5. Übungsblatt

Ksenia Klassen
ksenia.klassen@udo.edu

Dag-Björn Hering
dag.hering@udo.edu

Henning Ptaszyk
henning.ptaszyk@udo.edu

6. Dezember 2016

1 Aufgabe 1

1.1 a)

Zunächst wird für beide Populationen der Mittelwert berechnet.

$$\mu_{P0} \approx \begin{pmatrix} 0,022 \\ 3,026 \end{pmatrix} \tag{1}$$

$$\mu_{P1} \approx \begin{pmatrix} 5,986 \\ 3,026 \end{pmatrix} \tag{2}$$

1.2 b)

Die Kovantianzmatrix S_{P0} und S_{P1} berechnen sich über

$$S_j = \sum_i^{n_j} (\vec{x} - \vec{\mu}_j)(\vec{x} - \vec{\mu}_j)^T \tag{3}$$

zu

$$S_{P0} \approx \begin{pmatrix} 122904 & 81979 \\ 81979 & 67449 \end{pmatrix} \tag{4}$$

und

$$S_{P1} \approx \begin{pmatrix} 122344 & 73118 \\ 73118 & 53845 \end{pmatrix} \tag{5}$$

Aus $S_{P0P1} = S_{P0} + S_{P1}$ folgt

$$S_{P0P1} \approx \begin{pmatrix} 245248 & 155097 \\ 155097 & 121294 \end{pmatrix} \tag{6}$$

Desweiteren muss noch S_B über

$$S_B = (\vec{\mu}_1 - \vec{\mu}_2)(\vec{\mu}_1 - \vec{\mu}_2)^T \tag{7}$$

berechnet werden.

$$\Rightarrow S_B \approx \begin{pmatrix} 35.562 & 0.432 \\ 0.432 & 0.005 \end{pmatrix} \tag{8}$$

1.3 c)

Um nun die Fisher-Diskriminate zu berechen, müssen die Eigenwerte und Eigenvektoren der Matrix

$$A = S_{P0P1}^{-1} S_B \tag{9}$$

berechnet werden. Die Inverse von S_{P0P1} lautet:

$$S_{P0P1}^{-1} \approx \begin{pmatrix} 0.000021 & -0.000027 \\ -0.000027 & 0.000043 \end{pmatrix} \tag{10}$$

somit folgt für A

$$A \approx \begin{pmatrix} 0.000746 & 0.000009 \\ -0.000950 & -0.000012 \end{pmatrix} \tag{11}$$

Die Eigenwerte der Matrix A sind:

$$\lambda_1 \approx 0,0007 \qquad \lambda_2 \approx 0 \tag{12}$$

Die normierten Eigenvektoren sind:

$$\vec{v}_{\lambda_1} \approx \begin{pmatrix} 0,617 \\ -0,012 \end{pmatrix} \tag{13}$$

$$\vec{v}_{\lambda_2} \approx \begin{pmatrix} -0,787 \\ 1 \end{pmatrix} \tag{14}$$

Als Projektionsvektors wird der Eigenvektor zum Eigenwert λ_1 gewählt. Daraus folgt die Geradengleichung:

$$\vec{\lambda} = \lambda \cdot \begin{pmatrix} 0,617 \\ -0,012 \end{pmatrix} \tag{15}$$

1.4 d)

Nun kann mit Hilfe des Projektionsvektors $-v_{\lambda_1}$ ein eindimensionales Histogramm der Populationen 1 berechnete werden. Das Minus dreht die Orientierung so, dass $P0$ rechts von $P1$ liegt.

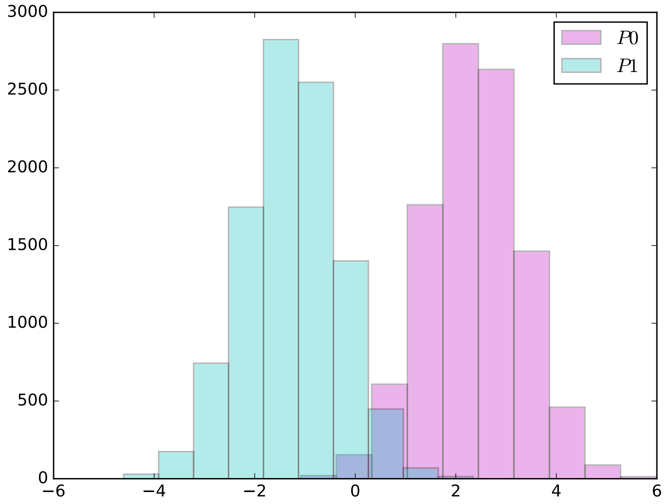


Abbildung 1: Histogramm der Projektion von den Populationen $P0$ und $P1$ auf die Projektionsgeraden $\vec{\lambda}$.

1.5 e)

Betrachtet man nun $P0$ als Signal und $P1$ als Untergrund kann die Effizienz und Reinheit des Signals als Funktion eines Schnitts λ_{cut} aufgetragen werden. Die folgende Abbildung 2 enthält die Effizienz und Reinheit in Abhängigkeit von λ_{cut}.

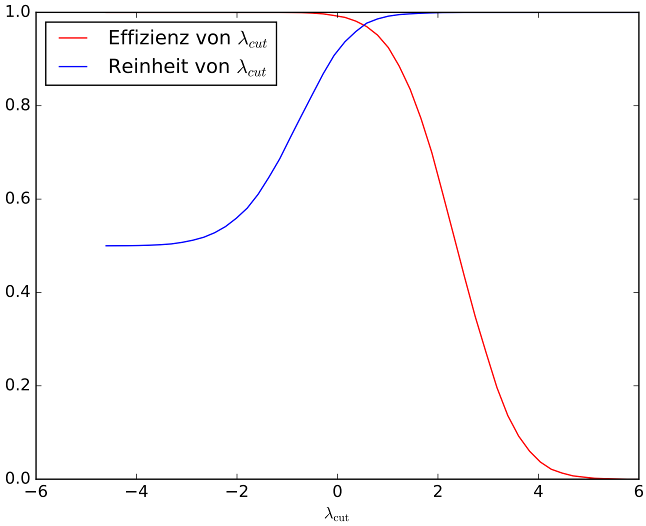


Abbildung 2: Effizienz und Reinheit des Schnittes ausgehend von der Gerade $\vec{\lambda}$ in Abhängigkeit von λ_{cut} .

1.6 f)

In der Abbildung 3 ist das Signal-zu-Untergrundverhältnis(t_p/f_p) in Abhängigkeit von dem Schnittes λ_{cut} aufgetragen.

Bei

$$\lambda_{cut} \approx 2,2 \tag{16}$$

wird das Signal-zu-Untergrundverhältis maximal.

1.7 g)

Ebenfalls kann die Signifikanz $t_p/\sqrt{t_p + f_p}$ in Abhängigkeit von dem Schnittes λ_{cut} aufgetragen werden 4. Diese besitzt bei

$$\lambda_{cut} \approx 0.58 \tag{17}$$

ein Maximum.

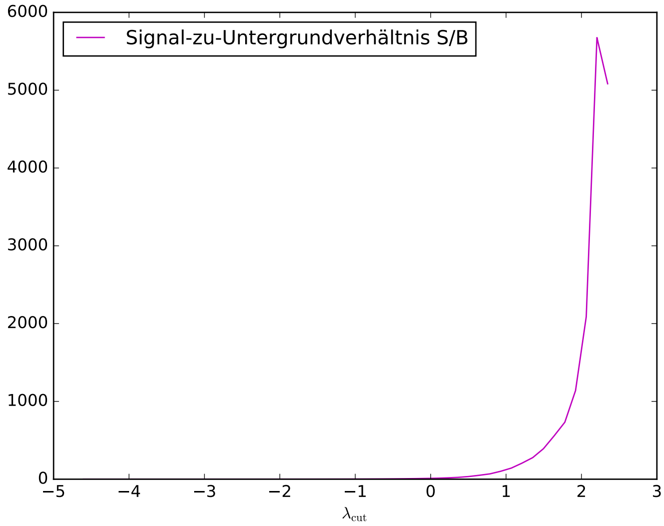


Abbildung 3: Das Signal-zu-Untergrundverhältnis in Abhängigkeit von λ_{cut} .

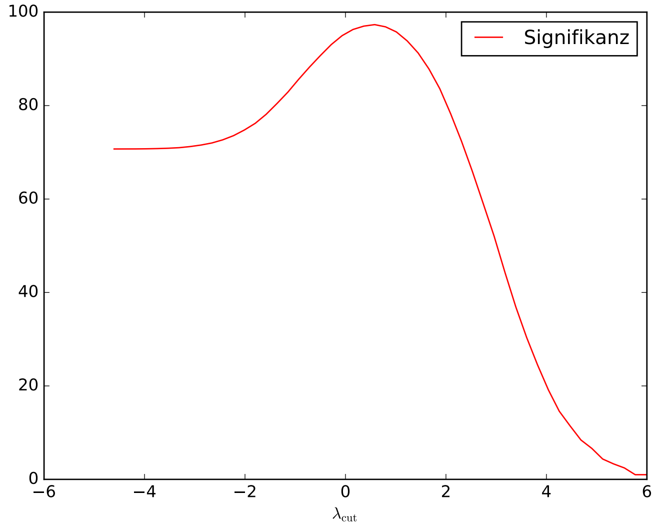


Abbildung 4: Die Signifikanz in Abhängigkeit von λ_{cut} .

1.8 h)

Die Schritte *e*) bis *g*) werden nun für den Fall, dass P0 nun die Population P_0_1000 bezeichnet wieder holt. Es ergibt sich die Projektion 5.

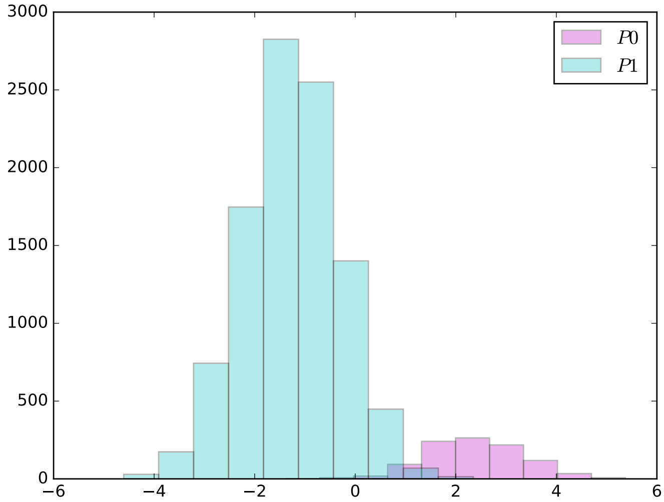


Abbildung 5: Histogramm der Projektion von den Populationen $P0$ und $P1$ auf die Projektionsgeraden $\vec{\lambda}$.

Wieder kann die Effizienz und Reinheit berechnet werden 6,
so wie das Signal-zu-Untergrundverhältnis 7, dass bei

$$\lambda_{cut} \approx 2,2 \tag{18}$$

ein Maximum besitzt. Die Signifikanz, die in der Abbildung 8 dagestellt ist, wird bei

$$\lambda_{cut} \approx 1,1 \tag{19}$$

maximal.

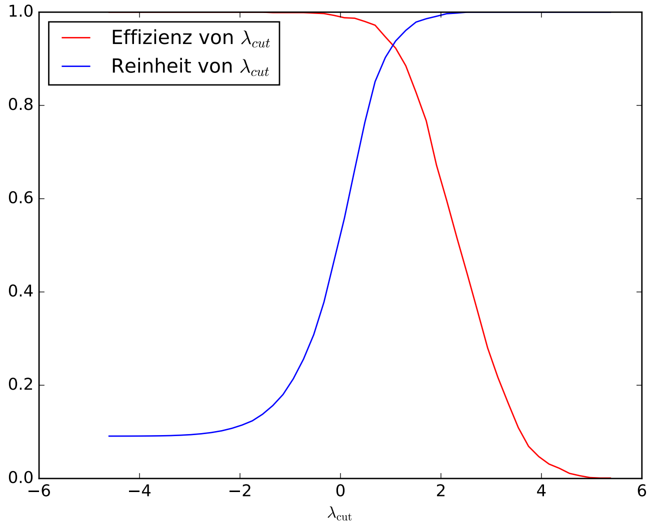


Abbildung 6: Effizienz und Reinheit des Schnittes ausgehend von der Gerade $\vec{\lambda}$ in Abhängigkeit von λ_{cut} .

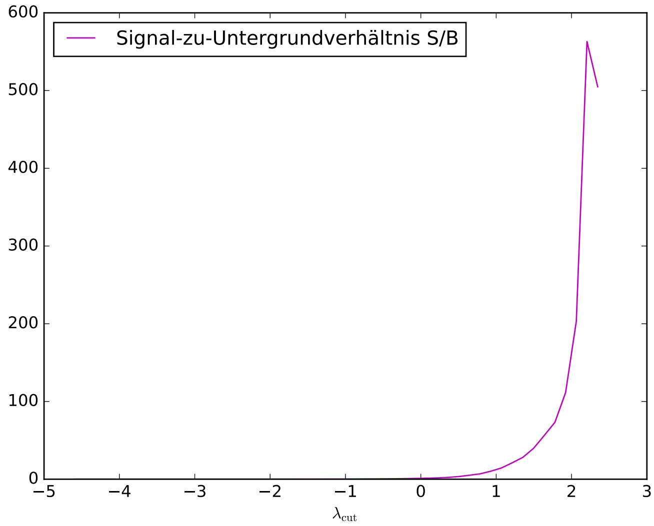


Abbildung 7: Das Signal-zu-Untergrundverhältnis in Abhängigkeit von λ_{cut} .

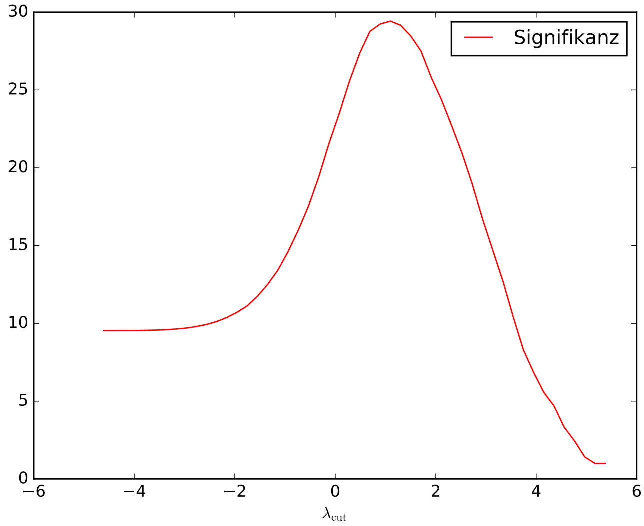
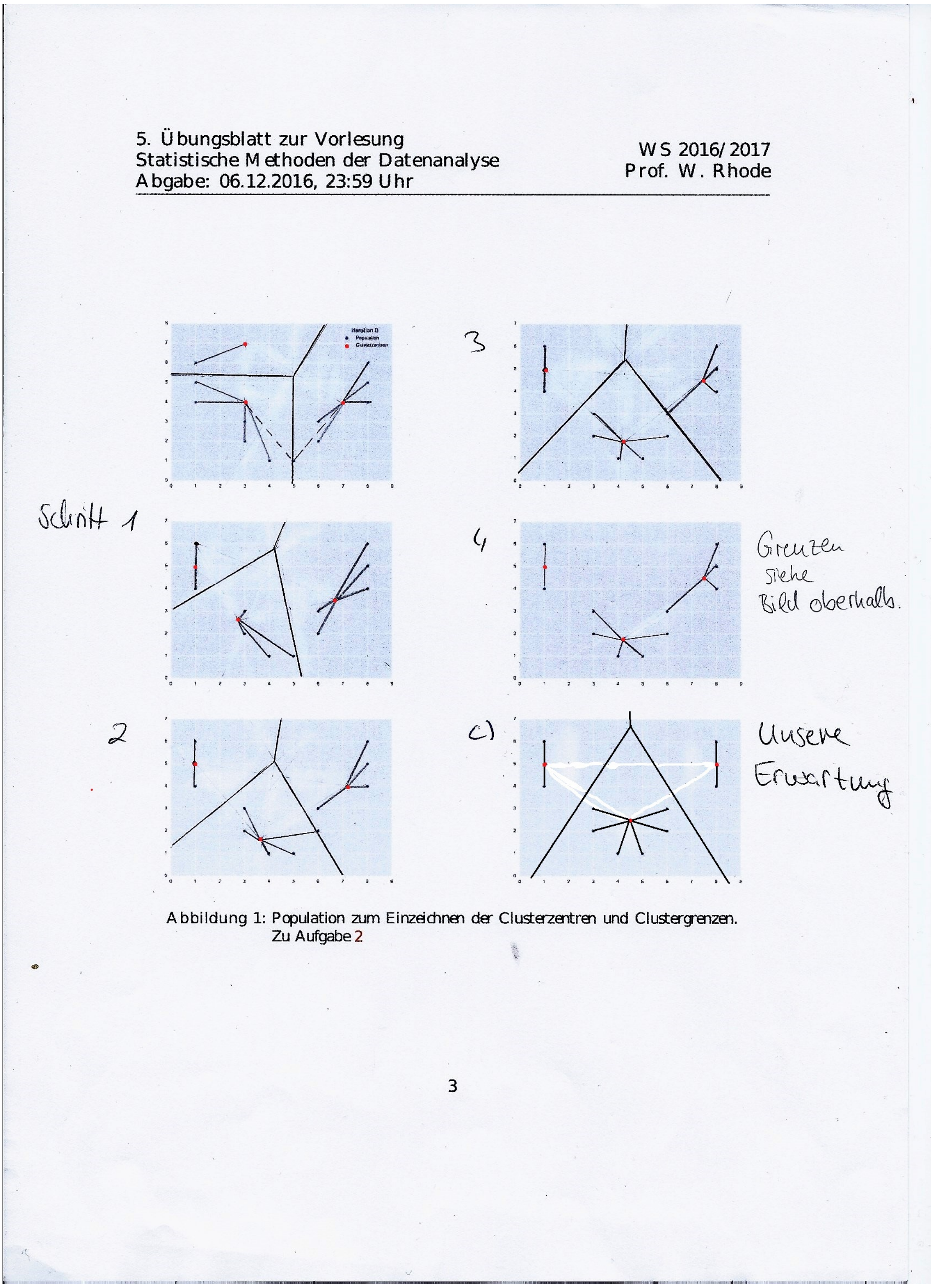


Abbildung 8: Die Signifikanz in Abhängigkeit von λ_{cut} .

2 Aufgabe2



5. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Abgabe: 06.12.2016, 23:59 Uhr

WS 2016/2017
Prof. W. Rhode

Abbildung 1: Population zum Einzeichnen der Clusterzentren und Clustergrenzen. Zu Aufgabe 2

3

Abbildung 9

.

a) Bestimmung des Zentrums mittels Mittelwertbildung:

Schritt 1: Zentrum 1: Zentrum 2: Zentrum 3:

Bsp: $x = \frac{1+1+3+3+4+5}{6} = \frac{17}{6}$ $z_2: \left(\frac{41}{6} ; \frac{7}{2}\right)$ $z_3: (1;6)$

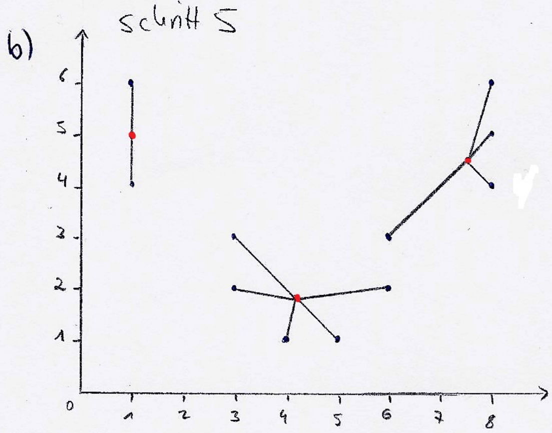
$y = \frac{8}{3}$ $\Rightarrow z_1: \left(\frac{17}{6}, \frac{8}{3}\right)$

Schritt 2: $z_1: \left(\frac{15}{4} ; \frac{7}{4}\right)$ $z_2: \left(\frac{36}{5} ; 4\right)$ $z_3: (1;5)$

Schritt 3: $z_1: \left(\frac{21}{5} ; \frac{9}{5}\right)$ $z_2: \left(\frac{15}{2} ; \frac{9}{2}\right)$ $z_3: (1;5)$

Schritt 4: $z_1: \left(\frac{21}{5} ; \frac{9}{5}\right)$ $z_2: \left(\frac{15}{2} ; \frac{9}{2}\right)$. $z_3: (1;5)$

b) Schritt 5



Nach vier weiteren Iterationsschritten, bekommt man das gleiche Bild.
Die Zentren liegen im gleichen Punkt.

c) Das Verfahren konvergiert nach drei Schritten.

Abbildung 10

.

3 Aufgabe3

Siehe beigelegte Ordner.