5. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Annika Burkowitz, Sebastian Bange, Alexander Harnisch

WS 2017/2018
Prof. W. Rhode

Aufgabe 16

Die Zahlenwerte hängen von der genauen Gestalt der Populationen ab und werden von unserem Programm ausgegeben.

d)

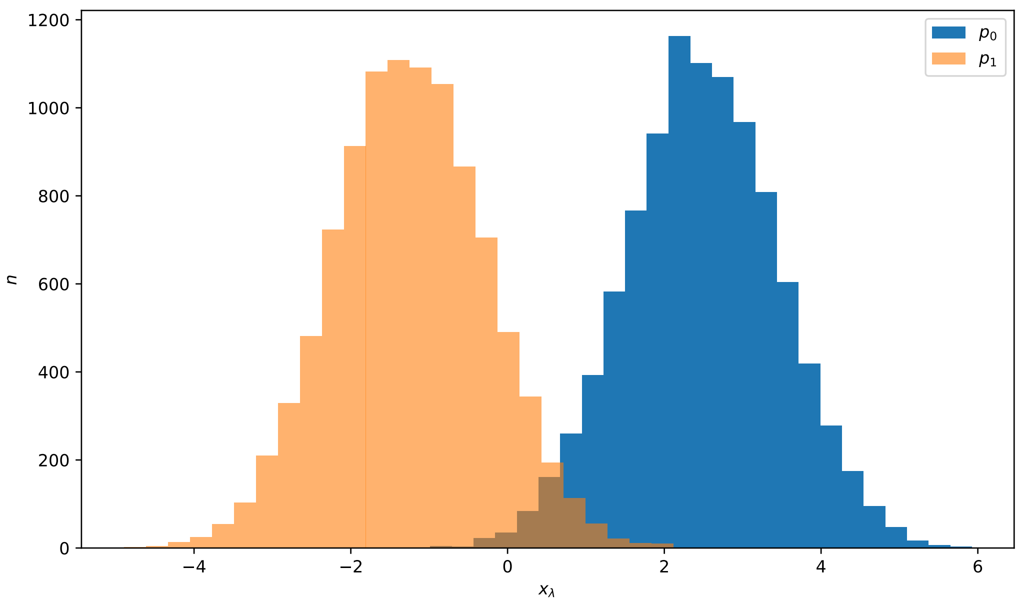

Abbildung 1: Histogramm der Projektion für $P_{0,10000}$.

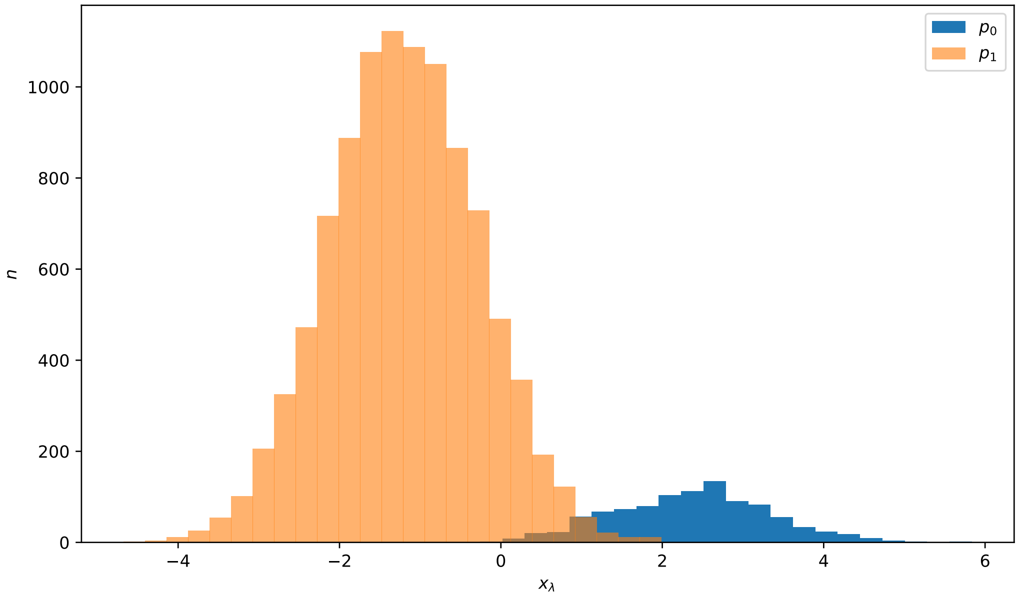

Abbildung 2: Histogramm der Projektion für $P_{0,1000}$.

e)

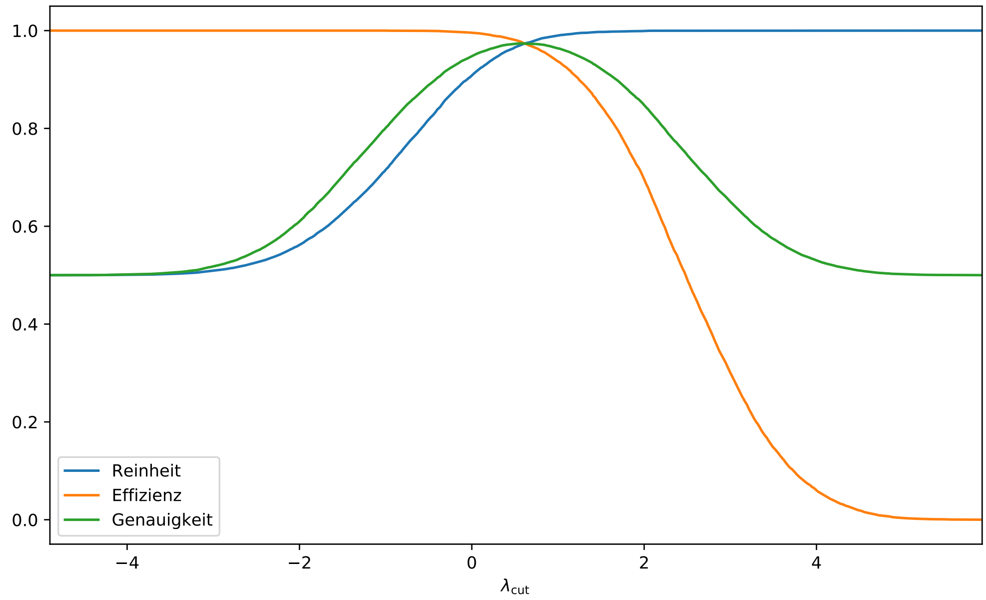

Abbildung 3: Effizienz, Reinheit und Genauigkeit für $P_{0,10000}$.

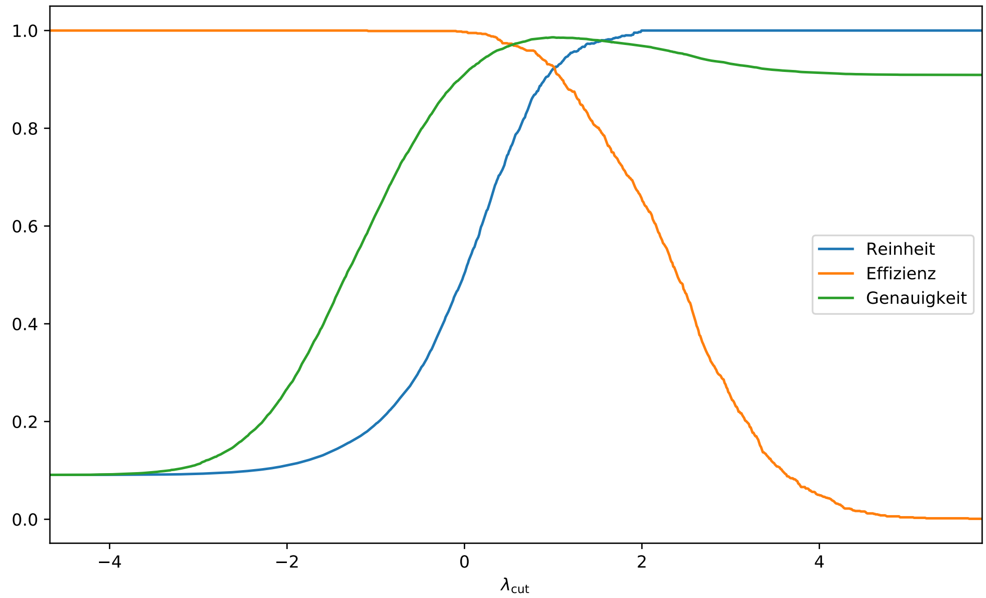

Abbildung 4: Effizienz, Reinheit und Genauigkeit für $P_{0,1000}$.

f)

Das Signal-zu-Untergrundverhältnis $\frac{S}{B}$ wird maximal für $B \to 0$ also für $\lambda_{\text{cut}} \to -\infty$.

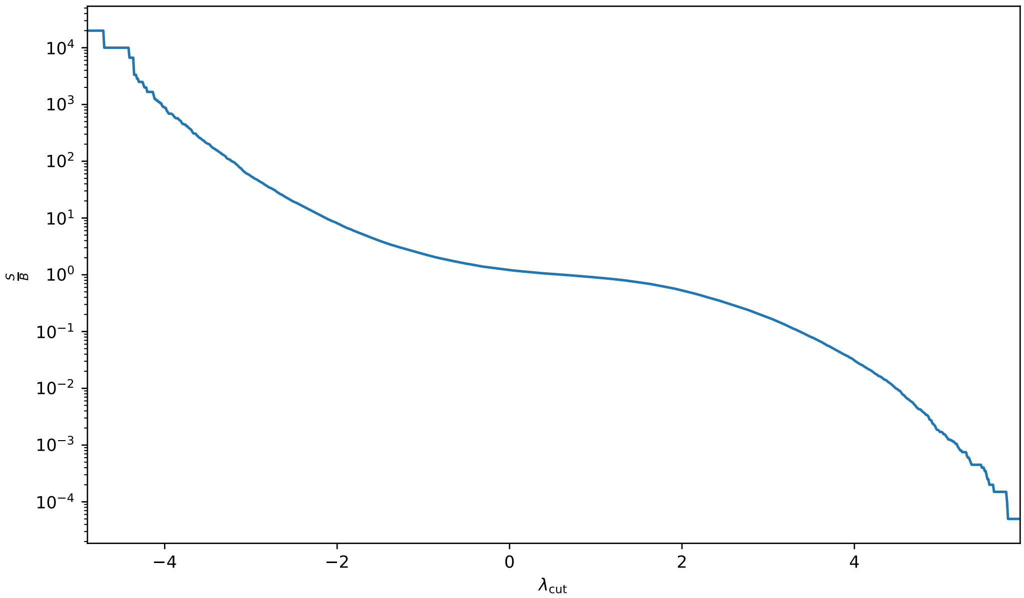

Abbildung 5: Signal-zu-Untergrundverhältnis für $P_{0,10000}$.

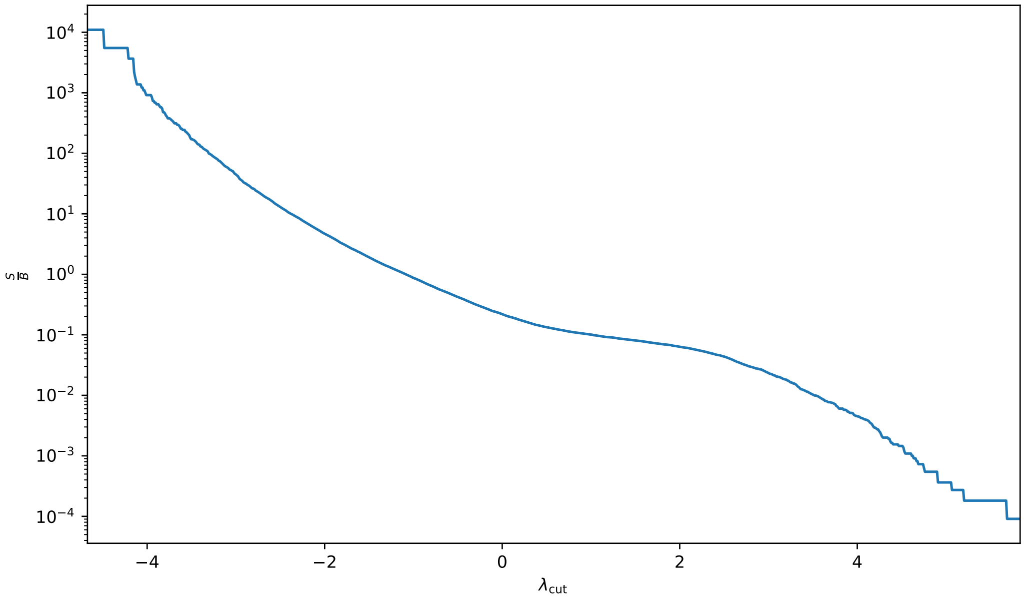

Abbildung 6: Signal-zu-Untergrundverhältnis für $P_{0,1000}$.

g)

Die Signifikanz $\frac{S}{\sqrt{S+B}}$ wird maximal für $S + B \to 0$ also für $\lambda_{\mathrm{cut}} \to -\infty$.

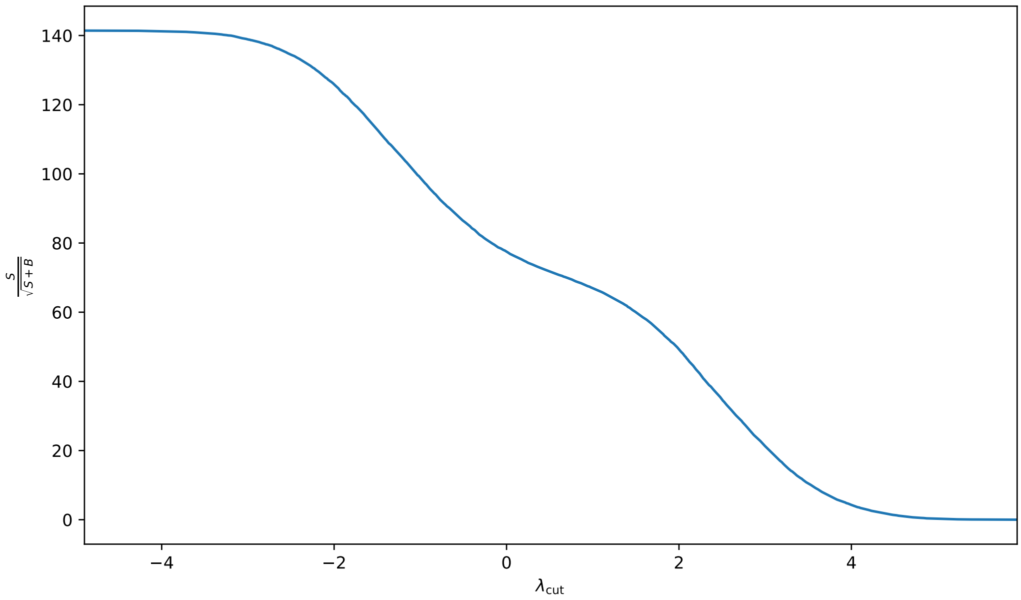

Abbildung 7: Signifikanz für $P_{0,10000}$.

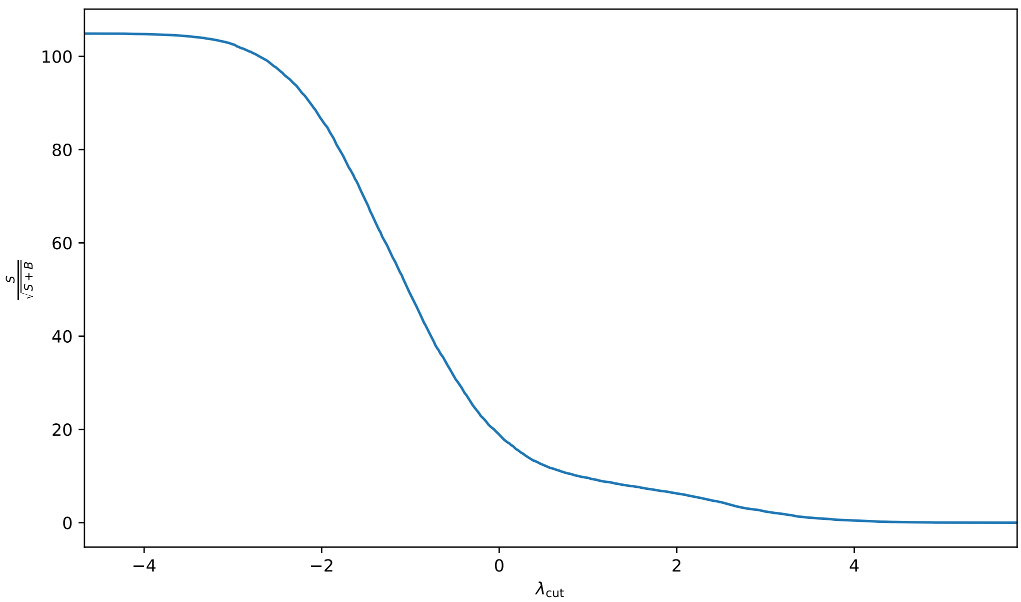

Abbildung 8: Signifikanz für $P_{0,1000}$.

Aufgabe 17: kMeans per Hand

Population: $\vec{x}_1 = \begin{pmatrix}1\\4\end{pmatrix}$, $\vec{x}_2 = \begin{pmatrix}1\\5\end{pmatrix}$, $\vec{x}_3 = \begin{pmatrix}1\\6\end{pmatrix}$, $\vec{x}_4 = \begin{pmatrix}3\\3\end{pmatrix}$, $\vec{x}_5 = \begin{pmatrix}3\\2\end{pmatrix}$

$\vec{x}_6 = \begin{pmatrix}4\\1\end{pmatrix}$, $\vec{x}_7 = \begin{pmatrix}5\\1\end{pmatrix}$, $\vec{x}_8 = \begin{pmatrix}6\\2\end{pmatrix}$, $\vec{x}_9 = \begin{pmatrix}6\\3\end{pmatrix}$, $\vec{x}_{10} = \begin{pmatrix}8\\4\end{pmatrix}$

$\vec{x}_{11} = \begin{pmatrix}8\\5\end{pmatrix}$, $\vec{x}_{12} = \begin{pmatrix}8\\6\end{pmatrix}$

a) Start-Clusterzentren: $S_1^{(0)} = \begin{pmatrix}3\\7\end{pmatrix}$, $S_2^{(0)} = \begin{pmatrix}3\\4\end{pmatrix}$, $S_3^{(0)} = \begin{pmatrix}7\\4\end{pmatrix}$

Punkte zuordnen:

Cluster 1: $\{\vec{x}_3\}$

Cluster 2: $\{\vec{x}_1, \vec{x}_2, \vec{x}_4, \vec{x}_5, \vec{x}_6\}$

Cluster 3: $\{\vec{x}_7, \vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

Alle eindeutig zuordenbar bis auf $\vec{x}_7$:

$$\|\vec{x}_7 - S_2^{(0)}\|^2 = \left\|\begin{pmatrix}5\\1\end{pmatrix} - \begin{pmatrix}3\\4\end{pmatrix}\right\|^2 = 2^2 + (-3)^2 = 13$$
$$\|\vec{x}_7 - S_3^{(0)}\|^2 = \left\|\begin{pmatrix}5\\1\end{pmatrix} - \begin{pmatrix}7\\4\end{pmatrix}\right\|^2 = (-2)^2 + 3^2 = 13$$

identisch → suche einen Cluster aus ⇒ Cluster 3

1. Iteration:

Neue Clusterzentren berechnen

$$S_1^{(1)} = \begin{pmatrix}1\\6\end{pmatrix}$$
$$S_2^{(1)} = \frac{1}{5}\left[\begin{pmatrix}1\\4\end{pmatrix} + \begin{pmatrix}1\\5\end{pmatrix} + \begin{pmatrix}3\\3\end{pmatrix} + \begin{pmatrix}3\\2\end{pmatrix} + \begin{pmatrix}4\\1\end{pmatrix}\right] = \frac{1}{5}\begin{pmatrix}12\\15\end{pmatrix} = \begin{pmatrix}2{,}4\\3\end{pmatrix}$$
$$S_3^{(1)} = \frac{1}{6}\left[\begin{pmatrix}5\\1\end{pmatrix} + \begin{pmatrix}6\\2\end{pmatrix} + \begin{pmatrix}6\\3\end{pmatrix} + \begin{pmatrix}8\\4\end{pmatrix} + \begin{pmatrix}8\\5\end{pmatrix} + \begin{pmatrix}8\\6\end{pmatrix}\right] = \frac{1}{6}\begin{pmatrix}41\\21\end{pmatrix} \approx \begin{pmatrix}6{,}833\\3{,}5\end{pmatrix}$$

Punkte neu zuordnen:

Cluster 1: $\{\vec{x}_2, \vec{x}_3\}$

Cluster 2: $\{\vec{x}_1, \vec{x}_4, \vec{x}_5, \vec{x}_6\}$

Cluster 3: $\{\vec{x}_7, \vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

$\vec{x}_7$ nicht offensichtlich zuordenbar:

$$\|\vec{x}_7 - S_2^{(1)}\|^2 = \left\|\begin{pmatrix}5\\1\end{pmatrix} - \begin{pmatrix}2{,}4\\3\end{pmatrix}\right\|^2 = 2{,}6^2 + (-2)^2 = 10{,}76$$
$$\|\vec{x}_7 - S_3^{(1)}\|^2 \approx \left\|\begin{pmatrix}5\\1\end{pmatrix} - \begin{pmatrix}6{,}833\\3{,}5\end{pmatrix}\right\|^2 \approx 3{,}361 + 6{,}25 = 9{,}611$$

$\vec{x}_7$ gehört zu Cluster 3

b) 2. Iteration:

$$S_1^{(2)} = \frac{1}{2}\left[\begin{pmatrix}1\\5\end{pmatrix} + \begin{pmatrix}1\\6\end{pmatrix}\right] = \begin{pmatrix}1\\5{,}5\end{pmatrix}$$
$$S_2^{(2)} = \frac{1}{4}\left[\begin{pmatrix}1\\4\end{pmatrix} + \begin{pmatrix}3\\3\end{pmatrix} + \begin{pmatrix}3\\2\end{pmatrix} + \begin{pmatrix}4\\1\end{pmatrix}\right] = \begin{pmatrix}2{,}75\\2{,}5\end{pmatrix}$$
$$S_3^{(2)} = S_3^{(1)} \approx \begin{pmatrix}6{,}833\\3{,}5\end{pmatrix}$$

$\vec{x}_7$ nicht eindeutig zuordenbar:

$$\|\vec{x}_7 - S_2^{(2)}\|^2 = \left\|\begin{pmatrix}5\\1\end{pmatrix} - \begin{pmatrix}2{,}75\\2{,}5\end{pmatrix}\right\|^2 = 2{,}25^2 + (-1{,}5)^2 = 7{,}3125$$
$$\|\vec{x}_7 - S_3^{(2)}\|^2 = 9{,}611 \quad \text{(s.ob.)}$$

⇒ $\vec{x}_7$ gehört zu Cluster 2

Cluster 1: $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$
Cluster 2: $\{\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7\}$
Cluster 3: $\{\vec{x}_8, \vec{x}_9, \vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

3. Iteration:
Neue Clusterzentren berechnen

$$S_1^{(3)} = \frac{1}{3}\left[\begin{pmatrix}1\\6\end{pmatrix}+\begin{pmatrix}1\\5\end{pmatrix}+\begin{pmatrix}1\\4\end{pmatrix}\right] = \begin{pmatrix}1\\5\end{pmatrix}$$

$$S_2^{(3)} = \frac{1}{4}\left[\begin{pmatrix}3\\3\end{pmatrix}+\begin{pmatrix}3\\2\end{pmatrix}+\begin{pmatrix}4\\1\end{pmatrix}+\begin{pmatrix}5\\1\end{pmatrix}\right] = \begin{pmatrix}3{,}75\\1{,}75\end{pmatrix}$$

$$S_3^{(3)} = \frac{1}{5}\left[\begin{pmatrix}6\\2\end{pmatrix}+\begin{pmatrix}6\\3\end{pmatrix}+\begin{pmatrix}8\\4\end{pmatrix}+\begin{pmatrix}8\\5\end{pmatrix}+\begin{pmatrix}8\\6\end{pmatrix}\right] = \begin{pmatrix}7{,}2\\4\end{pmatrix}$$

$\vec{x}_8$ nicht eindeutig zuordenbar

$$\left.\begin{array}{l} \|\vec{x}_8 - S_2^{(3)}\|^2 = \left\|\begin{pmatrix}6\\2\end{pmatrix}-\begin{pmatrix}3{,}75\\1{,}75\end{pmatrix}\right\|^2 = 2{,}25 + (-0{,}25)^2 = 5{,}125 \\ \|\vec{x}_8 - S_3^{(3)}\|^2 = \left\|\begin{pmatrix}6\\2\end{pmatrix}-\begin{pmatrix}7{,}2\\4\end{pmatrix}\right\|^2 = (-1{,}2)^2 + (-2)^2 = 5{,}44 \end{array}\right\} \rightarrow \vec{x}_8 \text{ gehört zu Cluster 2}$$

Cluster 1: $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$
Cluster 2: $\{\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8\}$
Cluster 3: $\{\vec{x}_9, \vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

4. Iteration:
Neue Clusterzentren berechnen

$$S_1^{(4)} = S_1^{(3)} = \begin{pmatrix}1\\5\end{pmatrix}$$

$$S_2^{(4)} = \frac{1}{5}\left[\begin{pmatrix}3\\3\end{pmatrix}+\begin{pmatrix}3\\2\end{pmatrix}+\begin{pmatrix}4\\1\end{pmatrix}+\begin{pmatrix}5\\1\end{pmatrix}+\begin{pmatrix}6\\2\end{pmatrix}\right] = \begin{pmatrix}4{,}2\\1{,}8\end{pmatrix}$$

$$S_3^{(4)} = \frac{1}{4}\left[\begin{pmatrix}6\\3\end{pmatrix}+\begin{pmatrix}8\\4\end{pmatrix}+\begin{pmatrix}8\\5\end{pmatrix}+\begin{pmatrix}8\\6\end{pmatrix}\right] = \begin{pmatrix}7{,}5\\4{,}5\end{pmatrix}$$

$\vec{x}_9$ nicht eindeutig zuordenbar:

$$\left.\begin{array}{l} \|\vec{x}_9 - S_2^{(4)}\|^2 = \left\|\begin{pmatrix}6\\3\end{pmatrix}-\begin{pmatrix}4{,}2\\1{,}8\end{pmatrix}\right\|^2 = 1{,}8^2 + 1{,}2^2 = 4{,}68 \\ \|\vec{x}_9 - S_3^{(4)}\|^2 = \left\|\begin{pmatrix}6\\3\end{pmatrix}-\begin{pmatrix}7{,}5\\4{,}5\end{pmatrix}\right\|^2 = (-1{,}5)^2 + (-1{,}5)^2 = 4{,}5 \end{array}\right\} \rightarrow \vec{x}_9 \text{ gehört zu Cluster 3}$$

Cluster 1: $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$
Cluster 2: $\{\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8\}$
Cluster 3: $\{\vec{x}_9, \vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

5. Iteration
→ keine Änderung mehr $\quad S_1^{(5)} = S_1^{(4)}, \quad S_2^{(5)} = S_2^{(4)}, \quad S_3^{(5)} = S_3^{(4)}$

c) Der Algorithmus konvergiert in diesem Fall nach der 4. Iteration.
Das Ergebnis entspricht nicht unseren Erwartungen. Unter der Annahme, dass es 3 Cluster gibt, hätten wir folgendes erwartet:

Cluster 1: $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, Cluster 2: $\{\vec{x}_4, \vec{x}_5, \vec{x}_6, \vec{x}_7, \vec{x}_8, \vec{x}_9\}$, Cluster 3: $\{\vec{x}_{10}, \vec{x}_{11}, \vec{x}_{12}\}$

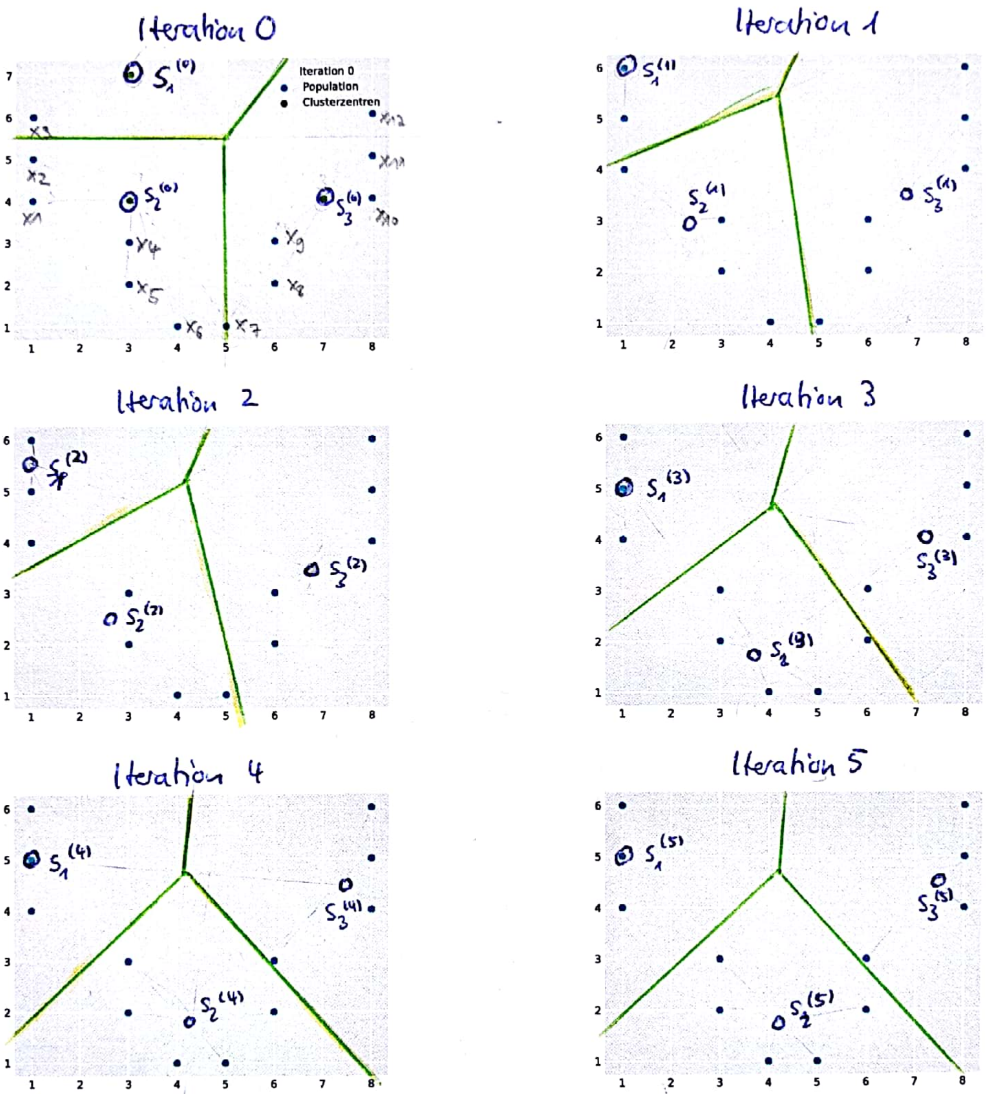

Abbildung 1: Population zum Einzeichnen der Clusterzentren und Clustergrenzen. Zu Aufgabe 17