

Multivariate Statistik, Übung 7

HENRY HAUSTEIN

Aufgabe 1

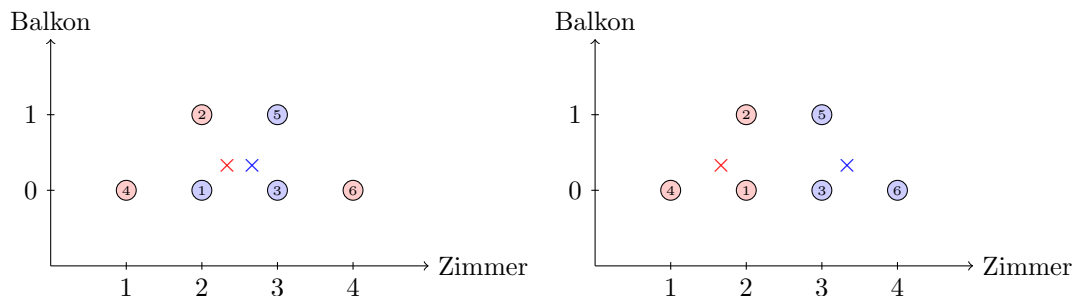
Ich nehme hier als Ähnlichkeitsmaß $s(x, y) = 1 - \frac{d(x, y)}{d_{max}}$, mit $d(x, y)$ als Manhattan-Distanz, weil es schnell berechenbar ist und nur schöne Zahlen produziert. Offensichtlich ist $d_{max} = 1 + 3$ (maximal ein Balkon und 3 Zimmer Unterschied zwischen den Wohnungen). Für die Distanz- und Ähnlichkeitsmatrix ergibt sich dann

$$D = \begin{pmatrix} 0 & 1 & 1 & 1 & 2 & 2 \\ & 0 & 2 & 2 & 1 & 3 \\ & & 0 & 2 & 1 & 1 \\ & & & 0 & 3 & 3 \\ & & & & 0 & 2 \\ & & & & & 0 \end{pmatrix} \quad S = \begin{pmatrix} 1 & 0.75 & 0.75 & 0.75 & 0.5 & 0.5 \\ & 1 & 0.5 & 0.5 & 0.75 & 0.25 \\ & & 1 & 0.5 & 0.75 & 0.75 \\ & & & 1 & 0.25 & 0.25 \\ & & & & 1 & 0.5 \\ & & & & & 1 \end{pmatrix}$$

Aufgabe 2

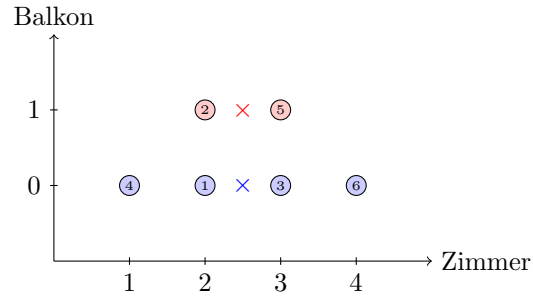
- (a) Ablauf des Algorithmus: C_1 und C_2 stehen für die Cluster 1 und 2, \bar{C}_1 und \bar{C}_2 für die Mittelwerte der Cluster 1 und 2

Iteration	C_1	C_2	\bar{C}_1	\bar{C}_2
0	1,3,5	2,4,6	$(\frac{8}{3}, \frac{1}{3})$	$(\frac{7}{3}, \frac{1}{3})$
1	3,5,6	1,2,4	$(\frac{10}{3}, \frac{1}{3})$	$(\frac{5}{3}, \frac{1}{3})$



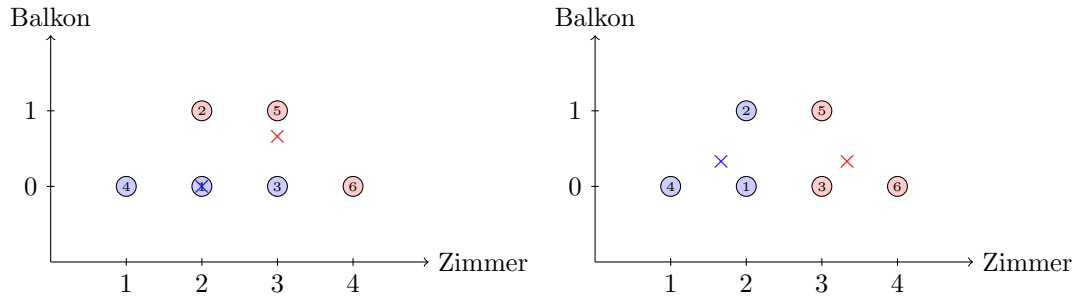
- (b) Ablauf des Algorithmus: C_1 und C_2 stehen für die Cluster 1 und 2, \bar{C}_1 und \bar{C}_2 für die Mittelwerte der Cluster 1 und 2

Iteration	C_1	C_2	\bar{C}_1	\bar{C}_2
0	1,3,4,6	2,5	$(\frac{5}{2}, 0)$	$(\frac{5}{2}, 1)$



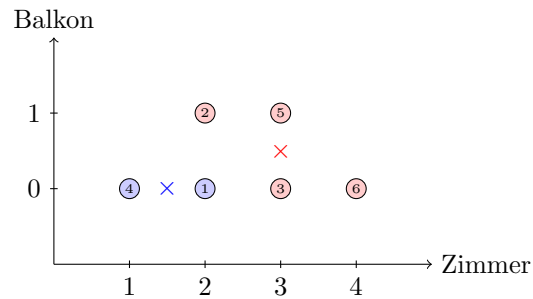
(c) Ablauf des Algorithmus: C_1 und C_2 stehen für die **Cluster 1** und **2**, \bar{C}_1 und \bar{C}_2 für die Mittelwerte der Cluster 1 und 2

Iteration	C_1	C_2	\bar{C}_1	\bar{C}_2
0	1,3,4	2,5,6	$(2, 0)$	$(3, \frac{2}{3})$
1	1,2,4	3,5,6	$(\frac{5}{3}, \frac{1}{3})$	$(\frac{10}{3}, \frac{1}{3})$



(d) Ablauf des Algorithmus: C_1 und C_2 stehen für die **Cluster 1** und **2**, \bar{C}_1 und \bar{C}_2 für die Mittelwerte der Cluster 1 und 2

Iteration	C_1	C_2	\bar{C}_1	\bar{C}_2
0	1,4	2,3,5,6	$(\frac{3}{2}, 0)$	$(3, \frac{1}{2})$



Aufgabe 3

Als allererstes Verändern wir unsere Datenmatrix, dass sie den Anforderungen der Aufgabenstellung entspricht:

Wohnung	Merkmale				
	x_1	x_2	x_3	x_4	x_5
1	1	1	0	0	0
2	1	1	0	0	1
3	1	1	1	0	0
4	1	0	0	0	0
5	1	1	1	0	1
6	1	1	1	1	0

(a) Wir starten im ersten Schritt mit 6 Clustern und berechnen zu jedem die Ähnlichkeit

	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{5\}$	$\{6\}$
$\{1\}$	1	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{3}{5}$	$\frac{3}{5}$
$\{2\}$		1	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{2}{5}$
$\{3\}$			1	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{4}{5}$
$\{4\}$				1	$\frac{2}{5}$	$\frac{2}{5}$
$\{5\}$					1	$\frac{3}{5}$
$\{6\}$						1

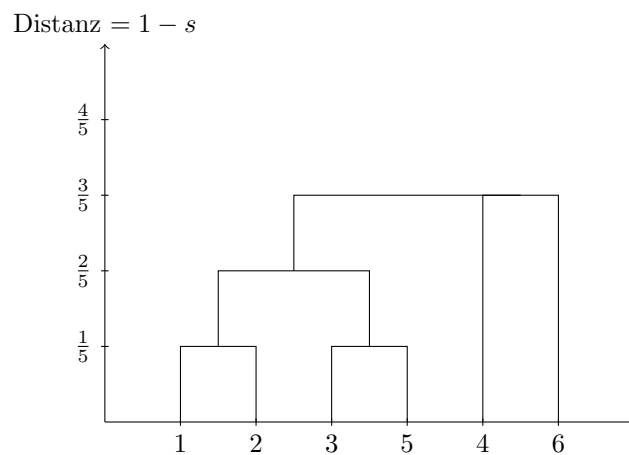
Jetzt bilden wir neue Cluster:

	$\{1, 2\}$	$\{3, 5\}$	$\{4, 6\}$
$\{1, 2\}$	1	$\frac{3}{5}$	$\frac{2}{5}$
$\{3, 5\}$		1	$\frac{2}{5}$
$\{4, 6\}$			1

Und nun werden noch einmal neue Cluster gebildet:

	$\{1, 2, 3, 5\}$	$\{4, 6\}$
$\{1, 2, 3, 5\}$	1	$\frac{2}{5}$
$\{4, 6\}$		1

Im letzten Schritt vereinigen wir nun auch noch diese beiden Cluster und erhalten $\{1, 2, 3, 4, 5, 6\}$. Das Dendrogramm ist



(b) Der erste Schritt ist hier identisch, für den zweiten Schritt ergibt sich

	$\{1, 2\}$	$\{3, 5\}$	$\{4, 6\}$
$\{1, 2\}$	1	$\frac{4}{5}$	$\frac{4}{5}$
$\{3, 5\}$		1	$\frac{4}{5}$
$\{4, 6\}$			1

Und nun werden noch einmal neue Cluster gebildet:

	$\{1, 2, 3, 5\}$	$\{4, 6\}$
$\{1, 2, 3, 5\}$	1	$\frac{4}{5}$
$\{4, 6\}$		1

Im letzten Schritt vereinigen wir nun auch noch diese beiden Cluster und erhalten $\{1, 2, 3, 4, 5, 6\}$. Das Dendrogramm ist

