

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

Bergische Universität Wuppertal

Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

SoSe 2023

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	25.09.2023
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar) Aufgabenblock A: beigelegte Formelsammlung.

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 8 (acht) Seiten.

Unterschrift: _____

Aufgabenblock A:

Knowledge-based systems

Aufgabe A.1: Entscheidungsbäume

[11 Punkte]

Zur Einschätzung der Qualität von Olivenölen in die Qualitätsstufen (Klassen) *hoch* und *niedrig* liegt der folgende Datensatz mit 9 Fällen und den Attributen *Schärfe*, *Bitterkeit*, *Fruchtigkeit* vor, die jeweils in drei Kategorien eingeordnet werden $\{1, 2, 3\}$.

i	Schärfe	Bitterkeit	Fruchtigkeit	Qualität
1	2	2	3	hoch
2	1	3	1	niedrig
3	1	1	1	niedrig
4	3	3	3	hoch
5	1	3	1	niedrig
6	2	3	2	hoch
7	3	2	3	hoch
8	2	1	1	niedrig
9	3	3	3	hoch

- a) Berechnen Sie die Werte $info(M, \text{Schärfe})$, $info(M, \text{Bitterkeit})$ und $info(M, \text{Fruchtigkeit})$. Zur schnelleren Lösungsfindung können Sie die folgenden vorberechneten Werte benutzen:
 $log_2(\frac{1}{3}) = -1,58$, $log_2(\frac{3}{5}) = -0,74$, $log_2(\frac{2}{3}) = -0,58$, $log_2(\frac{2}{5}) = -1,32$ (8 Punkte)
- b) Geben Sie den Entscheidungsbaum an, der bei Anwendung von (ID3) (Decision tree construction) mit dem Gain-Kriterium auf den gegebenen Datensatz resultiert. (3 Punkte)

Aufgabe A.2: Bayes Netzwerke

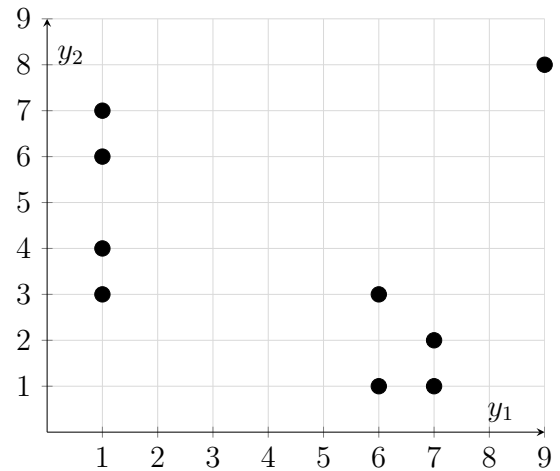
[12 Punkte]

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 $P(A) = 0.4$, $P(B|A) = 0.3$, $P(B|\neg A) = 0.1$, $P(C|A) = 0.2$,
 $P(D|B, C) = 0.2$, $P(D|\neg B, C) = 0.9$, $P(D|B, \neg C) = 0.5$, $P(D|\neg B, \neg C) = 0.7$.

- a) Zeichnen Sie das Netzwerk inklusive der CPTs. Geben Sie auch an, welche weitere Wahrscheinlichkeit noch angegeben werden müsste, um vollständige Tabellen zu erhalten. (6 Punkte)
- b) Es gilt $P(A|C) = 0.16$. Berechnen Sie mithilfe dieser Information die noch fehlende Wahrscheinlichkeit. (6 Punkte)

Aufgabe A.3: Clustering, Separation und nächste Nachbarn**[22 Punkte]**Wir betrachten die Datenpunkte x_1, \dots, x_9 .

i	y_1	y_2
1	1	4
2	1	7
3	1	6
4	1	3
5	7	2
6	6	3
7	6	1
8	7	1
9	9	8



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 $c_1 = (3, 5)$ und $c_2 = (8, 2)$ wie im k -means Verfahren. (4 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 . (4 Punkte)
- Zeigen Sie, dass die Punktmenge $M_+ := \{x_5, x_6, x_7, x_8, x_9\}$ und $M_- := \{x_1, x_2, x_3, x_4\}$ absolut linear separabel sind, indem Sie passende Gewichtparameter 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_- . (3 Punkte)
- Klassifizieren Sie den Punkt $x_{10} = (2, 5)$ mit Hilfe des einfachen 3-nearest-neighbor Verfahrens. Die Punkte x_1, x_2, x_3, x_4 sind dabei der Klasse A und die Punkte x_5, x_6, x_7, x_8 und x_9 der Klasse B zugeordnet. (4 Punkte)
- Bestimmen Sie geeignete Werte für ϵ und P_{min} , sodass der Algorithmus DBSCAN genau 2 Cluster findet und genau ein Datenpunkt als Noise eingeordnet wird. Zeigen Sie durch Anwendung des Verfahrens, dass diese Parameterbelegung zu dem gewünschten Ergebnis führt. (7 Punkte)

Formelsammlung zum Aufgabenblock A

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Aufgabenblock B:

Computer Hardware und Systembetrieb

(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

Aufgabe B.1: Pipeline

[5 Punkte]

Erläutern Sie die Begriffe Stalling, Bubble und (Pipeline-)Flush.

Aufgabe B.2: 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.3: 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. (10 Punkte)

Aufgabe B.4: 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)

Aufgabe B.5: HPC-Cluster/COTS-Produkte

[10 Punkte]

- a) Was versteht man unter COTS-Produkten? (4 Punkte)
- b) Wieso verwendet man heute oft Standardtechnologien anstelle spezieller Entwicklungen für den HPC-Bereich? (3 Punkte)
- c) Für welche Komponenten eines HPC-Clusters lassen sich COTS-Produkte nutzen? (3 Punkte)

Aufgabenblock C:

Kommunikationssysteme (Internet-Technologien)

Aufgabe C.1: Internet

[15 Punkte]

- a) Was versteht man unter dem Internet aus physischer Sicht, was aus logischer Sicht?
(4 Punkte)
- b) Erläutern Sie die Aufgaben und Eigenschaften der Schichten des Internet-Referenzmodells.
(4 Punkte)
- c) 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)
- d) Nennen Sie die drei Hauptziele der Sicherheit im Internet und erläutern Sie diese in jeweils einem Satz.
(3 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 (Mail User Agent) des Absenders zum E-Mail-Client des Empfängers.
(8 Punkte)
- b) Welche Ziele verfolgt Phishing?
Erläutern Sie eine der Methoden, die für Phishing-Angriffe genutzt werden.
(3 Punkte)
- c) Skizzieren Sie auf Anwendungsebene (Layer 4) die Nutzung eines Web-Mail-Dienstes per Webbrowser, wobei der Web-Mailer sowohl für den Abruf empfangener E-Mail von einem Posteingangsserver als auch für das Versenden von E-Mail über einen Postausgangsserver genutzt wird. Geben Sie in der Skizze die jeweils genutzten Anwendungsprotokolle zwischen den beteiligten Anwendungen und Diensten an.
(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, JavaScript und DOM der WWW-Architektur.
(7 Punkte)
- b) Stellen Sie in einer Skizze die HTTP-Interaktion zwischen einem Webclient (WWW-Browser) und einem Webserver dar, wobei eine angeforderte DHTML-Ressource dynamisch per PHP-Script generiert werden soll und clientseitig JavaScript zum Einsatz kommt. Die Skizze soll dabei insbesondere die für die Erzeugung der dynamisch generierten Inhalte genutzten Funktionseinheiten enthalten. Beschreiben Sie den gesamten Vorgang mit allen notwendi-

gen Arbeitsschritten von der Ressourcenanforderung bis zum Abschluss des Seitenaufbaus (Rendering). (8 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Aufgabe D.1: Transaktionsmanagement [15 Punkte]

- a) Beschreiben Sie die 3 typischen Probleme bei der Verarbeitung von Daten in Datenbanksystemen. (12 Punkte)
- b) Welches Konstrukt wird verwendet, um die 3 typischen Probleme zu vermeiden? (3 Punkte)

Aufgabe D.2: Erweiterbarkeit von Relationalen Datenbanksystemen [15 Punkte]

- a) Wie können Daten in XML-Dokumenten in Relationalen Datenbanksystemen abgebildet und gespeichert werden? Beschreiben Sie detailliert 2 differente Ansätze. (12 Punkte)
- b) Wird Erweiterbarkeit im Relationenmodell auf Ebene der Relationen oder auf Ebene der Datentypen realisiert? Begründen Sie Ihre Aussage. (3 Punkte)

Aufgabe D.3: Relationenalgebra [15 Punkte]

- a) Erläutern Sie die folgenden Operationen der Relationenalgebra: PRODUCT (Kartesisches Produkt), RESTRICT (Restriktion) und JOIN (Verbundoperation).
Welcher Zusammenhang besteht zwischen diesen 3 Operationen? (10 Punkte)
- b) Erläutern Sie das Geschlossenheitsprinzip der Relationenalgebra. (5 Punkte)

Aufgabe D.4: Interne Ebene [15 Punkte]

- a) Erläutern Sie die 3-Ebenen-Architektur Relationaler Datenbanksysteme. (6 Punkte)
- b) Erläutern Sie den Prozess des Zugriffs auf Daten in einem Datenbanksystem. (9 Punkte)