Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Fachbereich Informatik
Sommersemester 2024
Prof. Dr. Peter Becker

Mathematisch-algorithmische Grundlagen für Data Science

Aufgabenblatt 8

Abgabe zu **zweit** vor der Vorlesung am 4. Juni 2024.

Sollpunktzahl: 18 Punkte

Hinweis: Zur Lösung der Aufgaben auf diesem Übungsblatt benötigen Sie einen LP-Solver, z. B. das GNU Linear Programming Kit (GLPK). Es bieten sich nun die folgenden Möglichkeiten an:

- Sie installieren sich das GLPK auf Ihrem Rechner. Auf gängigen Linux-Systemen ist dies über das Paket-Management sehr einfach durchführbar. Beachten Sie hierfür die auf der Homepage verlinkten Folien zum GLPK.

  Falls Sie einen Mac nutzen, so empfehle ich, Homebrew

  `https://formulae.brew.sh/formula/glpk#default`

  oder MacPorts

  `https://ports.macports.org/port/glpk/`

  zu verwenden.

  Falls Sie einen Windows-Rechner nutzen, so schauen Sie mal unter

  `https://winglpk.sourceforge.net/`

  oder installieren sich eine virtuelle Linux-Maschine.

- Für die Python-Fans: `scipy` hat einen eingebauten LP-Solver, siehe
  `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.linprog.html`

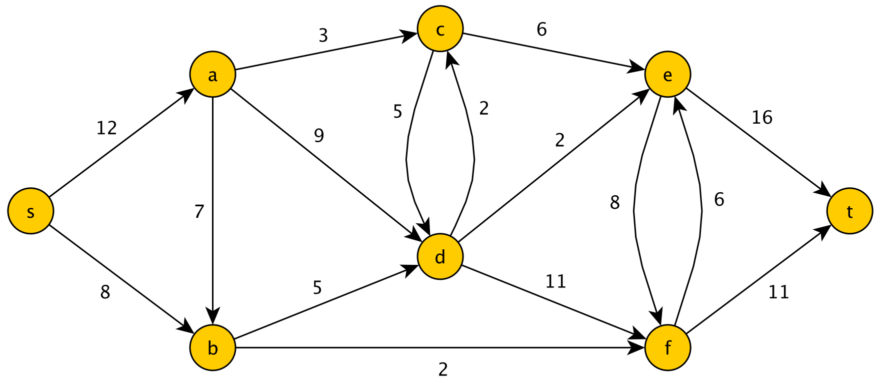
  Die Nutzung hatten wir in der Vorlesung besprochen (siehe Folien zum GLPK).

- Es gibt verschiedene Web-Seiten, auf denen man ein LP online lösen kann, z. B.
  `https://online-optimizer.appspot.com/`

  Das Format zur Eingabe ist zwar leicht unterschiedlich zum GLPK, aber mit den gegebenen Beispielen sollte es selbsterklärend sein. Intern wird wieder das GLPK genutzt.

Aufgabe 1 (Maximalfluss und minimaler Schnitt) — 4+4=8 Punkte

Wir wollen das folgende Flussnetzwerk analysieren und hierzu den minimalen Schnitt des Netzwerks bestimmen.



(a) Erstellen Sie zur Lösung des Maximalflussproblems ein LP und lösen Sie dieses. Wie lautet der Wert des maximalen Flusses? Geben Sie auch den Maximalfluss konkret an.

(b) Bestimmen Sie mittels der Lösung aus (a) einen minimalen Schnitt. Geben Sie alle Kanten an, die zum minimalen Schnitt gehören.

Aufgabe 2 (Ausgleichsrechnung mit $\|\cdot\|_1$- und $\|\cdot\|_\infty$-Norm) 4+4+4+4+4=20 Punkte

(a) Bestimmen Sie für die Daten von Aufgabe 2, Aufgabenblatt 5 die Regressionsgerade $y = ax + b$, die durch das Fehlerfunktional $E_1(a, b) = \sum_{i=1}^{n} |ax_i + b - y_i|$ bestimmt ist.

(b) wie (a), aber für das Fehlerfunktional $E_\infty(a, b) = \max_{i=1}^{n} |ax_i + b - y_i|$.

(c) Bestimmen Sie für die Daten von Aufgabe 1, Aufgabenblatt 5 die Ausgleichsfunktion $y = ax^2 + bx + c$, die durch das Fehlerfunktional $E_1(a, b, c) = \sum_{i=1}^{n} |ax_i^2 + bx_i + c - y_i|$ bestimmt ist.

(d) wie (c), aber für das Fehlerfunktional $E_\infty(a, b, c) = \max_{i=1}^{n} |ax_i^2 + bx_i + c - y_i|$.

(e) Erstellen Sie jeweils einen Plot für die Aufgaben (a), (b) und die Aufgaben (c), (d). Zeichnen Sie in die Plots auch die Ausgleichsfunktionen für das E_2-Fehlerfunktional.

Aufgabe 3 (Trennende Hyperebene) 4+4=8 Punkte

(a) Gegeben sind die folgenden Daten:

x_i	y_i	Klasse
0.6	2.5	1
1.0	0.9	1
1.4	−2.9	2
1.5	1.4	1
2.5	2.5	1
2.6	−2.5	2
2.7	−3.9	2
2.9	3.5	1
3.0	1.1	1
3.5	−1.5	2
3.6	−2.4	2
4.1	−3.6	2
4.4	2.3	1

x_i	y_i	Klasse
5.2	−2.4	2
5.5	−1.4	2
5.6	1.6	1
5.8	−3.2	2
6.0	3.6	1
7.2	−1.2	2
7.5	2.4	1
7.6	−1.9	2
8.6	3.4	1
9.6	0.6	2
9.9	−1.3	2
10.6	2.5	1
11.1	1.7	2

Berechnen Sie eine Gerade, die die beiden Klassen trennt, so dass der vertikale Abstand der Klassenpunkte, die der Geraden am nächsten liegen, maximiert wird (vgl. Folie 284).

(b) Wir erweitern die Datenmenge von (a) um die folgenden Daten:

x_i	y_i	Klasse
1.4	−1.4	2
4.0	1.0	2
4.9	−1.0	1
7.2	1.5	2
8.6	0.5	1
9.9	0.7	1

Es ist jetzt nicht mehr möglich, die beiden Klassen linear zu trennen.

Berechnen Sie eine Gerade, so dass die Summe der vertikalen Abstände zur Geraden für die falsch klassifizierten Punkte minimal wird (vgl. Folie 286).