

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

Bergische Universität Wuppertal
Schumpeter School of Business and Economics

Bachelor of Science / Bachelor of Arts

SS 2020

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:	09.09.2020
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar), Aufgabenblock A: beigelegte 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 8 (acht) Seiten.

Unterschrift: _____

Aufgabenblock A:

Knowledge-based systems

Aufgabe A.1: Entscheidungsbäume

[15 Punkte]

Gegeben ist ein Datensatz zur Analyse von abgelegten Führerscheinprüfungen. Das Attribut Fahrstunden (**w**enig/**m**ittel/**v**iele) kategorisiert die insgesamt absolvierten Fahrstunden. Das Attribut Alter kategorisiert die Prüflinge in die Alterstufen **j**ung und **a**lt. Die Daten werden hinsichtlich des Bestehens des Teilnehmers klassifiziert. Der Wert 0 steht für nicht-bestanden und der Wert 1 zeigt das Bestehen des Prüflings an.

Datensatz		
Fahrstunden	Alter	Bestanden?
w	j	0
w	j	1
m	j	0
m	a	0
m	a	1
v	j	0
v	a	1
v	a	1

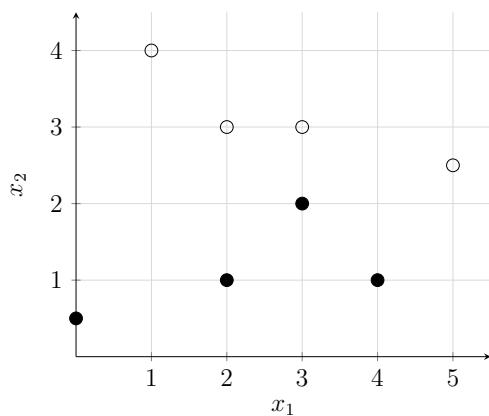
- a) Berechnen Sie die Werte $info(M, \text{Fahrstunden})$ und $info(M, \text{Alter})$. Zur schnelleren Lösungsfindung können Sie die folgenden vorberechneten Werte benutzen:
 $info(M) = 1$, $log_2(1/2) = -0,5$, $log_2(1/3) = -1,58$, $log_2(2/3) = -0,58$,
 $log_2(1/4) = -2$, $log_2(3/4) = -0,42$ (6 Punkte)
- b) Erzeugen Sie einen vollständigen Entscheidungsbaum mittels des ID3-Algorithmus unter Anwendung des Gain-Kriteriums mittels Ihrer in a) berechneten Werte. Sofern Sie die Teilaufgabe a) nicht gelöst haben, nehmen Sie an, dass $info(M, \text{Fahrstunden}) > info(M, \text{Alter})$ gilt. *Hinweis: Zur Lösung sind keine weiteren Rechnungen notwendig.* (4 Punkte)
- c) Das Verfahren ID3 erzeugt für jede vorhandene Ausprägung eines Attributs eine neue Verzweigung. Welche Problematik entsteht hierdurch bei Attributen mit vielen diskreten Ausprägungen oder einem kontinuierlichen Wertebereich? Erläutern Sie zudem kurz eine Möglichkeit das Entscheidungsbaumverfahren auch mit solchen Attributen sinnvoll durchführen zu können. (5 Punkte)

Aufgabe A.2: Das Perzeptron

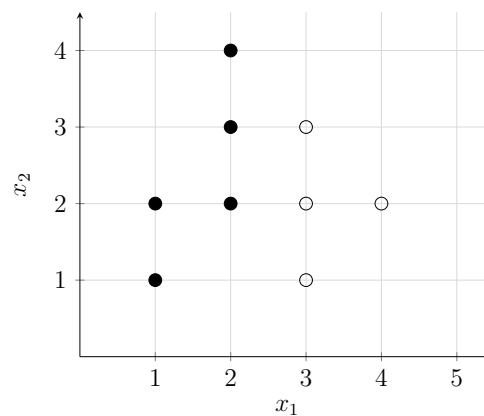
[12 Punkte]

a) Gegeben sind drei Skizzen von Datenpunkten mit binärer Klassifikation (durch unterschiedliche Färbung kenntlich gemacht).

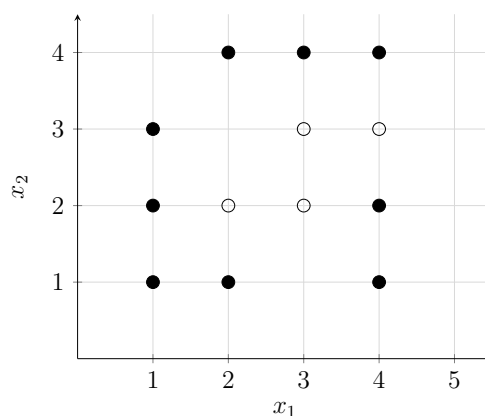
1)



2)



3)



Skizzieren Sie jeweils eine - wenn möglich lineare - Separation der Datenpunkte. Für welche der Skizzen ist keine lineare Separation möglich? (3 Punkte)

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

Für jeden Datensatz mit genau vier Datenpunkten und binärer Klassifikation existiert immer eine lineare Separation. (4 Punkte)

c) Der *Perzeptron-Lernalgorithmus* findet in endlich vielen Schritten eine Separation für linear separierbare Datensätze. Erläutern Sie kurz wie die Lernmethode des Perzeptrons auch auf nicht linear separierbare Datensätze angewendet werden kann. Ist auch für solche Datensätze eine Separation ohne Fehler möglich? Begründen Sie Ihre Antwort. (5 Punkte)

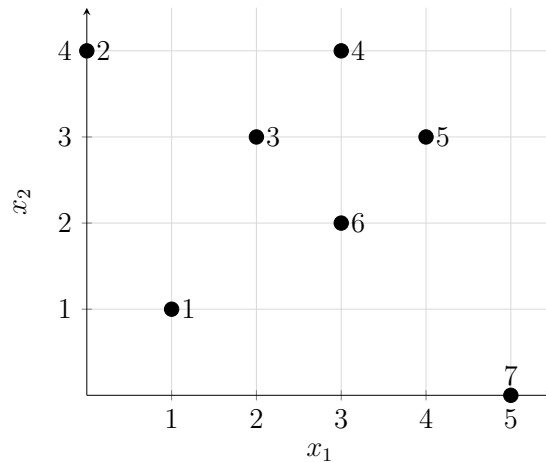
Aufgabe A.3: Ensemble Learning**[5 Punkte]**

Betrachtet wird eine Iteration des klassischen Adaboost-Verfahrens unter Anwendung von schwachen Schätzern. In einer beliebigen Iteration t generiert der beste schwache Schätzer keine Fehler für den zu lernenden Datensatz ($\epsilon_t = 0$). Erläutern Sie wie das Lernverfahren sinnvoll auf dieses Szenario reagiert. Beurteilen Sie zudem die resultierende Schätzfunktion, die durch die in den vorherigen Iterationen $1, \dots, t - 1$ ausgewählten Schätzern h_1, \dots, h_{t-1} definiert ist.

Aufgabe A.4: Density-Based Spatial Clustering**[13 Punkte]**

Im Folgenden wird das Clusteringverfahren *DBSCAN* betrachtet. Gegeben sind die Parameter $p^{min} = 5$ (Hinweis: Ein Punkt ist sein eigener Nachbar) und $\epsilon = 3$. Als Distanzmaß kommt die Manhattan-Distanz zum Einsatz.

Gegeben seien *sieben* Punkte x_1, \dots, x_7 , die untenstehend skizziert sind.



- Kategorisieren Sie alle gegebenen Punkte entweder als 1. Kernpunkte (core points), als 2. Grenzpunkte (border points) oder 3. als Rauschen (noise). (4 Punkte)
- Bestimmen Sie nun den oder die resultierenden Cluster durch Einsatz des Algorithmus' DBSCAN. (5 Punkte)
- Nehmen Sie zu der folgenden These begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort erhält keine Punkte. *Für zwei Punkte q und p gilt: Ist q direkt dichte-erreichbar (directly density reachable) von p aus, dann gilt auch, dass p direkt dichte-erreichbar (directly density reachable) von q aus ist.* (4 Punkte)

Formelsammlung zum Aufgabenblock A

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

$$gain(M, A_j) = info(M) - info(M, A_j)$$

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

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

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

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

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

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

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

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: Superskalare CPUs [10 Punkte]

Was ist eine Superskalare CPU?

Mit welchen Architekturmerkmalen erreicht man Superskalarität?

Welche neuen Probleme (Flaschenhälse) entstehen bei superskalaren CPUs?

Aufgabe B.3: Instruction Reordering [10 Punkte]

In welchen Fällen ist Instruction Reordering sinnvoll?

Welche Voraussetzungen an die Prozessorhardware müssen dabei erfüllt sein?

Was ist zu beachten, um keine falschen Resultate zu erhalten?

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: Speicherhierarchien [10 Punkte]

- a) Skizzieren Sie eine typische Speicherhierarchie vom Prozessor bis zu Archivierungsmedien einschließlich der wichtigsten Attribute. (6 Punkte)
- b) Welche Auswirkungen von Memory Wall und Power Wall könnten den Hauptspeicher zukünftig betreffen? (4 Punkte)

Aufgabenblock C:

Kommunikationssysteme (Internet-Technologien)

Aufgabe C.1: Internet

[15 Punkte]

- a) Was versteht man unter dem Internet aus physischer und was aus logischer Sicht?
(4 Punkte)
- b) Wie ist eine Internet-Adresse (IPv4) aufgebaut? Was ist eine Netzwerkmaske und wofür wird sie eingesetzt? Welche Vor- und Nachteile hat das CIDR-Schema im Gegensatz zum klassenbezogenen Adressierschema?
(3 Punkte)
- c) Skizzieren Sie das Internet-Referenzmodell und erläutern Sie kurz die Aufgaben und Eigenschaften der einzelnen Schichten.
(4 Punkte)
- d) Nennen Sie die wichtigsten Einträge im Header eines TCP-Segments (TCP-Header) und erläutern Sie jeweils deren Zweck.
(4 Punkte)

Aufgabe C.2: E-Mail

[15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren (Store-And-Forward-Übertragung) für die Übermittlung einer E-Mail vom E-Mail-Client des Absenders zum E-Mail-Client des Empfängers.
(8 Punkte)
- b) Beschreiben Sie, wie sich E-Mail-Anhänge für sichere und vertrauenswürdige E-Mail 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?
(7 Punkte)

Aufgabe C.3: World Wide Web

[15 Punkte]

- a) Erläutern Sie kurz folgende Standards und Technologien im WWW: HTTPS, RIA, SOAP, XHTML, URL, CSS, JavaScript, DOM.
(8 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 (incl. JavaScript) dynamisch per Java-Servlet generiert wird. Beschreiben Sie kurz alle Arbeitsschritte von der Ressourcenanforderung bis zum Abschluss des Seitenaufbaus (Rendering).
(7 Punkte)

Aufgabenblock D:

Datenorganisation (Datenbankmanagementsysteme)

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

Aufgabe D.1: Interne Ebene [15 Punkte]

- a) Erläutern Sie die Gründe dafür, warum ein Datenbanksystem eine automatische Anfrageoptimierung vornimmt. (6 Punkte)
- b) Erläutern Sie die Grundkonzepte von Indizes bei der Speicherung von Daten. (6 Punkte)
- c) Erläutern Sie das Konzept des Mehrebenenindex (Multilevel-Index). (3 Punkte)

Aufgabe D.2: Sicherheit [15 Punkte]

- a) Erläutern Sie die Grundkonzepte von Authentifizierung und Autorisierung. (6 Punkte)
- b) Was bedeutet im Kontext der Autorisierung „Discretionary Access Control“? (3 Punkte)
- c) Erläutern Sie das Konzept des „Role Based Access Control“. (6 Punkte)

Aufgabe D.3: Architektur [15 Punkte]

- a) Erläutern Sie 4 typische Komponenten eines Datenbanksystems. (8 Punkte)
- b) Nennen Sie 7 Ziele verteilter Datenbanksysteme in jeweils einem Satz! (7 Punkte)

Aufgabe D.4: Relationenalgebra [15 Punkte]

- a) Erläutern Sie die folgenden 3 Operationen der Relationenalgebra: UNION (Vereinigungsmenge), INTERSECT (Schnittmenge), DIFFERENCE (MINUS, Differenzmenge) anhand eines selbst gewählten Beispiels. Was ist der Kernaspekt der alles 3 Operationen gemein ist? (10 Punkte)
- b) Diskutieren Sie das Geschlossenheitsprinzip der Relationenalgebra! (5 Punkte)