

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

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

SS 2021

Prüfungsgebiet: BWiWi 2.8 Wissensbasierte Systeme und
Informationstechnologien
BWiGes 5.8 Operations Management und
Informationstechnologien
Grundlagen der Wirtschaftsinformatik Modul I
(PO Neufassung 2014)

Tag der Prüfung: 24.09.2021

Name des Prüfers: Prof. Dr. Bock

Erlaubte Hilfsmittel: Taschenrechner (nicht programmierbar),
Aufgabenblock A: beigegefügte Formelsammlung,

Bearbeiten Sie 2 der angegebenen 4 Aufgabenblöcke vollständig!

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. Insgesamt können **90 Punkte** erreicht werden. Für eine erfolgreiche Bearbeitung müssen wenigstens **45 Punkte** erworben werden.

Für Bachelor-Studiengänge ist der Aufgabenblock A verbindlich!

Bei Bearbeitung von mehr als zwei Aufgabenblöcken wird die Bearbeitung des Aufgabenblocks A und des ersten weiteren bearbeiteten Blocks gewertet.

Die Klausur besteht mit diesem Deckblatt aus insgesamt 7 (sieben) Seiten.

Unterschrift: _____

Aufgabenblock A:

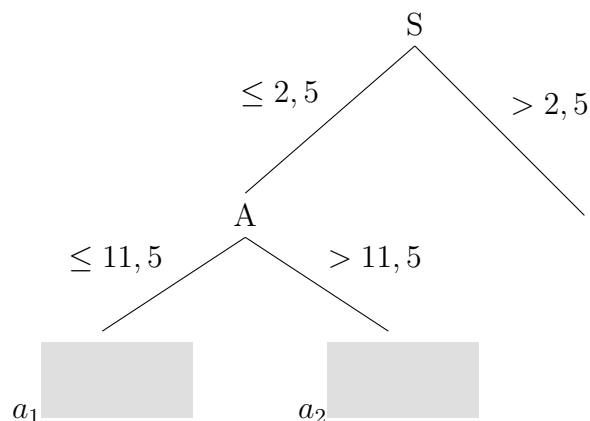
Knowledge-based systems

Aufgabe A.1: Entscheidungsbäume

[10 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. Zusätzlich hat er angefangen, mit der CART Methodik einen Entscheidungsbaum aufzubauen.

Datensatz					
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) Vervollständigen Sie die linke Seite des Entscheidungsbaums, indem Sie für die beiden Blattknoten a_1 und a_2 die resultierende Klassifikation in den grauen Kästchen notieren. (2 Punkte)
- b) Bestimmen Sie mittels des *Gini impurity*-Maßes den Test, der zur nächsten Verzweigung der rechten Baumhälfte am geeignetsten ist. Testen Sie einerseits das Attribut F und verwenden Sie andererseits für die Attribute J, A, und S jeweils als einzigen Test:
- $J \leq 2015,5 / J > 2015,5$
 - $S \leq 3,5 / S > 3,5$
 - $A \leq 13,5 / A > 13,5$

Erweitern Sie anschließend den obigen Entscheidungsbaum. Klassifiziert der um diese Verzweigung erweiterte Entscheidungsbaum den Datensatz vollständig und fehlerfrei?

(8 Punkte)

Aufgabe A.2: Thesen

[15 Punkte]

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

- a) In vielen Fällen kann die Anwendung des Gain-Ratio-Kriteriums anstelle des Gain-Kriteriums vorteilhaft sein. Dieser positive Effekt kann jedoch nicht erreicht werden, wenn ausschließlich binäre Tests vorliegen. (5 Punkte)
- b) Wir betrachten die Punkte p und q . Ist der Punkt q density-reachable von p aus, ist p immer auch density-reachable von q aus. (5 Punkte)
- c) Im Verfahren AdaBoost.M2 kann man für den in der Iteration t verwendeten besten schwachen Schätzer annehmen, dass $\epsilon_t \leq 1/2$ gilt. (5 Punkte)

Aufgabe A.3: Clustering und Separation

[20 Punkte]

Im Folgenden wird zunächst das k -means-Clusteringverfahren betrachtet.

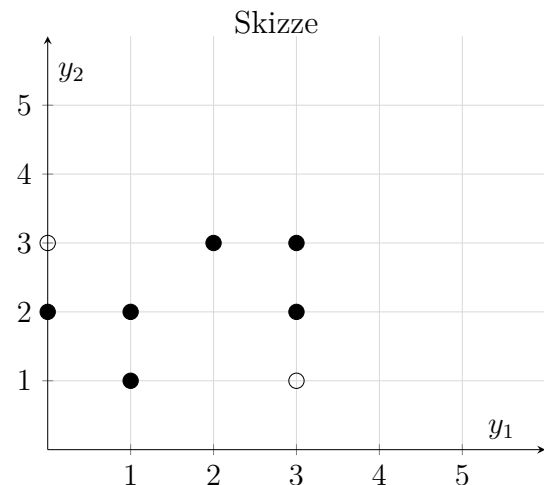
Gegeben sind Datenpunkte x_1, \dots, x_6 (skizziert als ●) sowie die beiden Zentren $C = \{c_1, c_2\}$ (skizziert als ○).

Datenpunkte x_i

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

Zentren c_i

i	y_1	y_2
1	0	3
2	3	1



Hinweis: Verwenden Sie zur Bearbeitung der Teilaufgaben das *Manhattan-Distanzmaß*.

- a) Bestimmen Sie die aktuellen Cluster C_1 und C_2 . (2 Punkte)
- b) Führen Sie eine weitere Iteration des k-means-Verfahrens durch und geben sie abschließend die neuen Cluster C_1 und C_2 sowie die neuen Zentren c_1 und c_2 an. Skizzieren Sie die beiden Zentren zusätzlich in obiger Skizze mit dem Symbol \otimes . (9 Punkte)
- c) Für das k-means++ Verfahren sei x_1 als erstes Zentrum ausgewählt worden. Welcher Datenpunkt hat die höchste Wahrscheinlichkeit, als zweites Zentrum ausgewählt zu werden? Geben Sie unter Verwendung dieses Punktes die resultierenden Cluster an. (4 Punkte)
- d) Zeigen Sie, dass die Punktmenge $M^+ := \{x_1, x_2, x_3\}$ und $M^- := \{x_4, x_5, x_6\}$ linear separabel sind, indem Sie passende Gewichtparameter und den Wert θ für ein lineares Perzeptron angeben (es ist keine Durchführung eines Algorithmus notwendig!). Zeigen Sie dann, dass Ihr Perzeptron die Punkte korrekt klassifiziert. (5 Punkte)

Formelsammlung zum Aufgabenblock A

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

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

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

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

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

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

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

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

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

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

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

Aufgabenblock B:

Computer Hardware und Systembetrieb

(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

Aufgabe B.1: Superskalare CPUs **[5 Punkte]**

- a) Was ist eine superskalare CPU? (3 Punkte)
- b) Mit welchen Architekturmerkmalen erreicht man Superskalarität? (2 Punkte)

Aufgabe B.2: 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.3: 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? (4 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.

Aufgabe B.5: Copy-on-Write **[10 Punkte]**

- a) Erläutern Sie kurz die Idee hinter Copy-on-Write. (6 Punkte)
- b) Wo wird dieses Konzept typischerweise eingesetzt? (4 Punkte)

Aufgabenblock C:

Kommunikationssysteme (Internet-Technologien)

Aufgabe C.1: Internet

[15 Punkte]

- a) Stellen Sie in einer Skizze dar, auf welchem Weg eine WWW-Ressourcenanfrage (HTTP-Request) von einem Webbrowser auf einem Internet-Host zu einem Webserver auf einem anderen Internet-Host durch die beiden Protokollstapel und das Internet übertragen wird. Geben Sie dabei insbesondere auch zu jeder Schicht außer der untersten (Netzwerk-Hardware) typische Protokolle für diese Internet-Nutzung an. (7 Punkte)
- b) Stellen Sie in einer Skizze dar, in welcher Weise WWW-Anwendungsdaten beim Durchlaufen der Schichten des Internet-Protokollstapels bis hin zum Netzwerkraum (Netzzugangsschicht) gekapselt werden. (3 Punkte)
- c) Erläutern Sie die Aufgaben und Eigenschaften der einzelnen Schichten des Internet-Referenzmodells. (5 Punkte)

Aufgabe C.2: Netzzugang und Routing

[15 Punkte]

- a) Welche Aufgabe hat ein DHCP-Server in einem Netzwerk? Welche Netzwerk-Konfigurationsdaten auf welcher Schicht des Internet-Protokollstapels eines Hosts sind unbedingt für die Kommunikation im Internet erforderlich? (4 Punkte)
- b) Erläutern Sie die im Internet für die Vermittlung der IP-Datagramme genutzten Methoden Speichervermittlung und Teilstreckenvermittlung. Was versteht man dabei unter Quellennunabhängigkeit? (3 Punkte)
- c) Welchem Zweck dient die IP-Adresshierarchie? Welche Informationen benötigt ein IP-Router für universelles und optimales Routing? (4 Punkte)
- d) Stellen Sie in einer Skizze dar, wie ein Datenpaket aus einem privaten Netz modifiziert wird, wenn es einen NAT-Router hin zum Internet passiert (Masquerading), und wie die Antwortpakete beim Eintritt in das private Netz modifiziert werden. (4 Punkte)

Aufgabe C.3: IPv4 vs. IPv6

[15 Punkte]

- a) Wie ist eine Internet-Adresse (IPv4) aufgebaut? (3 Punkte)
- b) Welche Vor- und Nachteile hat das CIDR-Schema im Gegensatz zum klassenbezogenen Adressierschema? (2 Punkte)
- c) Nennen und erläutern Sie mindestens fünf Vorteile von IPv6 gegenüber IPv4. (10 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Aufgabe D.1: Architektur [15 Punkte]

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

Aufgabe D.2: Interne Ebene [15 Punkte]

- a) Erläutern Sie den Begriff des Index. Welche Vorteile und ggf. Nachteile sind mit der Nutzung eines Index verbunden? (6 Punkte)
- b) Beschreiben Sie das Konzept von Zeigerketten (Pointer Chains) als Implementierungsmechanismus für einen Index. Nennen Sie einen typischen Verwendungszweck für die Verwendung von Zeigerketten sowie Vor- und Nachteile dieses Ansatzes. (9 Punkte)

Aufgabe D.3: Transaktionsmanagement [15 Punkte]

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

Aufgabe D.4: Anfrageoptimierung [15 Punkte]

- a) Beschreiben Sie das Komponentenmodell für die Anfrageverarbeitung eines Relationalen Datenbanksystems. (10 Punkte)
- b) Erläutern Sie das Konzept der Semantischen Anfrageoptimierung und nennen Sie mindestens vier (4) ihrer Voraussetzungen. (5 Punkte)