

# Klausur Angewandte Informatik

## Lösungshinweise

Prüfungsdatum: 23. Januar 2019 – Prüfer: Etschberger, Jansen, Wesp

Studiengang: BW, IM

Punkte: 20, 12, 8, 8, 12, 21, 9 ; Summe der Punkte: 90

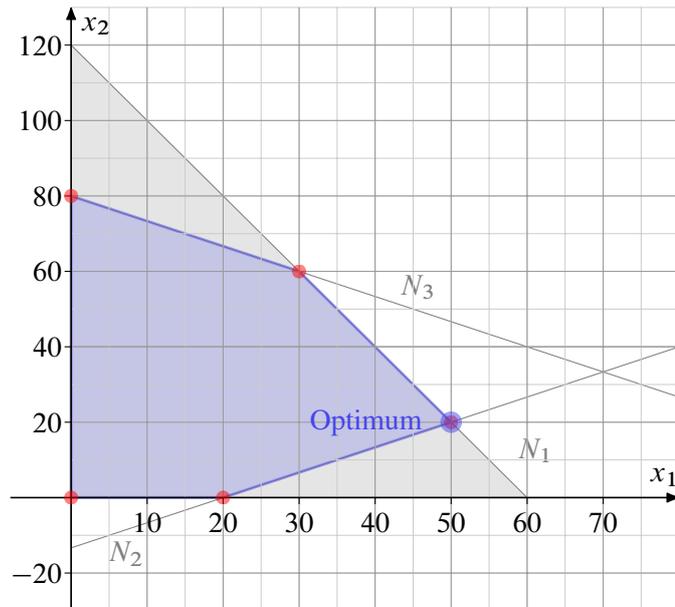
### Aufgabe 1

20 Punkte

Gegeben ist das folgende lineare Optimierungsproblem mit

- ▶ den Strukturvariablen  $x_1, x_2 \in \mathbb{R}_+$ ,
- ▶ der Zielfunktion  $f: \mathbb{R}_+^2 \rightarrow \mathbb{R}$  und
- ▶ den Nebenbedingungen  $N_1, N_2, N_3$  (dabei ist  $N_1$  nur teilweise gegeben).

$f$	$8x_1 +$	$3x_2$	$\rightarrow$	$\max$
$N_1$	$2x_1$	$+ x_2$	$\leq$	$120$
$N_2$	$2x_1 -$	$3x_2$	$\leq$	$40$
$N_3$	$2x_1 +$	$3x_2$	$\leq$	$240$



- a) Die graphische Repräsentation von  $N_1$  ist im Koordinatensystem eingezeichnet. Füllen Sie in der Tabelle oben die fehlenden Felder von  $N_1$  aus.
- b) Zeichnen Sie  $N_2, N_3$  sowie den Zulässigkeitsbereich des Problems in das Koordinatensystem ein.
- c) Markieren Sie die für ein Optimum in Frage kommenden Ecken.
- d) Zeichnen Sie alle Punkte der Zielfunktion für  $f(x_1, x_2) = 240$  ein (Isonutzengerade).
- e) Berechnen Sie die Koordinaten im Optimum sowie den dazugehörigen Wert der Zielfunktion.
- f) Nach einem Schritt des Simplexalgorithmus ergibt sich das folgende Tableau:

	$x_1$	$x_2$	$y_1$	$y_2$	$y_3$		Operation
⑤	0	-15	0	4	0	160	① + 4 · ③
⑥	0	4	1	-1	0	80	② - 1 · ③
⑦	1	$-3/2$	0	$1/2$	0	20	$+1/2 \cdot ③$
⑧	0	6	0	-1	1	200	④ - 1 · ③

Dabei bezeichnet die Zeile Nummer ⑤ die Zielfunktion. Führen sie einen weiteren Simplexschritt durch und geben Sie den Wert aller Struktur-, Schlupfvariablen sowie der Zielfunktion nach diesem Schritt an.

### Lösungshinweis:

Z	$8x_1 + 3x_2$	$\rightarrow$	max
$N_1$	$2x_1 + x_2$	$\leq$	120
$N_2$	$2x_1 + -3x_2$	$\leq$	40
$N_3$	$2x_1 + 3x_2$	$\leq$	240

	$x_1$	$x_2$	$y_1$	$y_2$	$y_3$		Operation
①	-8	-3	0	0	0	0	
②	2	1	1	0	0	120	
③	2	-3	0	1	0	40	
④	2	3	0	0	1	240	
⑤	0	-15	0	4	0	160	① + 4 · ③
⑥	0	4	1	-1	0	80	② - 1 · ③
⑦	1	$-3/2$	0	$1/2$	0	20	$+1/2 \cdot ③$
⑧	0	6	0	-1	1	200	④ - 1 · ③
⑨	0	0	$15/4$	$1/4$	0	460	⑤ + $15/4 \cdot ⑥$
⑩	0	1	$1/4$	$-1/4$	0	20	$+1/4 \cdot ⑥$
⑪	1	0	$3/8$	$1/8$	0	50	⑦ + $3/8 \cdot ⑥$
⑫	0	0	$-3/2$	$1/2$	1	80	⑧ - $3/2 \cdot ⑥$

## Aufgabe 2

12 Punkte

Es sollen kürzeste Wege durch einen Graph nach dem Algorithmus von *Floyd-Warshall* gefunden werden.

Gegeben ist dazu die unvollständige Zeichnung eines Graphen (siehe Abbildung 1) sowie die dazugehörige ebenfalls jeweils unvollständige Kosten- beziehungsweise Vorgängermatrix  $C_0$  bzw.  $P_0$ .

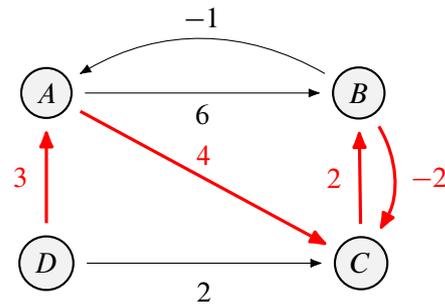


Abbildung 1: Graph zu Teilaufgabe a)

$$C_0 = \begin{pmatrix} 0 & 6 & 4 & \infty \\ -1 & 0 & -2 & \infty \\ \infty & 2 & 0 & \infty \\ 3 & \infty & 2 & 0 \end{pmatrix},$$

$$P_0 = \begin{pmatrix} A & A & A & 0 \\ B & B & B & 0 \\ 0 & C & C & 0 \\ D & 0 & D & D \end{pmatrix}$$

Ergänzen Sie die fehlenden Kanten des Graphen mit deren Kantengewichten in der Abbildung, sowie jeweils die fehlenden Einträge in der Kosten- und der Vorgängermatrix.

### Lösungshinweis:

s.o.

### Aufgabe 3

8 Punkte

Gegeben ist der Graph in Abbildung 2. Auf diesen soll der Algorithmus nach *Prim* angewendet werden.

- Tabelle 1 dokumentiert einen Iterationsschritt des Algorithmus. Um den wievielten Schritt handelt es sich dabei? Begründen Sie Ihre Antwort.
- Führen Sie gemäß *Prim* den nächsten Iterationsschritt durch und tragen Sie das Ergebnis in die fehlenden Felder von Tabelle 2 ein.
- Markieren Sie ohne weitere Berechnungen die Kanten des minimal-aufspannenden Baums in Abbildung 2 farbig.

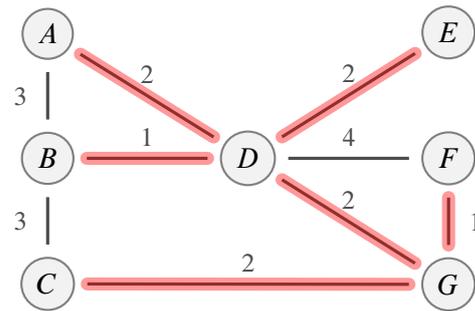


Abbildung 2: Graph zum *Prim*-Algorithmus

	A	B	C	D	E	F	G
$g$	0	1	3	2	2	4	2
$p$		D	B	A	D	D	D
$\beta$	1	1	0	1	1	0	0

Tabelle 1: Zu Aufgabe a)

	A	B	C	D	E	F	G
$g$	0	1	2	2	2	1	2
$p$		D	G	A	D	G	D
$\beta$	1	1	0	1	1	0	1

Tabelle 2: Nächster Iterationsschritt

### Lösungshinweis:

- siehe oben
- Es war der vierte Iterationsschritt. Das sieht man an der Anzahl der Einser.
- siehe oben
- siehe oben

## Aufgabe 4

8 Punkte

Wachtmeister Willi wertet die Videoaufzeichnung einer Radarfalle aus, um Geschwindigkeitsüberschreitungen von Autofahrern zu ahnden. Nachdem er die Aufzeichnungen der letzten 26 Tage durchgesehen hat, erstellt er eine Liste mit der Anzahl der pro Tag beobachteten Geschwindigkeitsüberschreitungen. Es ergibt sich:

Tag	1	2	3	4	5	6	7	8	9	10	11	12	13
Anzahl Überschreitungen	10	10	2	14	4	11	16	15	20	10	4	3	15
Tag	14	15	16	17	18	19	20	21	22	23	24	25	26
Anzahl Überschreitungen	24	4	3	14	17	16	7	13	10	9	18	13	17

Gehen Sie im Folgenden davon aus, dass die Anzahl der Geschwindigkeitsdelikte pro Tag eine einfache Stichprobe aus einer normalverteilten Grundgesamtheit darstellen.

- Wie müsste die Nullhypothese  $H_0$  und die Gegenhypothese  $H_1$  lauten, wenn Willi mit einem Test statistisch bestätigen möchte, dass die durchschnittliche Anzahl der Geschwindigkeitsdelikte höher als 10 ist?
  - Würden Sie eher ein hohes oder ein niedriges Signifikanzniveau wählen, wenn Sie diese Vermutung statistisch bestätigen wollen (Begründung)?
  - Führen Sie den Test zu einem Signifikanzniveau von  $\alpha = 5\%$  durch.
  - Was bedeutet der Fehler 1. Art hier?
- R e) Beim Ausführen eines Tests in R ergibt sich als Ausgabe:

```
t.test(x, mu = 9.5, alternative = "greater")
##
## One Sample t-test
##
## data: x
## t = 1.7464, df = 25, p-value = 0.04651
## alternative hypothesis: true mean is greater than 9.5
## 95 percent confidence interval:
## 9.543796      Inf
## sample estimates:
## mean of x
##      11.5
```

Was bedeutet in dieser Ausgabe  $t = \dots$ ,  $df = \dots$ ,  $p\text{-value} = ?$

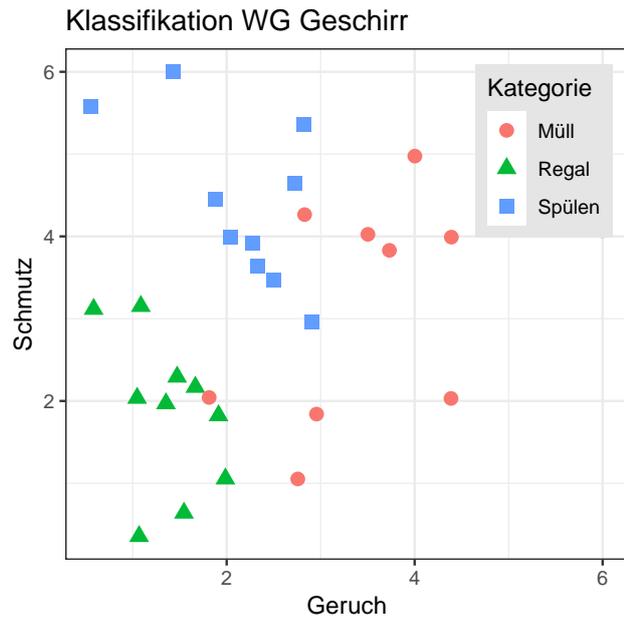
### Lösungshinweis:

- $H_0 : \mu = 10$  gegen  $H_1 : \mu > 10$ .
- Je größer  $\alpha$ , desto eher wird  $H_0$  abgelehnt, also sollte ein hohes Signifikanzniveau gewählt werden (dafür: höheres Risiko für Fehler 1. Art)
- $\bar{x} = 11.5$ ,  $s = 5.84$ ,  $x_{1-\alpha} = x_{0.95} = 1.708$ ,  $v = \frac{\bar{x} - 10}{s} \cdot \sqrt{26} \approx 1.31$ .  
 $B = (1.708; \infty)$
- Fehler 1. Art: Die Stichprobe führt zu einer Ablehnung der Nullhypothese ( $\mu = 10$ ), obwohl  $H_0$  stimmt, d.h. hier würde Willi nach dem Test davon ausgehen, dass es im Durchschnitt mehr als 10 Delikte pro Tag gibt, obwohl das nicht stimmt.

Walter, Willi und Wolfgang wollen in ihrer WG den Umgang mit gebrauchtem Geschirr neu regeln. Sie haben in einer Sitzung alle gemeinsam bei 30 benutzten Tellern entschieden, was mit diesen passieren soll, nämlich

- ▶ ohne weitere Maßnahme zurück ins *Regal* stellen,
- ▶ den Teller *spülen* oder
- ▶ da sich die Reinigung nicht mehr lohnt sofort in den *Müll* werfen.

Außerdem haben die drei jeweils pro Teller den Geruch ( $x_1$ ) sowie den Verschmutzungsgrad ( $x_2$ ) bewertet. Dabei entspricht 0 jeweils keiner, 6 einer äußerst starken Ausprägung.



Diese Daten sollen als Grundlage eines linearen Diskriminanzmodells dienen, um neue Objekte einer der drei Klassen zuzuordnen. Für die Klassenmittelwerte bzw. die zwei nötigen Gewichtungsvektoren  $g^1, g^2$  ergibt sich

Klasse	Klassenmittelwerte	
	$\bar{x}_1$	$\bar{x}_2$
Regal	1.37	1.86
Müll	3.64	3.21
Spülen	2.15	4.4

bzw.

$$g^1 = \begin{pmatrix} -1.035 \\ -0.437 \end{pmatrix},$$

$$g^2 = \begin{pmatrix} -0.627 \\ 0.797 \end{pmatrix}.$$

R.A. Fisher hat zur Identifikation neuer Objekte mittels dieser Gewichtungsvektoren folgende Entscheidungsregel vorgeschlagen:

„Berechne für ein neues  $x^T = (x_1, x_2)$  mit  $y^1 = x^T \cdot g^1$  und  $y^2 = x^T \cdot g^2$  für jede Klasse  $K$  die Distanz  $d(K) = (y^1 - \bar{y}_1^K)^2 + (y^2 - \bar{y}_2^K)^2$  und ordne  $x$  dann der Klasse  $K$  mit dem kleinsten  $d(K)$  zu.“

Dabei bezeichnet  $\bar{y}_1^K = \bar{x}^K g^1$  bzw.  $\bar{y}_2^K = \bar{x}^K g^2$  den Mittelwert der beiden Diskriminanzvariablen für die Klasse  $K$ . Hier im Beispiel kann  $K$  die 3 Werte *Regal*, *Müll*, *Spülen* annehmen.

Willi ist alleine zu Hause. Ein Teller liegt auf dem Tisch, den er mit

$$x^T = (\text{Geruch}, \text{Schmutz}) = (x_1, x_2) = (2.5, 2)$$

bewertet. Er möchte diesen Teller gemäß dieser Regel klassifizieren.

- a) Berechnen Sie  $\bar{y}_1^K$  bzw.  $\bar{y}_2^K$  für alle drei Klassen mit den WG-Daten.
- b) Geben Sie für jede der drei Klassen die Distanzfunktion  $d(K)$  an.
- c) Was passiert mit dem Teller auf Basis der neuen WG-Entscheidungsregel?

**Lösungshinweis:**

```
Y.Klassenmittel
```

```
##          LD1  LD2
## Müll    -5.17 0.277
## Regal   -2.23 0.621
## Spülen -4.14 2.159
```

```
X.new = data.frame(Geruch=c(2.5), Schmutz=c(2), Gruppe=c("neu"))
Fisher.predict(X.new, X.lda=X.lda)
```

```
##   Geruch Schmutz Gruppe Prognose  y1  y2 d.Spuelen
## 1    2.5      2   neu    Regal -3.46 0.025    5.02
##   d.Muell d.Regal Fisher.pred
## 1    2.98    1.86    Regal
```

## Aufgabe 6

21 Punkte

- R a) Gegeben ist ein Dataframe (lieferung) mit drei Spalten (postleitzahl, menge und kosten). Dabei sind die ersten beiden je ein Faktor und die letzte Spalte ist numerisch. Geben sie jeweils die R-Befehle an mit denen der dataframe wie jeweils beschrieben geändert wird:

1. Anlegen eines neuen Eintrags am Ende des Dataframes (mit passenden Daten ihrer Wahl!)

```
rbind(lieferung,
      data.frame(postleitzahl='99999',
                 menge='1000',
                 kosten=20000))
```

2. Sie wollen eine Spalte mit einem logischen Vektor hinzufügen, der genau dann TRUE ist falls die kosten größer als 10000 sind. Die neue Spalte soll den Namen teuer haben.

```
cbind(lieferung, teuer=lieferung$kosten>10000)
```

3. Sie möchten die 42. Zeile ihres Dataframes löschen.

```
lieferung[-42,]
```

4. Nachdem Sie alle obigen Änderungen gemacht haben, stellen Sie fest, dass Ihnen die Spaltenüberschriften nicht gefallen. Wie Ändern Sie diese in PLZ, Menge, Kosten und Teuer um?

```
names(daten) <-c('PLZ', 'Menge', 'Kosten', 'Teuer')
```

- R b) Sie erzeugen einen numerischen Vektor mittels `x <- seq(1, 11, length.out = 11)`. Welche der folgenden Befehle erzeugt den gleichen Vektor?

	wahr	falsch		wahr	falsch
<code>rep(1, 11)</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<code>1:11</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<code>seq(1, 11, 1)</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<code>seq(to=11, from=1)</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<code>10:110 / 10</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<code>seq(1:11:1)</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

- R c) Seien a und b zwei gleich lange numerische Vektoren der Länge n. Sie führen den Befehl `a==b` aus.

	wahr	falsch
Das Resultat ist ein numerischer Vektor.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Das Resultat ist ein logischer Vektor.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Das Resultat ist ein Vektor der Länge 1.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Das Resultat ist ein Vektor der Länge n.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Es wird ein Fehler ausgegeben.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Es wird eine Warnung ausgegeben.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Der Vektor a wird überschrieben.	<input type="checkbox"/>	<input checked="" type="checkbox"/>

- R d) Welche Aussagen bezüglich Dataframes in R sind korrekt?

	wahr	falsch
Ein Dataframe ist eine rechteckige Datenstruktur	X	
Jeder Eintrag einer Spalte in einem Dataframe muss vom gleichen Datentyp sein	X	
Jeder Eintrag einer Zeile in einem Dataframe muss vom gleichen Datentyp sein		X
In Dataframes kann man mit dem \$-Operator auf einzelne Spalten zugreifen	X	
In Dataframes sind Zeichenketten standardmäßig Faktoren	X	
is.data.frame() wandelt Matrizen in Dataframes um		X
Dataframes und Listen sind der gleiche Datentyp		X

R e) Geben Sie an, was die Ausführung der folgenden Zeilen R-Code ausgibt (drei Antworten!).

```
x <- c(3,1,4,1,5,9,pi)
y <- 3 <= x

y
all(y)
any(y)
```

f) Welchen Vorteil hat  $\text{\LaTeX}$  gegenüber einem Textverarbeitungsprogramm wie Word?

R

g) Aus der Statistikvorlesung kennen Sie das *Geburtstagsparadoxon*, in dem es darum geht, wie groß die Wahrscheinlichkeit ist, dass unter  $n$  Leuten mindestens 2 Leute am gleichen Tag Geburtstag haben. In R gibt es zum Berechnen dafür die Funktion `pbirthday()` (Hilfe: siehe Abb. 3).

Wie berechnen Sie mit diesem Vorwissen und der Hilfe von R die Wahrscheinlichkeit dafür, dass in einer Gruppe von 80 Leuten genau 3 Leute die gleiche vierstellige EC-Karten PIN haben?

birthday (stats)

R Documentation

## Probability of coincidences

### Description

Computes answers to a generalised *birthday paradox* problem. `pbirthday` computes the probability of a coincidence and `qbirthday` computes the smallest number of observations needed to have at least a specified probability of coincidence.

### Usage

```
qbirthday(prob = 0.5, classes = 365, coincident = 2)
pbirthday(n, classes = 365, coincident = 2)
```

### Arguments

<code>classes</code>	How many distinct categories the people could fall into
<code>prob</code>	The desired probability of coincidence
<code>n</code>	The number of people
<code>coincident</code>	The number of people to fall in the same category

Abbildung 3: Hilfe zur R-Funktion `pbirthday()`



R

h) Wozu dient das R-Paket `knitr`?



Lösungshinweis:

## Aufgabe 7

9 Punkte

R Gegeben ist die R-Funktion `test()` und sowie die Vektoren `x`, `y` und `z` mit

```
test <- function(x){
  if(is.logical(x)) {x <- is.numeric(x)}
  if(!is.numeric(x)) {x <- seq(2,10,length.out = length(x))}
  rval <- x^2 + 1
  return(rval)
}

x <- c(2,3,5,7)
y <- c("A","E","I","O","U")
z <- c(TRUE, FALSE, TRUE, TRUE, FALSE)
```

Geben Sie jeweils die Ausgabe der folgenden Funktionsaufrufe an.

```
test()
## Error in test(): Argument "x"
## fehlt (ohne Standardwert)
```

```
test(y = x)
## Error in test(y = x): unbenutztes
## Argument (y = x)
```

```
test(x)
## [1] 5 10 26 50
```

```
test(x = y)
## [1] 5 17 37 65 101
```

```
test
## function(x){
##   if(is.logical(x)) {x <- is.numeric(x)}
##   if(!is.numeric(x)) {x <- seq(2,10,length.out
##   rval <- x^2 + 1
##   return(rval)
## }
## <bytecode: 0x00000000195fb438>
```

```
test(z)
## [1] 5
```

Lösungshinweis: