

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

Bergische Universität Wuppertal
Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

WS 2019/2020

Prüfungsgebiet:	BWiWi 2.8 / BWiGes 5.8 Wissensbasierte Systeme und Informationstechnologien Operations Management und Informationstechnologien Grundlagen der Wirtschaftsinformatik Modul I (PO Neufassung 2014)
Tag der Prüfung:	18.02.2020
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 Studierende der 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 11 (elf) Seiten.

Unterschrift: _____

Aufgabenblock A:

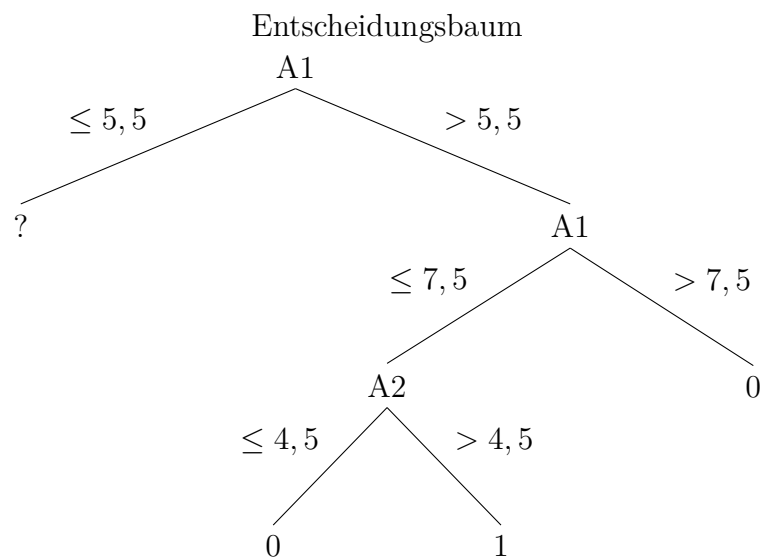
Knowledge-based systems

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

Aufgabe A.1: Entscheidungsbäume [15 Punkte]

Gegeben ist ein Datensatz mit nicht näher spezifizierten Attributen A1, A2 und A3 und Klassen C, sowie ein zugehöriger, unvollständiger Entscheidungsbaum. Dieser Entscheidungsbaum wurde mittels der CART-Methodik erzeugt.

Datensatz				
i	A1	A2	A3	C
1	6	4	1	0
2	3	1	0	0
3	1	8	3	2
4	7	5	2	1
5	2	0	9	2
6	4	2	7	2
7	9	9	6	0
8	8	6	4	0
9	0	7	8	1
10	5	3	5	2



- Betrachten Sie den als „?“ bezeichneten Knoten des unvollständigen Entscheidungsbaums in der linken Baumhälfte. Markieren Sie die Fälle des Datensatzes, die für die weitere Verzweigungen nach der CART-Methodik herangezogen werden (z.B. farblich oder durch Streichen der übrigen Fälle). (2 Punkte)
- Bestimmen Sie mittels des *Gini impurity*-Maßes den Test, der zur weiteren Verzweigung der in Teilaufgabe a) kenntlich gemachten Daten am geeignetsten ist. Testen Sie jedes Attribut (A1, A2, A3) nur einmal für die Verzweigung, indem Sie für den Schwellwert eines Tests jeweils den Mittelwert der relevanten Werte des jeweiligen Attributs verwenden.

Hinweis: Führen Sie nur eine Verzweigung durch, auch wenn weitere Verzweigungen möglich wären. (9 Punkte)

Auf der nächsten Seite geht es weiter ...

- c) Nehmen Sie zu der folgenden These begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort enthält keine Punkte.

In der Informationstheorie ist die *Entropie* H ein Maß für den mittleren Informationsgehalt eines Ereignisses eines Zufallsexperimentes unter einer gegebenen Verteilung. Die *Entropie* H nimmt also dann einen maximalen Wert an, wenn ein Ereignis als sicher gilt, d.h. die Eintrittswahrscheinlichkeit eines Ereignisses besitzt eine Wahrscheinlichkeit von 1. (4 Punkte)

Aufgabe A.2: Boosting mit AdaBoost**[15 Punkte]**

Bearbeiten Sie die folgenden Teilaufgaben.

- a) In der Veranstaltung haben sie die Varianten des AdaBoost-Verfahrens *AdaBoost* (für binäre Klassifikationen) und *AdaBoost.M1* kennen gelernt. Beschreiben sie kurz die Grundideen der Verfahren bezüglich der Aktualisierung der gegebenen Verteilung (Anpassung der Gewichte $D_t(i)$ für alle $1, \dots, m$ Fälle). Gehen Sie hierbei auch darauf ein, wie korrekt und falsch klassifizierte Fälle in einer Iteration t bewertet werden.

(6 Punkte)

- b) Im Folgenden wurde der Algorithmus *AdaBoost.M1* auf einen medizinischen Datensatz angewandt. Klassifiziert wurden Personen, beschrieben durch verschiedene Attribute, durch ihren Diabetes-Typ bzw. dem Befund keiner vorliegenden Erkrankung („Nein“, „Typ I“, „Typ II“). Unten dargestellt sind die Ergebnisse nach drei Iterationen des AdaBoost.M1-Verfahrens. Angegeben sind für jede Iteration der gewählte Schätzer $h_t(x_i)$ und die Gewichte der einzelnen Fälle $D_t(i)$ für alle Fälle $i = 1, \dots, 5$. Die bekannte Klassifikation eines Falles i ist mit y_i notiert.

Tabelle A.2.2: Gewichte $D_t(i)$ und Ausgaben der Schätzer $h_t(x_i)$ der Iterationen 1 – 3

i	$D_1(i)$	$h_1(x_i)$	$D_2(i)$	$h_2(x_i)$	$D_3(i)$	$h_3(x_i)$	y_i
1	0,2	Typ I	0,125	Typ II	0,25	Typ I	Typ I
2	0,2	Typ I	0,125	Typ II	0,25	Typ I	Typ I
3	0,2	Typ II	0,125	Typ II	0,0833	Typ I	Typ II
4	0,2	Typ II	0,125	Typ II	0,0833	Typ I	Typ II
5	0,2	Typ II	0,5	Nein	0,3333	Nein	Nein

Zusätzlich gegeben sind die β_t Werte für $t = 1, 2, 3$:

$$\beta_1 = 0,25; \beta_2 = 0,3333; \beta_3 = 0,1999.$$

- Überprüfen Sie die Korrektheit des Wertes für β_2 indem Sie diesen Wert eigenständig berechnen. Geben Sie hierzu Ihre Rechnung an. (2 Punkte)
- Bestimmen Sie mittels der Schätzfunktion $h_{\text{fin}}(x)$ die geschätzten Klassifikationen für die Werte $i = 1, \dots, 5$ über die Schätzer der ersten drei Iterationen. Wieviele Fälle werden korrekt klassifiziert? (7 Punkte)

Aufgabe A.3: Clustering

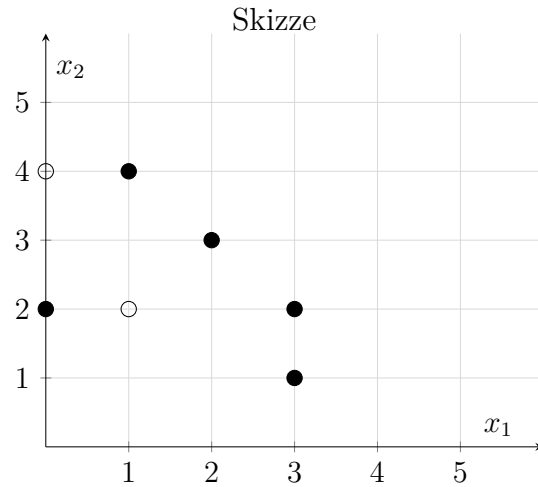
[15 Punkte]

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

Gegeben sind Datenpunkte x_1, \dots, x_5 (skizziert als ●) sowie die beiden Zentren $C = \{c_1, c_2\}$ (skizziert als ○).

Daten x_i		
i	x_1	x_2
1	0	2
2	1	4
3	2	3
4	3	1
5	3	2

Zentren c_i		
i	x_1	x_2
1	0	4
2	1	2



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

- a) In welche Cluster sind die Punkte x_i mit $i = \{1, 2, \dots, 5\}$ durch die beiden Zentren c_1 und c_2 aktuell eingeteilt?

Hinweis: Für die Ermittlung der Distanzen können Sie untenstehende Tabelle verwenden. Eine Rechnung ist nicht notwendig.

Manhattan-Distanzen					
c_i	$d(c_i, x_1)$	$d(c_i, x_2)$	$d(c_i, x_3)$	$d(c_i, x_4)$	$d(c_i, x_5)$
c_1					
c_2					

(3 Punkte)

- b) Rechnen Sie eine Iteration des *k-means* Verfahrens für $k = 2$ unter Verwendung der aktuellen Zentren c_1 und c_2 . Aktualisieren Sie die beiden Zentren und Bestimmen Sie abhängig von den resultierenden Werten die neuen Cluster. Wie wirkt sich ihre ermittelte Lösung auf die Fortführung des Algorithmus aus?

Hinweis: Für die Ermittlung der Distanzen können Sie untenstehende Tabelle verwenden. Eine Rechnung ist nicht notwendig.

Manhattan-Distanzen					
c_i	$d(c_i, x_1)$	$d(c_i, x_2)$	$d(c_i, x_3)$	$d(c_i, x_4)$	$d(c_i, x_5)$
c_1					
c_2					

(8 Punkte)

Auf der nächsten Seite geht es weiter ...

- c) Es seien unabhängig von den Teilaufgaben a) und b) die beiden Cluster $C_1 = \{x_1, x_5\}$ und $C_2 = \{x_2, x_3, x_4\}$ definiert. Die Manhattan-Distanzen zwischen den einzelnen Punkten seien in folgender Tabelle gegeben.

Manhattan-Distanzen					
$d(x_i, x_j)$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
$i = 1$	0	2	2	4	3
$i = 2$		0	2	5	4
$i = 3$			0	3	2
$i = 4$				0	1
$i = 5$					0

Bestimmen Sie die Manhattan-Distanzen zwischen den beiden Clustern nach dem *Average Linkage*-Maß und dem *Complete Linkage*-Maß. (4 Punkte)

Formelsammlung zum Aufgabenblock A

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

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

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

$$D_{t+1}(i) = \frac{D_t(i) \cdot e^{-\alpha_t \cdot y_i \cdot h_t(x_i)}}{Z_t}$$

$$H(x) = \operatorname{sign} \left(\sum_{t=1}^T \alpha_t \cdot h_t(x) \right)$$

$$f(x) = \sum_{t=1}^T \alpha_t \cdot h_t(x)$$

$$\epsilon_t = \sum_{i|h_t(x_i) \neq y_i}^m D_t(i)$$

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \cdot \begin{cases} \beta_t & \text{if } h_t(x_i) = y_i \\ 1 & \text{otherwise} \end{cases}$$

$$h_{fin}(x) = \operatorname{argmax} \left\{ \sum_{t|h_t(x)=y} \log\left(\frac{1}{\beta_t}\right) \mid y \in Y \right\}$$

$$\forall x, y \in \mathbb{R}^n : d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

$$D^{sl}(C_1, C_2) = \min\{d(a, b) \mid a \in C_1 \wedge b \in C_2\}$$

$$D^{cl}(C_1, C_2) = \max\{d(a, b) \mid a \in C_1 \wedge b \in C_2\}$$

$$D^{al}(C_1, C_2) = \frac{1}{|C_1| \cdot |C_2|} \cdot \sum_{a \in C_1} \sum_{b \in C_2} d(a, b)$$

$$D^{cm}(C_1, C_2) = d(c(C_1), c(C_2)) \text{ with}$$

$$c(C) = (c_1, c_2, \dots, c_m) \text{ where } \forall i \in \{1, \dots, m\} : c_i = \frac{1}{m} \cdot \sum_{a \in C} a_i.$$

Aufgabenblock B:

Computer Hardware und Systembetrieb

(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

Bearbeiten Sie alle der folgenden 5 Aufgaben! (insgesamt 45 Punkte)

Aufgabe B.1: Von Neumann Konzept [10 Punkte]

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

Aufgabe B.2: Pipeline [5 Punkte]

Beschreiben Sie das Prinzip einer Befehls-Pipeline.

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)

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

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.
(4 Punkte)

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
1. bei der Interaktion zwischen Webclient (WWW-Browser) und Webserver,
 2. bei der Interaktion zwischen dem Webserver und einem weiteren Webserver bei der Nutzung eines Webservice. (4 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Aufgabe D.1: Relationales Modell

[15 Punkte]

- a) Beschreiben und erläutern Sie die Grundkonzepte des Relationenmodells (Typ, Tupel, Relation) einzeln und in einem geeigneten Gesamtzusammenhang. (10 Punkte)
- b) Verdeutlichen Sie die Konzepte Kandidatschlüssel und Fremdschlüssel anhand selbstgewählter Beispiele. (5 Punkte)

Aufgabe D.2: Transaktionsmanagement

[15 Punkte]

Beschreiben und erläutern Sie das Konzept und den Ablauf einer Transaktion in einem Datenbanksystem. Geben Sie hierfür zudem ein geeignetes Beispiel an und gehen Sie dabei auch auf das Problem möglicher Deadlocks ein.

Aufgabe D.3: Anfrageoptimierung

[15 Punkte]

Beschreiben und erläutern Sie die Kernphasen der Anfrageoptimierung.