

# 2

Sicherstellen der Datenqualität

**Teil 2 der Abschlussprüfung**

## Allgemeine Korrekturhinweise

Die Lösungs- und Bewertungshinweise zu den einzelnen Handlungsschritten sind als Korrekturhilfen zu verstehen und erheben nicht in jedem Fall Anspruch auf Vollständigkeit und Ausschließlichkeit. Neben hier beispielhaft angeführten Lösungsmöglichkeiten sind auch andere sach- und fachgerechte Lösungsalternativen bzw. Darstellungsformen mit der vorgesehenen Punktzahl zu bewerten. Der Bewertungsspielraum des Korrektors (z. B. hinsichtlich der Berücksichtigung regionaler oder branchenspezifischer Gegebenheiten) bleibt unberührt.

Zu beachten ist die unterschiedliche Dimension der Aufgabenstellung (nennen – erklären – beschreiben – erläutern usw.).

Für die Bewertung gilt folgender Punkte-Noten-Schlüssel:

Note 1	=	100 – 92 Punkte	Note 2	=	unter	92 – 81 Punkte	
Note 3	=	unter	81 – 67 Punkte	Note 4	=	unter	67 – 50 Punkte
Note 5	=	unter	50 – 30 Punkte	Note 6	=	unter	30 – 0 Punkte

## 1. Aufgabe (25 Punkte)

a) 6 Punkte

<b>Künstliche Intelligenz</b>	Teilbereich der Informatik, Computerprogramme können Probleme eigenständig lösen, indem sie menschliche Intelligenz nachahmen.
<b>Maschinelles Lernen</b>	Teilbereich der künstlichen Intelligenz, das Verfahren beinhaltet, die aus Erfahrung lernen können. Anhand von Beispieldaten werden Muster erkannt, mit denen unbekannte Daten richtig beurteilt werden können.
<b>Deep Learning</b>	Teilbereich des maschinellen Lernens, der künstliche neuronale Netze mit einer oder mehreren Schichten verwendet

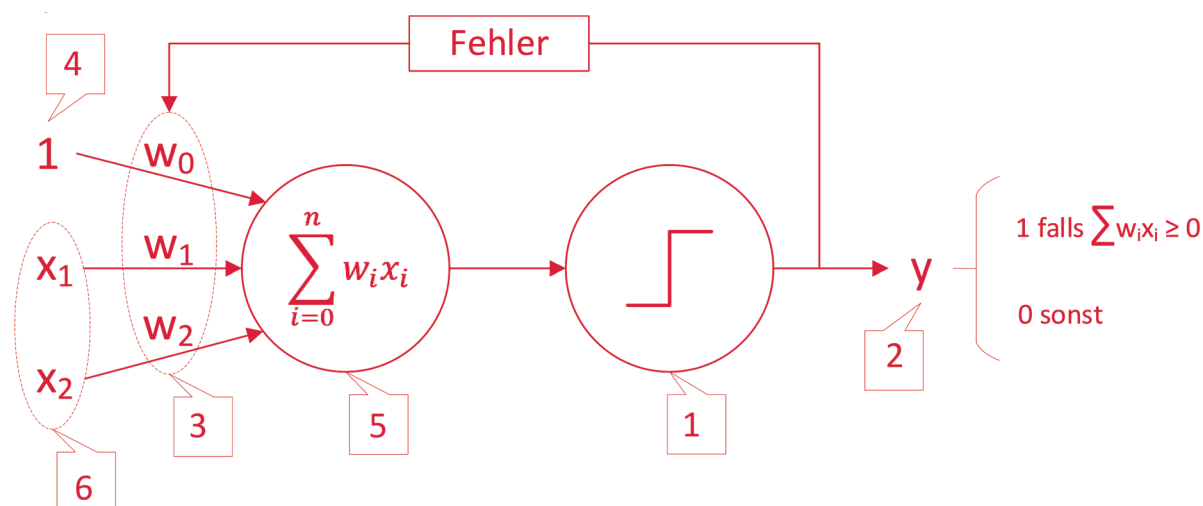
b) 4 Punkte

<b>Supervised Learning</b>	Voraussetzung für SL sind beschriftete Daten (labeled Data). Ziel: Parameter einer Funktion so einzustellen, dass die richtigen Ausgaben generiert werden. Dazu werden in der Lernphase Daten in eine Funktion gegeben. Die Ausgabe wird mit der Zielausgabe verglichen. Je nach Abweichung werden die Parameter der Funktion angepasst. Sind die Parameter optimal eingestellt kann die Funktion verwendet werden.
<b>Unsupervised Learning</b>	Ziel: Etwas über die Daten zu lernen. Hier wird ausschließlich durch die Muster in den Eingabedaten gelernt. Z. B. werden mit statistischen Clustering-Verfahren Datenwerte gruppiert. Es werden keine Zieldaten benötigt.

c) 2 Punkte

Die Daten sind linear separierbar, d. h. es kann eine Trennlinie eingezeichnet werden.

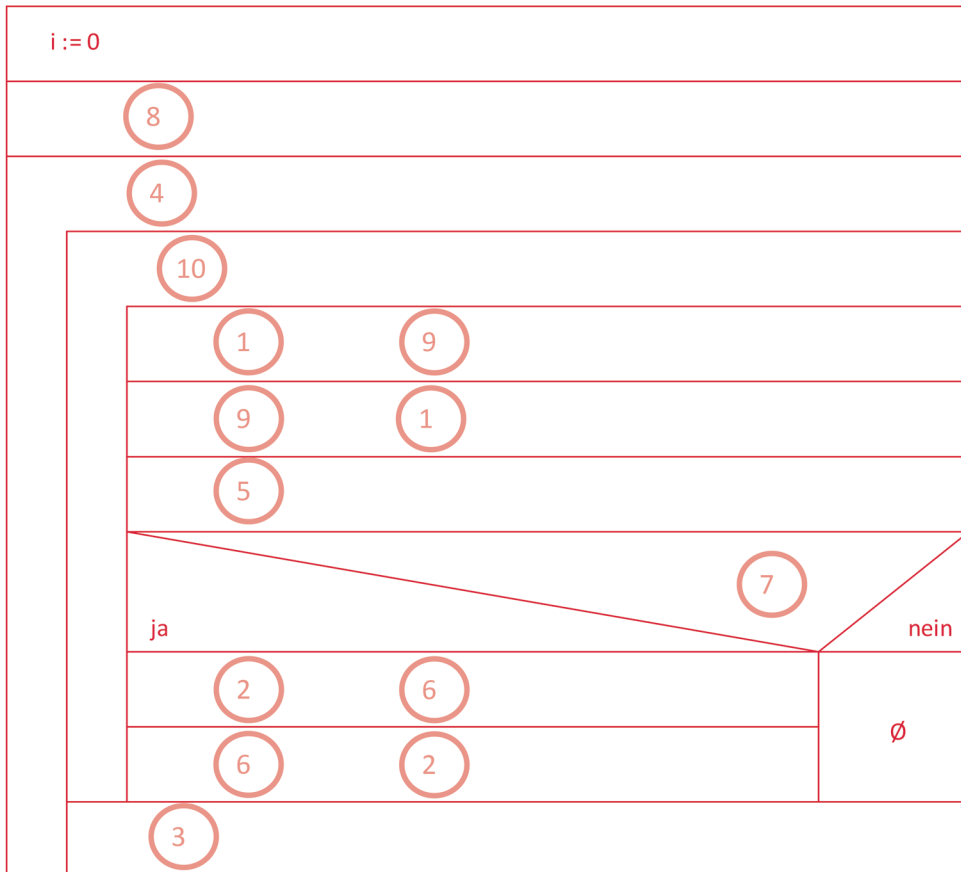
da) 3 Punkte



db) 5 Punkte

Die Eingabedaten  $x_1$  und  $x_2$  und der Bias mit dem Wert 1 werden mit ihren Gewichten multipliziert und in der Nettoeingabefunktion aufaddiert. Ist das Ergebnis größer gleich 0, so wird eine 1 ausgegeben, ansonsten 0. Mit Trainingseingabedaten, deren Ausgaben bekannt sind, werden bei Falschausgaben die Gewichte angepasst. Nach erfolgreichem Training liefert das Perzeptron für neue Eingabedaten die richtige Ausgabe.

e) 5 Punkte



## 2. Aufgabe (26 Punkte)

a) 13 Punkte

Je Tabelle 1 Punkt = 3 Punkte

Je PK 1 Punkt = 3 Punkte

Je FK 1 Punkt = 2 Punkte

Name getrennt = 1 Punkt

Je weiteres Attribut 0,5 Punkte = 2 Punkte

Je Beziehung 1 Punkt = 2 Punkte

Beispielüberführung



Die Kontaktperson des Kunden kann auch in eine eigene Tabelle ausgelagert werden.

ba) 4 Punkte

Erklärung: 2 Punkte, ein Vorteil: 2 Punkte

In Datenbankmanagementsystemen (DBMS) ist ein Prepared Statement eine parametrisierte Anweisung oder Abfrage eine Funktion zur Vorkompilierung von SQL-Code, die ihn von den Daten trennt. Die Vorteile von vorbereiteten Anweisungen sind:

- Effizienz, da sie ohne Neukompilierung wiederholt verwendet werden können
- Sicherheit, da SQL-Injektionsangriffe reduziert oder eliminiert werden

bb) 9 Punkte

Die Tabellen, die in a) erstellt wurden, sollen hier verwendet werden, d. h. die SQL-Lösung kann abweichen.

Folgefehler

```
# 3P Create Table
cursor.execute("CREATE TABLE `Kunden` (
    `id` INTEGER NOT NULL,
    `name` TEXT NOT NULL,
    `vorname` TEXT NOT NULL,
    `tel` TEXT NOT NULL,
    PRIMARY KEY (`id`)
);")

# 4P Insert Data
sql = "INSERT INTO `Kunden` VALUES (%s, %s, %s, %s)"
values = (1, "Panter", "Rosa", "030/22459326")
cursor.execute(sql, values)

# 2P DB Verbindung schließen
cursor.close()
```

Wenn a) nicht gemacht wurde, soll mit dem Beispieldatensatz gearbeitet werden und dann ergibt sich folgende Lösung:

MDE-Gerät {seriennummer="669-818", typ="ScanFix 5L", letztes\_Update=2021-01-01, zertifikat="IP44"}

```
# 3P Create Table
cursor.execute("CREATE TABLE `MDEGeraet` (
    `seriennummer` VARCHAR(7) NOT NULL,
    `typ` TEXT NOT NULL,
    `letztes_Update` TEXT NOT NULL,
    `zertifikat` TEXT NOT NULL,
    PRIMARY KEY (`seriennummer`)
);")

# 4P Insert Data
sql = "INSERT INTO `MDEGeraet` VALUES (%s, %s, %s, %s)"
values = ("669-818", "ScanFix 5L", 2021-01-01, "IP44")
cursor.execute(sql, values)

# 2P DB Verbindung schließen
cursor.close()
```

### 3. Aufgabe (28 Punkte)

a) 2 Punkte

```
{  
  „artikelnr“: 1,  
  „zeitpunkt“: „01-APR-23 12.41.08 PM“,  
  „lagerbewegung“: 2  
}
```

b) 2 Punkte

Bei Restful Services liegt das Konzept zugrunde, dass eine Ressource über einen Webserver verfügbar ist und eindeutig über eine URL identifiziert wird.

So können die Daten vom Gerät an den verfügbaren Webservice übertragen werden.

c) 4 Punkte

HTTP-Methode	Operation für die Lagerbewegungen
GET	Lagerbewegungen abfragen
PUT	Lagerbewegungen verändern
POST	Lagerbewegungen neu erzeugen
DELETE	Lagerbewegungen löschen

da) 4 Punkte

```
SELECT artikelnr, MAX(menge) AS Max, MIN(menge) AS Min  
FROM lagerbestaende  
GROUP BY artikelnr;
```

db) 7 Punkte

```
SELECT artikelnr, ROUND(AVG(Menge),0) AS Erwartungswert,  
ROUND(STDDEV(menge),0) AS Standardabweichung  
FROM lagerbestaende  
GROUP BY artikelnr;
```

Hinweis: Falls GROUP BY in da) und db) fehlt, als Wiederholungsfehler werten.

dc) 6 Punkte

Erwartungswert:

$(8 + 16 + 9 + 4 + 12) / 5 = 49 / 5 = 9,8$  – wird auf 10 gerundet

Standardabweichung:

$(8 - 10)^2 + (16 - 10)^2 + (9 - 10)^2 + (4 - 10)^2 + (12 - 10)^2 =$   
 $4 + 36 + 1 + 36 + 4 = 81$

$81 / 5 = 16,2$  Wurzel aus 16,2 = 4,02 gerundet 4

Der Erwartungswert gibt an, welcher Wert im Durchschnitt erwartet wird.

Hier ist im Durchschnitt ein Lagerbestand von 10 zu erwarten.

Die Standardabweichung misst die durchschnittliche Abweichung der einzelnen Datenpunkte von ihrem Erwartungswert. Je größer die Standardabweichung ist, desto stärker sind die Datenpunkte von diesem Durchschnittswert gestreut. In diesem Beispiel ist die Standardabweichung mit 4 relativ hoch.

dd) 3 Punkte

Werte nach Größe sortieren. Bei ungerader Anzahl liegt der Median in der Mitte.

Sortierte Menge: 4, 8, 9, 12, 16

Die Anzahl der Werte ist ungerade → Der Wert liegt in der Mitte bei 9.

Der Median ist der Wert, der genau in der Mitte einer sortierten Datenreihe liegt und teilt die Daten in zwei gleich große Hälften.

Weniger anfällig für Ausreißer

Einblick in die zentrale Position der Daten

#### 4. Aufgabe (21 Punkte)

a) 3 Punkte

Die Datenaufbereitung ist ein Prozess, bei dem z. B.

- Rohdaten bereinigt und
- Rohdaten normalisiert werden, um hochwertige Daten für Analysen und weitere Datenmanagementaufgaben zu generieren.
- fehlerhafte Daten durch Default-Werte ersetzt oder
- fehlerhafte Daten entfernt werden.

ba) 8 Punkte

	AID	Lagerhalle	Regal	Fach	Durchschnittliche wöchentliche Entnahmen	Maximale wöchentliche Entnahme	Regalfachkapazität
1.	1	1	4	5	253,7	356	null
2.	1	1	3	5	187,6	200	200
3.	1	1	3	5	187,6	200	20
4.	2	1	3	5	null	null	400
5.	2	1	1	2	777,2	550	500
6.	3	2	1	3	-6	300	300
7.	3	2	1	4	250	300	300
8.	4	2	2	null	40,5	55	600

Vier aus fünf:

- Daten fehlen – 1. 4. 8. Zeile
- Daten sind nicht plausibel
  - Ausreißer Zeile 3
- Daten sind nicht plausibel
  - Zeile 5: Mehr entnehmen als vorhanden geht nicht.
- Daten sind nicht plausibel
  - Zeile 6: negative Entnahmen
- Daten liegen bis auf den letzten Wert doppelt vor – Zeile 2 und 3

bb) 6 Punkte

Drei aus vier:

- Zeile 3 streichen, da doppelte und fehlerhafte Daten enthalten sind.
- Zeile 4 streichen, da wesentliche Daten fehlen.
- Zeile 1 und 8: Ermittlbare Daten nachtragen – z. B. Regalkapazität und Fach, indem der Artikel liegt.
- Zeile 6: negative Zahl – Ermittlungsprozess überprüfen

c) 4 Punkte

- Es werden personenbezogene Daten erfasst. Dabei ist die DSGVO einzuhalten.
- Die freiwillige Einwilligung der Mitarbeiter kann erforderlich sein.
- Die Zweckgebundenheit muss erfüllt sein.
- Die Daten sind sparsam und anonymisiert zu erfassen.
- Die Daten sind vor unberechtigtem Zugriff zu schützen.
- u. a.



