

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

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

WS 2021/2022

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	08.02.2022
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 9 (neun) 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

- Geben Sie einen weiteren Wein mit seinen Attributsausprägungen und einer Klassifikation an, die den obigen Datensatz inkonsistent werden lässt (Dieser Wein soll in den folgenden Aufgaben keine Rolle mehr spielen). (2 Punkte)
- Bestimmen Sie das *Gini impurity*-Maß für die Verzweigung über das Attribut F im Wurzelknoten eines Entscheidungsbaums. Warum können Sie ohne weitere Zusatzinformationen dieses Maß berechnen? (4 Punkte)
- Bestimmen Sie die resultierende Reduktion oder Erhöhung des *Gini impurity*-Maßes im Vergleich zu Teilaufgabe b) bei einer weiteren Verzweigung aller Rotweine über das binäre Kriterium $S \leq 2,5 \vee S > 2,5$. (3 Punkte)
- Welche Eigenschaft hat der Teildatensatz, der alle Rotweine mit bis zu 2,5 Sternen betrachtet? Zeichnen Sie auch den dazugehörigen Entscheidungsbaum für diesen Teildatensatz. (2 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 $P(A) = 0,5$, $P(B|A) = 0,3$, $P(B|\neg A) = 0,1$, $P(C|A) = 0,2$ und $P(D|B, C) = 0,4$.

- 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. (5 Punkte)
- Berechnen Sie die Wahrscheinlichkeit $P(A|B)$. (4 Punkte)

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

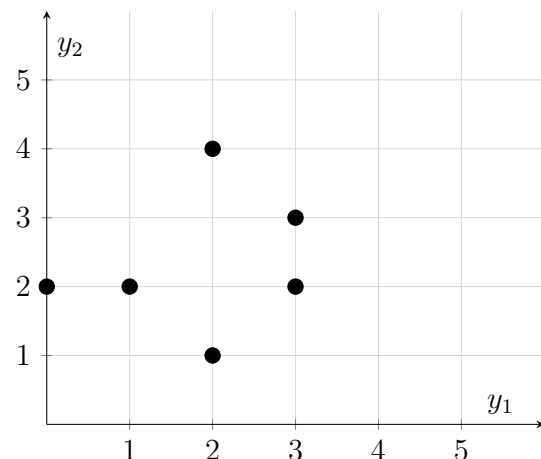
Nehmen Sie zu den folgenden Thesen begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort erhält keine Punkte.

- a) Im Algorithmus DB-Scan kann ein Punkt auch nachträglich (nachdem er explizit in der Hauptschleife untersucht wurde) noch einem Cluster zugeordnet werden. Allerdings gilt das nicht für Randpunkte (border points). (5 Punkte)
- b) Wir betrachten das Verfahren AdaBoost.M1 bei einem Trainingsdatensatz mit mehr als zwei Klassen. Die Anwendung eines „Weak Learners“, der auf reinem Raten (guessing) basiert, führt automatisch zum Abbruch des Algorithmus?. (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).

- a) Bestimmen Sie die aktuellen Cluster C_1 und C_2 für die Zentren $c_1 = (1, 3)$ und $c_2 = (3, 2)$ 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 Punktmenge $M_+ := \{x_5, x_6\}$ und $M_- := \{x_2, x_3\}$ 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_- und klassifizieren Sie dann die Punkte x_1 und x_4 . (6 Punkte)
- d) Klassifizieren Sie den Punkt $x_7 = (2, 2)$ mit Hilfe des einfachen 3-nearest-neighbor Verfahrens. Die Punkte x_1, x_2 und x_5 sind dabei der Klasse A und die Punkte x_3, x_4 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: Von Neumann Flaschenhals

[5 Punkte]

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

Aufgabe B.2: Instruction-Pipeline

[10 Punkte]

- a) Was ist eine Instruction-Pipeline? (4 Punkte)
- b) Welche Vor- und welche Nachteile sind mit dem Einsatz einer CPU mit Pipelintechologie verbunden? (6 Punkte)

Aufgabe B.3: CISC vs. RISC

[10 Punkte]

- a) Was ist unter der Bezeichnung *Minimierung der semantischen Lücke* (bridging the semantic gap) zu verstehen? (2 Punkte)
- b) Im Rahmen welcher Rechnerarchitektur spielt sie eine grundlegende Rolle und weshalb kam es zu diesem „Ziel“? (2 Punkte)
- c) Nennen Sie die technologischen Konzepte (Design-Prinzipien), die den zwei Architekturen RISC und CISC zugrunde liegen. (6 Punkte)

Aufgabe B.4: Prozesse vs. Fäden

[10 Punkte]

Worin unterscheiden sich Prozesse und Fäden/Threads?

Aufgabe B.5: Mikroprogrammierte Prozessoren

[10 Punkte]

- a) Beschreiben Sie das Konzept von mikroprogrammierten Prozessoren. (6 Punkte)
- b) Welche Vorteile bieten (ggf. mehrstufig) mikroprogrammierte CPUs? (4 Punkte)

Aufgabenblock C:

Kommunikationssysteme (Internet-Technologien)

Aufgabe C.1: Internet

[15 Punkte]

- a) Was versteht man unter dem Internet aus physischer 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: E-Mail

[15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren (Store-And-Forward) 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)

Aufgabe C.3: World Wide Web

[15 Punkte]

- a) Erläutern Sie kurz folgende fünf Standards und Technologien im WWW: XHTML, SOAP, RIA, CSS und DOM. (5 Punkte)
- b) Stellen Sie in einer Skizze die HTTP-Interaktion zwischen einem WWW-Browser und einem Webserver mit allen beteiligten Funktionseinheiten dar, wobei die angeforderte DHTML-Ressource dynamisch per Java-Servlet generiert wird. Beschreiben Sie kurz alle Arbeitsschritte von der Ressourcenanforderung bis zum Abschluss des Seitenaufbaus (Rendering). (6 Punkte)
- c) Erstellen Sie eine vergleichende Skizze der jeweils genutzten Protokollstapel im WWW
 - bei Nutzung regulärer unverschlüsselter HTTP-Kommunikation,
 - bei Nutzung eines Webservice auf SOAP-Basis,
 - bei HTTP-Kommunikation in einem Overlay-Network via SSL-VPN mit in HTTPS gekapselten IP-Paketen und
 - bei HTTP-Kommunikation in einem VPN auf IPsec-Basis mit Tunnelmodus.

(4 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Hinweis: Bei Bearbeitung aller vier Aufgaben werden nur die Aufgaben D.1, D.2 und D.3 gewertet.

Aufgabe D.1: Architektur **[15 Punkte]**

- a) Erläutern Sie die vier Kernkomponenten eines Relationalen Datenbanksystems: Systemkatalog, Optimierungskomponente, Transaktionsmanager und Run-time Manager. (10 Punkte)
- b) Gegeben sei die folgende Aussage: in Verteilten Relationalen Datenbanksystemen erfolgt die Verteilung der Daten auf der Konzeptuellen Ebene.
Ist diese Aussage korrekt? Begründen Sie Ihre Entscheidung! (5 Punkte)

Aufgabe D.2: Interne Ebene **[15 Punkte]**

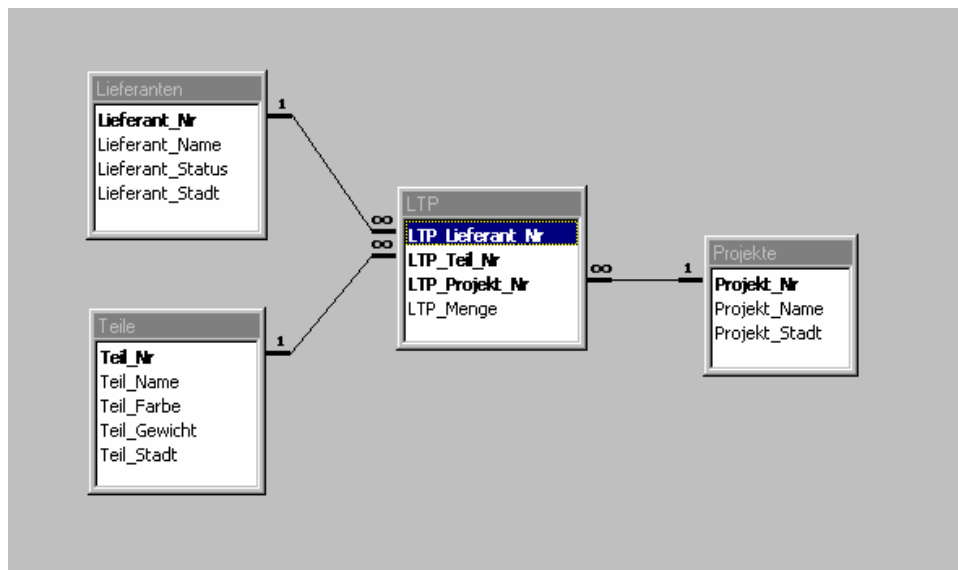
- a) Erläutern Sie die Begriffe File, Fileset, Page und Page Set anhand eines selbst gewählten Beispiels. (9 Punkte)
- b) Charakterisieren Sie kurz den B-Baum und erläutern Sie seinen Aufbau. (6 Punkte)

Aufgabe D.3: Sicherheit **[15 Punkte]**

- a) Erläutern Sie den Prozess der Authentifizierung und Authorisierung im Kontext eines Datenbanksystems. (10 Punkte)
- b) Erläutern Sie das Konzept „Mandatory Access Control“ (MAC) am Beispiel des Label Based Access Control (LBAC). Verwenden Sie bei Bedarf ein selbst gewähltes Beispiel. (5 Punkte)

Aufgabe D.4: Relationenmodell **[15 Punkte]**

- a) Erläutern Sie die strukturellen Elemente des Relationenmodells ((Daten-) Typ, Tupel und Relation) anhand eines selbst gewählten Beispiels. (10 Punkte)
- b) Was besagt das Konzept der „Tupeläquivalenz“? Wofür wird diese Konzept im Kontext des Relationenmodells benötigt? (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