-Referenzmodell und erläutern Sie kurz die Aufgaben und Eigenschaften der einzelnen Schichten. (4 Punkte)
- c) Nennen Sie die wichtigsten Einträge im Header eines IP-Datagramms (IP-Header) und erläutern Sie jeweils deren Zweck. (4 Punkte)
- d) Stellen Sie in einer Skizze dar, in welcher Weise WWW-Anwendungsdaten beim Durchlaufen der Schichten des Internet-Protokollstapels bis hin zum Netzwerkrahmen (Netzzugangsschicht) gekapselt werden. (3 Punkte)

Aufgabe C.2: Domain Name System

[15 Punkte]

- a) Welche Vorteile bieten symbolische Host-Namen im Vergleich zu IP-Adressen? (2 Punkte)
- b) Welche Aufgabe hat das Domain Name System? Was versteht man unter *Reverse DNS Lookup* und unter einem *Fully Qualified Domain Name*? (3 Punkte)
- c) Skizzieren Sie den Aufbau des Domain Name Space und zeichnen Sie insbesondere folgende Host-Namen bzw. Repräsentationen ein:
 - `www.wiwi.uni-wuppertal.de`
 - `www.google.com`
 - `135.195.132.in-addr.arpa`
 - `1115.studs.math.uni-wuppertal.de`(6 Punkte)
- d) Beschreiben Sie zwei Angriffsszenarien, die das Ziel haben, Hosts gefälschte DNS-Informationen unterzuschieben. Erläutern Sie kurz, wie durch *Domain Name System Security Extensions* (DNSsec) Sicherheit hergestellt wird. (4 Punkte)

Bitte beachten Sie für Aufgabe C.3 die Folgeseite!

Aufgabe C.3: Webservices

[15 Punkte]

- a) Welche Aufgaben erfüllen die drei Standards SOAP, WSDL und UDDI beim Einsatz von Webservices? (6 Punkte)
- b) Erläutern Sie anhand einer Skizze den Mechanismus *Finden, Binden, Datenaustausch* aller beteiligten Komponenten bei der Anforderung einer Ressource durch einen Webbrowser, für deren Erstellung durch den Webserver ein Webservice genutzt wird. Ihre Skizze soll dabei insbesondere enthalten, wo die drei Standards zum Tragen kommen. (5 Punkte)
- c) Skizzieren Sie die jeweilig genutzten Internet-Protokollstapel
- bei der Interaktion zwischen Webclient (WWW-Browser) und Webserver,
 - bei der Interaktion zwischen dem Webserver und einem weiteren Webserver bei der Nutzung eines Webservice.

(4 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Aufgabe D.1: Relationenmodell [15 Punkte]

Erläutern Sie ein Klassifikationsschema für Integritätsbedingungen. Orientieren Sie sich dabei an den strukturellen Elementen des Relationenmodells ((Daten-) Typ, Tupel und Relation). Geben Sie für jeden der Klassifikationsbereiche ein selbst gewähltes Beispiel an.

Aufgabe D.2: Transaktionsmanagement [15 Punkte]

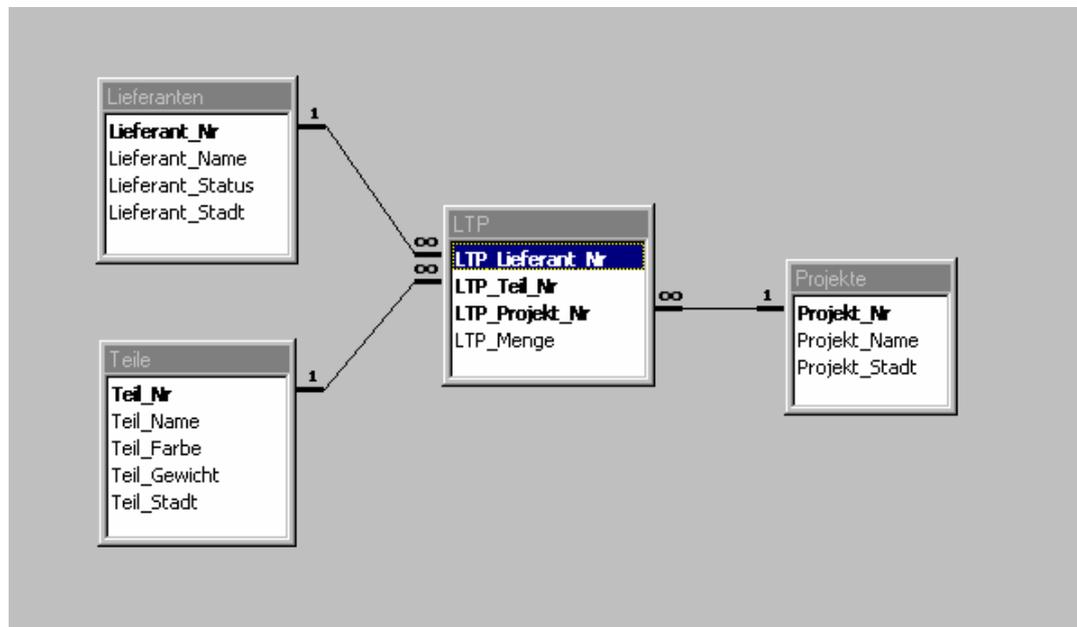
- Definieren Sie den Transaktionsbegriff, auch unter Verwendung der Eigenschaften einer Transaktion. (9 Punkte)
- Beschreiben Sie den Ablauf einer Transaktion. (6 Punkte)

Aufgabe D.3: Relationenalgebra [15 Punkte]

- Erläutern Sie das Geschlossenheitsprinzip der Relationenalgebra. (5 Punkte)
- Erläutern Sie die folgenden Operationen der Relationenalgebra: DIFFERENCE (Bildung einer Differenzmenge), INTERSECT (Bildung einer Differenzmenge) und UNION (Bildung einer Vereinigungsmenge).
Welche Besonderheit müssen die Ausgangsrelationen für die vorgenannten Operationen erfüllen? (10 Punkte)

Aufgabe D.4: Interne Ebene [15 Punkte]

- Erläutern Sie den Begriff des Index. Welche Vorteile und ggf. Nachteile sind mit der Nutzung eines Index verbunden? (6 Punkte)
- Erläutern Sie das Konzept des Mehrebenenindex (Multilevel-Index) am Beispiel des B^* -Baums. Stellen Sie dabei auch heraus, welche Teile des B^* -Baums einen *dense index* respektive einen *non-dense index* repräsentieren. (9 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

