

Klausur Angewandte Informatik

Lösungshinweise

Prüfungsdatum: 23. Januar 2019 – Prüfer: Etschberger, Jansen, Wesp
 Studiengang: BW, IM
 Punkte: 15, 15, 5, 6, 14, 26, 9 ; Summe der Punkte: 90

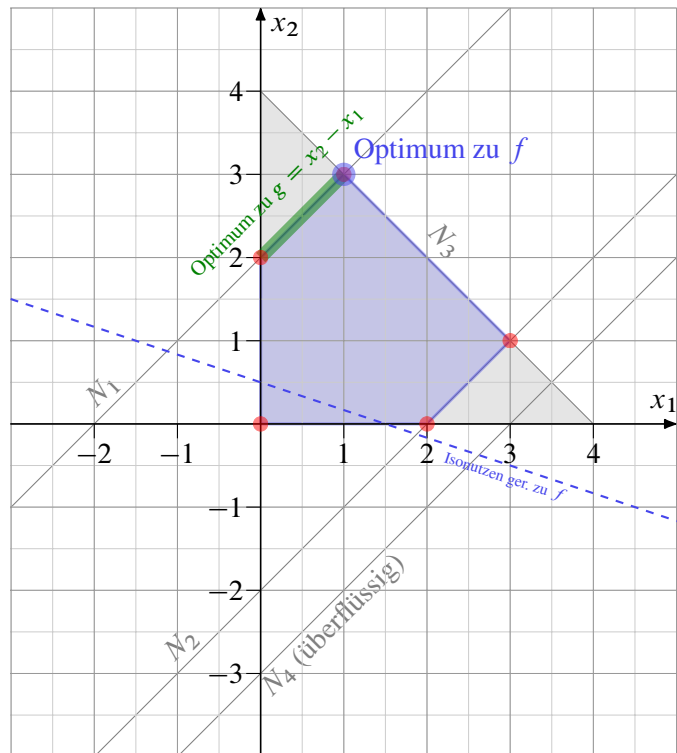
Aufgabe 1

15 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, \dots, N_4 (dabei ist N_3 nur teilweise gegeben).

f	x_1	+	$3x_2$	\rightarrow	max
N_1	$-x_1$	+	x_2	\leq	2
N_2	x_1	-	x_2	\leq	2
N_3	<input type="text" value="x1"/>	+	<input type="text" value="x2"/>	\leq	<input type="text" value="4"/>
N_4	x_1	-	x_2	\leq	3

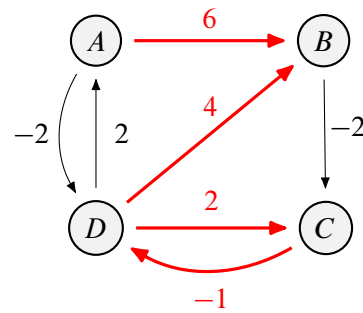


- a) Die graphische Repräsentation von N_3 ist im Koordinatensystem rechts eingezeichnet. Füllen Sie in der Tabelle oben die fehlenden Felder () von N_3 aus.
- b) Zeichnen Sie N_1, N_2, N_4 sowie den Zulässigkeitsbereich des Problems in das Koordinatensystem ein.
- c) Begründen Sie, ob und ggf. welche Nebenbedingungen zu vernachlässigen sind.
- d) Bestimmen Sie graphisch (Berechnung der Schnittpunkte ist nötig) die optimale Lösung des linearen Optimierungsproblems.
- e) Geben Sie den Funktionsterm $g(x_1, x_2)$ einer alternativen Zielfunktion an, so dass alle Punkte des zulässigen Bereichs optimal sind, welche auf dem Rand von N_1 liegen.

Lösungshinweis:

- a) siehe Zeichnung
- b) s. Z.
- c) N_3 ist eine Verschärfung von N_4 , die man damit vernachlässigen kann.
- d) Optimal: $(x_1, x_2) = (1, 3)$ mit Zielfunktionswert $f(1, 3) = 10$.
- e) siehe Zeichnung.

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



- a) Gegeben ist dazu die unvollständige Zeichnung eines Graphen und die dazugehörige ebenfalls jeweils unvollständige Kosten- bzw. Vorgängermatrix C_0 bzw. P_0 .

$$C_0 = \begin{pmatrix} 0 & 6 & \infty & -2 \\ \infty & 0 & -2 & \infty \\ \infty & \infty & 0 & -1 \\ 2 & 4 & 2 & 0 \end{pmatrix}, \quad P_0 = \begin{pmatrix} A & A & 0 & A \\ 0 & B & B & 0 \\ 0 & 0 & C & C \\ D & D & D & D \end{pmatrix}$$

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

- b) Der vorletzte Iterationsschritt wurde bereits durchgeführt und Sie haben C_3 vorliegen. Führen Sie nun den letzten Iterationsschritt durch, indem Sie die Matrix C_4 an den fehlenden Stellen ergänzen.

$$C_3 = \begin{pmatrix} 0 & 6 & 4 & -2 \\ \infty & 0 & -2 & -3 \\ \infty & \infty & 0 & -1 \\ 2 & 4 & 2 & 0 \end{pmatrix}, \quad C_4 = \begin{pmatrix} 0 & 2 & 0 & -2 \\ -1 & 0 & 3 & -3 \\ 1 & 3 & 0 & -1 \\ 2 & 4 & 2 & 0 \end{pmatrix}$$

- c) Gegeben ist jetzt zu einem *neuen* Graph mit 5 Knoten die Vorgängermatrix P_5 nach der letzten Iteration des Algorithmus mit

$$P_5 = \begin{pmatrix} A & B & D & C & A \\ B & A & C & E & D \\ C & D & C & E & D \\ B & E & D & D & C \\ E & B & A & C & E \end{pmatrix}.$$

Dabei ist bekannt, dass die Zeilen zu Beginn alphabetisch aufsteigend mit den Bezeichnungen der Knoten von A bis E angeordnet waren. Geben Sie dazu den kürzesten Weg vom Knoten D nach A an.

Lösungshinweis:

- a) s.o.
- b) s.o.
- c) Der kürzeste Weg ist $D \rightarrow C \rightarrow E \rightarrow B \rightarrow A$.

Aufgabe 3

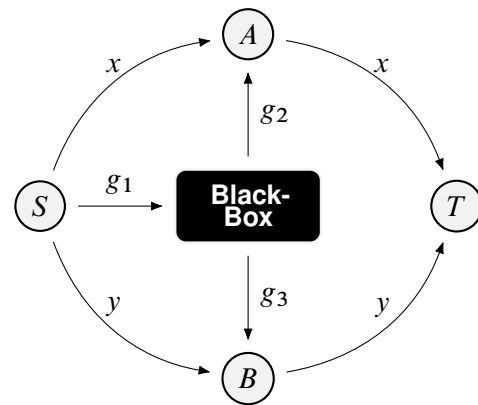
5 Punkte

Auf den nebenstehenden Graph wurde der Algorithmus von Dijkstra angewendet, wobei ein kürzester Weg von S nach T gesucht wird.

Alle Kanten des Graphen sind gerichtet. Die Gewichte g_1 , g_2 und g_3 sind unbekannt. Die Black-Box (BB) beschreibt einen Teil des Graphen, der nicht näher bekannt ist.

Außerdem ist bekannt, dass der kürzeste Weg von S

- ▶ nach A bzw. B stets über BB führt,
- ▶ nach A ein Drittel von y entspricht,
- ▶ nach B die Hälfte von x beträgt,
- ▶ nach T sowohl über A wie über B gleich ist.



Nach dem letzten Iterationsschritt steht nur noch (A, a) und (B, b) , mit $a, b \in \mathbb{N}$ in der Warteschlange.

Entscheiden Sie, ob Sie für $y = 9$ sagen können, welcher Knoten als nächstes laut Algorithmus abgearbeitet wird, oder begründen Sie, warum dies nicht möglich ist.

Lösungshinweis:

- ▶ Den kürzesten Weg von S nach A erhält man aus der Warteschlange, also zu a – analog für B .
- ▶ Laut Angabe gilt $\frac{x}{2} = b$ und $\frac{y}{3} = a$ und da der kürzeste Weg über beide gleich ist, gilt $a + x = b + y$ und damit $\frac{y}{3} + x = \frac{x}{2} + y \implies x = 4/3 \cdot y$.
- ▶ Mit $y = 9$ folgt damit $x = 4/3 \cdot 9 = 12$ und damit ist $b = 12/2 = 6$ und für $a = 9/3 = 3$. Damit gilt in der Warteschlange $(A,3)$ und $(B,6)$, also wäre A der nächste Knoten.

Aufgabe 4

6 Punkte

Eine Marketingberatung untersucht die Aufenthaltszeit von potentiellen Kunden auf Webshop-Seiten. Dazu wird die Verweildauer in Sekunden (Merkmal X) auf den Webshop-Seiten eines Unternehmens bei 600 Besuchern der Webseite analysiert und aufgezeichnet. Außerdem ist noch das Geschlecht dieser Leute bekannt (Merkmal Y). Die Beobachtungen (x_i, y_i) können als Ergebnis einer einfachen Stichprobe angesehen werden. Getrennt nach dem Geschlecht der Surfer ergibt sich folgende Häufigkeitstabelle für die Verweildauern (in Sekunden) und die Anzahl der Leute:

	Verweildauer auf Seite des Web-Shops									
	40	60	80	100	120	140	160	180	200	220
Frau	1	4	31	42	49	46	16	6	4	1
Mann	2	17	46	83	119	88	41	3	1	0

Getrennt nach Geschlecht bleiben die Untersuchten $\bar{x}_{\text{Frau}} = 119.7$ bzw. $\bar{x}_{\text{Mann}} = 117.45$ Sekunden auf der Webshop-Seite (muss nicht nachgerechnet werden).

Es soll mit einem χ^2 -Test zu einem Signifikanzniveau von 5 % getestet werden, ob das Geschlecht in der Grundgesamtheit abhängig von der Verweildauer ist. Dazu werden alle Beobachtungen von X in 5 Intervalle eingruppiert. Es ergibt sich folgende Kontingenztabelle:

Geschlecht	[40,76]	(76,112]	(112,148]	(148,184]	(184,220]
Frau	5	73	95	22	5
Mann	19	129	207	44	1

- Füllen Sie die Kontingenztabelle aus.
- Berechnen Sie den Summand der Testgröße v für die Frauen und die Dauerklasse [40,76].
- Für die Testgröße ergibt sich $v \approx 9.63$ (muss nicht nachgerechnet werden). Führen Sie den Test durch und leiten Sie daraus ab, ob die beiden Merkmale in der Grundgesamtheit unabhängig sind.
- Warum wäre eine Abhängigkeit der beiden Parameter kein Widerspruch zum fast identischen Wert von \bar{x}_{Frau} und \bar{x}_{Mann} ?

Lösungshinweis:

- Kontingenztabelle: s.o.
- bei Unabh. erwartete Häufigkeiten:

	[40,76]	(76,112]	(112,148]	(148,184]	(184,220]
Frau	8	67.33	100.67	22	2
Mann	16	134.67	201.33	44	4

Quadrierte Residuen:

	[40,76]	(76,112]	(112,148]	(148,184]	(184,220]
Frau	1.12	0.48	0.32	0	4.50
Mann	0.56	0.24	0.16	0	2.25

Daraus: Testwert $v \approx 9.6313$

χ^2 -Verteilung mit 4 Freiheitsgraden: $B = (9.4877; \infty)$

Also: Nullhypothese ablehnen, Dauer und Geschlecht sind abhängig.

- Auch wenn die Mittelwerte gleich sind, können die Verteilungen in den Dauersklassen höchst unterschiedlich sein. Hier: Streuung bei den Frauen viel höher.

Gegeben ist die $(1\,000\,000 \times 2)$ -Datenmatrix A mit zwei metrischen Merkmalen X_1, X_2 (Graphische Darstellung siehe Streuplot rechts). Die Varianz-Kovarianzmatrix S von A ergibt sich mit

$$S = \begin{pmatrix} 2.0824 & -1.4464 \\ -1.4464 & 2.9292 \end{pmatrix}.$$

Es soll eine Hauptkomponentenanalyse zu A durchgeführt werden.

- a) Berechnen Sie näherungsweise die Eigenwerte λ_1, λ_2 zu S .

(Hinweis: Falls Sie Teilaufgabe a) nicht lösen konnten, rechnen Sie bitte mit der (alternativen) Varianz-Kovarianz-matrix

$$S = \begin{pmatrix} 1.4704 & -1.6128 \\ -1.6128 & 6.5296 \end{pmatrix}$$

und den zugehörigen Eigenwerten $\lambda_1 = 7, \lambda_2 = 1$ weiter.)

- b) Wieviel Prozent der Information können in der ersten Hauptkomponente konzentriert werden?
 c) Berechnen Sie die Eigenvektoren von S .
 d) Zeichnen Sie mit den Eigenvektoren die Hauptachsen der Hauptkomponentenanalyse in die Grafik ein. Benutzen Sie dabei den Schwerpunkt der Daten $(x_1, x_2) = (5, 8)$ als Ursprung.

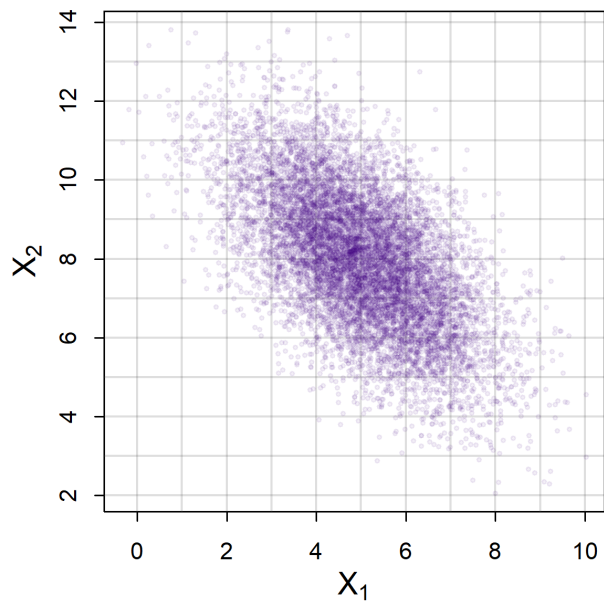


Abbildung 1: Streuplot von A

Lösungshinweis:

```
S
##           [,1]      [,2]
## [1,]  2.082418 -1.446406
## [2,] -1.446406  2.929195
F = eigen(S)$vectors
L = eigen(S)$values
F
##           [,1]      [,2]
## [1,] -0.5996126 -0.8002904
## [2,]  0.8002904 -0.5996126
L
## [1]  4.0129052  0.9987076
```

```
S.f
##           [,1]      [,2]
## [1,]  1.4704   -1.6128
## [2,] -1.6128   6.5296
F.f = eigen(S.f)$vectors
L.f = eigen(S.f)$values
F.f
##           [,1]      [,2]
## [1,] -0.28   -0.96
## [2,]  0.96   -0.28
L.f
## [1]  7  1
```

R

a) Gegeben ist ein Dataframe (daten) mit drei Spalten (nachname, vorname und geburtsjahr). Geben sie jeweils die R-Befehle an mit denen der dataframe wie jeweils beschrieben geändert wird:

1. Sie wollen eine weitere Zeile mit Ihren persönlichen Daten hinzufügen.

```
rbind(daten, data.frame(vorname='Hugo', name='Habicht', geburtsjahr=1904))
```

2. Sie wollen eine Spalte mit einem logischen Vektor hinzufügen, der genau dann TRUE ist falls die Person nach 1989 geboren ist. Die neue Spalte soll den Namen u30 haben.

```
cbind(daten, u30=daten$geburtsjahr>1989)
```

3. Die 42. Zeile sei nun die mit Ihrem persönlichen Eintrag. Geben Sie zwei Möglichkeiten an, wie Sie Ihren Nachnamen ändern können.

```
daten$nachname[42] = 'Meyer' oder daten[42,1] = 'Meyer'
```

4. Wie könnte man die Daten im Dataframe (erste Spalte, alphabetisch aufsteigend) sortieren?

```
daten[order(daten$nachname),]
```

5. Nachdem Sie alle obigen Änderungen gemacht haben, stellen Sie fest, dass Ihnen die Spaltenüberschriften nicht gefallen. Wie Ändern Sie diese in NAM, VOR, GEB und U30 um?

```
names(daten) <-c('NAM', 'VOR', 'GEB', 'U30')
```

b) Sie erzeugen einen numerischen Vektor mittels `x <- seq(1, 2, by = 0.1)`. Welche der folgenden Befehle erzeugt den gleichen Vektor?

	wahr	falsch
<code>c(1, 0.1, times = 10)</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<code>seq(1, 2, length = 10)</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<code>seq(1, 2, length.out = 11)</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<code>seq(to=2, from=1, 0.1)</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<code>rep(1, 0.1, times = 11)</code>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<code>10:20/10</code>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

c) Seien a und b zwei gleich lange numerische Vektoren der Länge n. Sie führen den Befehl `a!=b` aus. Was passiert?

	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"/>
dem Vektor b wird die Fakultät von a zugewiesen	<input type="checkbox"/>	<input checked="" type="checkbox"/>
das Ergebnis ist TRUE falls jeweils die Fakultät von a gleich b ist, sonst FALSE	<input type="checkbox"/>	<input checked="" type="checkbox"/>

d) Welche Aussagen bezüglich Faktoren in R sind korrekt?

	wahr	falsch
Faktoren benötigt man für ordinale Merkmale	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Faktoren benötigt man für kardinale Merkmale	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Faktoren benötigt man für nominale Merkmale	<input checked="" type="checkbox"/>	<input type="checkbox"/>
in Data Frames sind Zeichenketten standardmäßig Faktoren	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<code>is.factor()</code> wandelt numerische Werte in Faktoren um	<input type="checkbox"/>	<input checked="" type="checkbox"/>
es gibt keine Faktoren in R	<input type="checkbox"/>	<input checked="" type="checkbox"/>

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 <- 2 < x

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

f) 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. 2). Wie berechnen Sie mit diesem Vorwissen und der Hilfe von R die Wahrscheinlichkeit dafür, dass in einer Gruppe von 250 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 2: Hilfe zur R-Funktion `pbirthday()`



g) Was versteht man unter der *Monte-Carlo-Methode*? Geben Sie *kein* Beispiel aus der Vorlesung an, sondern erläutern Sie die zugrunde liegende Idee.



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) | is.numeric(x)) {
    x <- as.numeric(x)
  } else {x <- seq(2,8,2)}
  rval <- cumsum(x)
  return(rval)
}

x <- c(3,4,5,6)
y <- c("A", "B", "C")
z <- c(TRUE, FALSE, TRUE, TRUE)
```

Hinweis:

```
cumsum(c(2,3,5,7,11,13))
## [1] 2 5 10 17 28 41
```

Geben Sie jeweils die Ausgabe der folgenden Funktionsaufrufe an.

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

```
test(x)
## [1] 3 7 12 18
```

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

```
test(y)
## [1] 2 6 12 20
```

```
test(x=z)
## [1] 1 1 2 3
```

```
test2
## Error in eval(expr, envir, enclos):
Objekt 'test2' nicht gefunden
```

Lösungshinweis: