Übungsblatt 7

Julia Sobolewski, Lars Kolk und Jannine Salewski

Abgabe: 06.12.2018

TU Dortmund - Fakultät Physik

Aufgabe 19 Siehe zusätzlich Jupyter Notebook.

SMD - Blatt 7

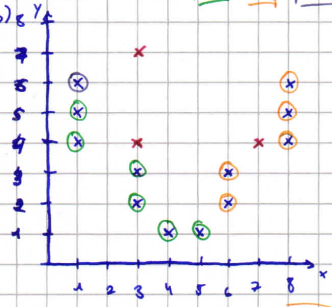
Aufgabe 19

Population: $\begin{pmatrix}1\\4\end{pmatrix}, \begin{pmatrix}1\\5\end{pmatrix}, \begin{pmatrix}1\\6\end{pmatrix}, \begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}3\\2\end{pmatrix}, \begin{pmatrix}4\\1\end{pmatrix}, \begin{pmatrix}5\\1\end{pmatrix}, \begin{pmatrix}6\\2\end{pmatrix}, \begin{pmatrix}6\\3\end{pmatrix}, \begin{pmatrix}8\\4\end{pmatrix}, \begin{pmatrix}8\\5\end{pmatrix}, \begin{pmatrix}8\\6\end{pmatrix}$

a)/ Clusterzentren: $\boxed{\begin{pmatrix}3\\4\end{pmatrix}}, \boxed{\begin{pmatrix}7\\4\end{pmatrix}}, \boxed{\begin{pmatrix}3\\7\end{pmatrix}}$

b)

0. Iteration

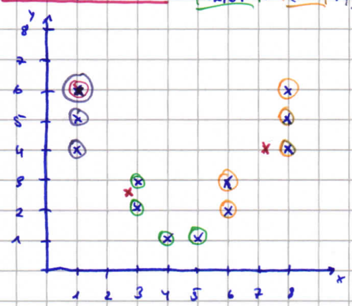


Abstände:	$\begin{pmatrix}5\\1\end{pmatrix}$
$\begin{pmatrix}3\\4\end{pmatrix}$	3,61
$\begin{pmatrix}7\\4\end{pmatrix}$	3,61

← $\begin{pmatrix}5\\1\end{pmatrix}$ wird $\begin{pmatrix}3\\4\end{pmatrix}$ zugeordnet

Neue Clusterzentren: $\boxed{\begin{pmatrix}2{,}83\\2{,}67\end{pmatrix}}, \boxed{\begin{pmatrix}7{,}2\\4\end{pmatrix}}, \boxed{\begin{pmatrix}1\\6\end{pmatrix}}$

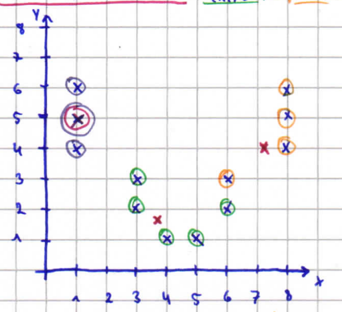
1. Iteration



Abstände	$\begin{pmatrix}1\\4\end{pmatrix}$	$\begin{pmatrix}5\\1\end{pmatrix}$
$\begin{pmatrix}1\\5\end{pmatrix}$	2	/
$\begin{pmatrix}2{,}83\\2{,}67\end{pmatrix}$	2,27	2,73
$\begin{pmatrix}7{,}2\\4\end{pmatrix}$	/	3,72

Neue Clusterzentren: $\boxed{\begin{pmatrix}3{,}75\\1{,}75\end{pmatrix}}, \boxed{\begin{pmatrix}7{,}2\\4\end{pmatrix}}, \boxed{\begin{pmatrix}1\\5\end{pmatrix}}$

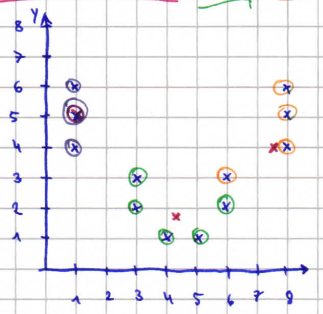
2. Iteration



Abstände	$\begin{pmatrix}6\\2\end{pmatrix}$
$\begin{pmatrix}3{,}75\\1{,}75\end{pmatrix}$	2,26
$\begin{pmatrix}7{,}2\\4\end{pmatrix}$	2,33

Neue Clusterzentren: $\boxed{\begin{pmatrix}4{,}2\\1{,}8\end{pmatrix}}, \boxed{\begin{pmatrix}7{,}5\\4{,}5\end{pmatrix}}, \boxed{\begin{pmatrix}1\\5\end{pmatrix}}$

3. Iteration



Abstände:	$\begin{pmatrix}6\\3\end{pmatrix}$
$\begin{pmatrix}4{,}2\\1{,}8\end{pmatrix}$	2,16
$\begin{pmatrix}7{,}5\\4{,}5\end{pmatrix}$	2,12

c) Nach der 3. Iteration werden sich die Clusterzentren nicht weiter ändern → Algorithmus ist konvergiert
Das Ergebnis entspricht ~~nicht~~ ganz den Erwartungen, da man intuitiv Punkt $\begin{pmatrix}6\\3\end{pmatrix}$ dem grünen Cluster zuordnen würde.

Aufgabe 20

a) Die Lossfunktion ist ein Maßfür die Ungenauigkeit zwischen einem vorhergesagten Wert $\hat{y}$ und einem richtigen Wert y. Eine gebräuchliche Kostenfunktion ist in etwa die Kreuzentropie

$$H(p,q) = -\sum_k p(k) \log q(k) \tag{1}$$

($p(x) \hat{=}$ wahre Wahrscheinlichkeitsdichte, $q(x) \hat{=}$ geschätzte Wahrscheinlichkeitsdichte)

Die Gleichung (1) liefert kleinere Werte je näher p und q sind.

b) Die Lossfunktion kann minimiert werden, indem man der Richtung des Gradienten in jedem Schritt folgt.

c) Aktivierungsfunktionen haben die Funktion, die Anregung einer Zelle im Zellkern zu simulieren und somit die Ausgabefunktion des Neurons zu bestimmen. 3 gängige Aktivierungsfunktionen sind:

1. sigmoid-Funktion:$\sigma(x) = \frac{1}{1+\exp -x}$
2. Tangens hyperbolicus :$\tanh(x)$
3. Rectified Linear Unit: $max(0,x)$

d) Ein Neuron bildet die Basis eines neuronalen Netzwerkes und ist einem Neuron in der Biologie nachempfunden. Mit den zugeweisenden Eingaben x_i und Gewichte W_i wird die Ausgabefunktion

$$\mathrm{net}_j = \sum_{i=0}^{n} x_i w_{ij} \tag{2}$$

definiert. Die Aktivierung dagegen ist definiert durch

$$o_j = \phi(\mathrm{net}_j - \theta_j) \tag{3}$$

($\phi \hat{=}$ Aktivierungsfunktion, $\theta \hat{=}$ 'Schellenwert' zur Verschiebung der Gewichtung)

e) Anwendungsbeispiele sind zum Beispiel:

1. Gesichts/Bilderkennung
2. Texterkennung
3. Spracherkennung

Diese Beispiele eignen sich besonders, da kein (oder wenig) explizites Wissen vorhanden sein muss, um diese zu identifizieren.

1 Aufgabe 21

1.1 a)-c)

siehe Anhang.

1.2 d)

Siehe Python-Datei.

$$W = \begin{pmatrix} -0,58 & 1,42 \\ 1,83 & -0,56 \end{pmatrix}$$

$$b = \begin{pmatrix} -0,48 \\ 0,61 \end{pmatrix}$$

1.3 e)

Mit

$$\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \tag{4}$$

und $f_1 = f_2$ folgt

$$\begin{aligned} W_{11}x + W_{12}y + b_1 &= W_{21}x + W_{22}y + b_2 \\ => y(x) &= \frac{b_2 - b_1 + x \cdot (W_{21} - W_{11})}{W_{12} - W_{22}} \end{aligned}$$

In Abbildung 1 sind die zwei Populationen dargestellt und die oben berechnete Trennung der beiden Populationen. $f_1 = f_2$ kann angenommen werden, da die Punkte auf der Geraden keiner Population zugeordnet werden können, da sie weder unter noch über der Geraden liegen, deshalb müssen f_1 und f_2 gleich sein.

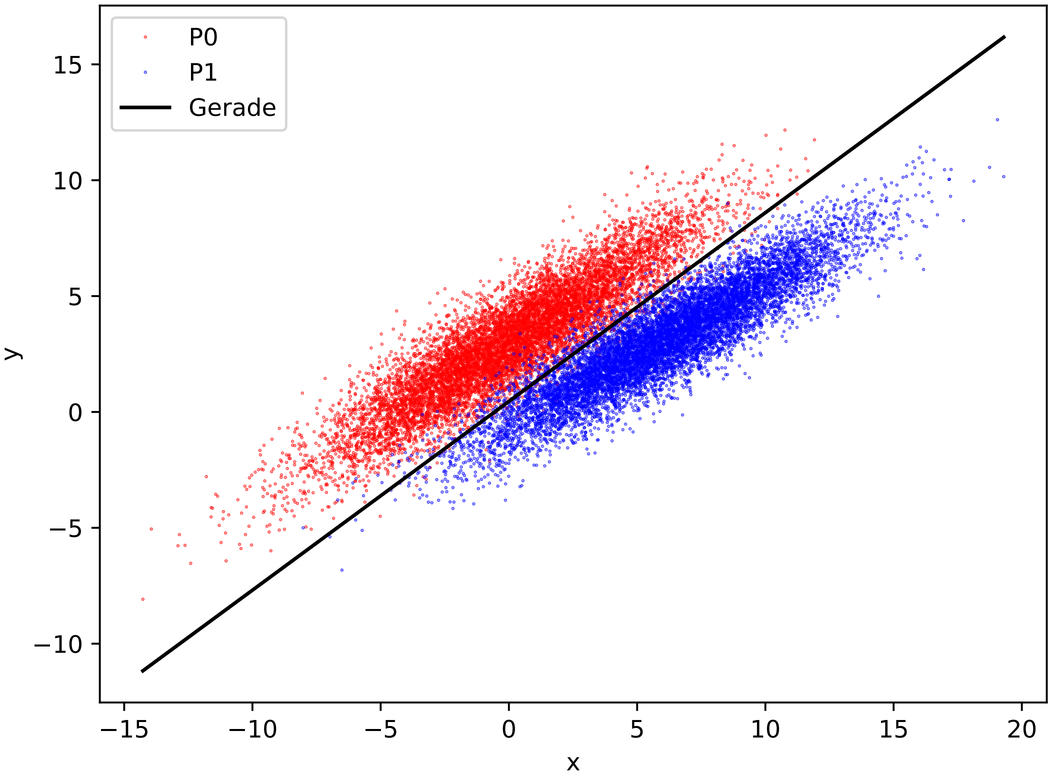


Abbildung 1: Populationen mit der berechneten Trennung.

SMD-Blatt 7 - Aufgabe 21

Lars Kolk
Julia Sobolewski
Jannine Salewski

a) - K Klassen
- m Beispiele x_i mit jeweils M Komponenten

$$\begin{pmatrix} \omega_{11} & \dots & \omega_{1,M} \\ \vdots & \ddots & \vdots \\ \omega_{K,1} & \dots & \omega_{K,M} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_K \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_K \end{pmatrix}$$

x_i : Spaltenvektor $x_i \in \mathbb{R}^{M\times 1}$

C : Kostenfunktion $C \in \mathbb{R}$

W: Gewichtungsmatrix $W \in \mathbb{R}^{K\times M}$

b : Bias-Vektor $b \in \mathbb{R}^{K\times 1}$

$$\nabla_W \hat{C} = \sum_{k=1}^{K} \underbrace{\frac{\partial \hat{C}}{\partial f_{ki}}}_{\in \mathbb{R}} \cdot \underbrace{\frac{\partial f_{ki}}{\partial W}}_{\in \mathbb{R}^{K\times M}} \Rightarrow \nabla_W \hat{C} \in \mathbb{R}^{K\times M}$$

$$\nabla_{f_i} \hat{C} = \sum_{k=1}^{K} \underbrace{\frac{\partial \hat{C}}{\partial f_{ki}}}_{\mathbb{R}} \underbrace{\frac{\partial f_{ki}}{\partial f_i}}_{\mathbb{R}^{K\times 1}} \Rightarrow \nabla_{f_i} \hat{C} \in \mathbb{R}^{K\times 1}$$

$$\frac{\partial f_{ki}}{\partial W} \in \mathbb{R}^{K\times M}$$

$$\frac{\partial f_{ki}}{\partial b} \in \mathbb{R}^{K\times 1}$$

b) $\nabla_{f_a} C(f) = \frac{1}{m} \sum_{i=1}^{m} \nabla_{f_a} \hat{C}(f)$

$$\nabla_{f_a} \hat{C}(f) = - \nabla_{f_a} \cdot \sum_{k=1}^{K} \mathbb{1}(y_i = k) \log\left(\frac{e^{f_{ki}}}{\sum_j e^{f_{ji}}}\right)$$

$$= - \nabla_{f_a} \log\left(\frac{e^{f_{y_i,i}}}{\sum_j e^{f_{ji}}}\right)$$

$$= \underbrace{\frac{-\sum_j e^{f_{ji}}}{e^{f_{y_i,i}}}}_{\text{äußere Ableitung}} \cdot \underbrace{\left[\frac{(\nabla_{f_a} e^{f_{y_i,i}}) \cdot \sum_j e^{f_{ji}} - e^{f_{y_i,i}} \cdot (\nabla_{f_a} \sum_j e^{f_{j,i}})}{(\sum_j e^{f_{j,i}})^2}\right]}_{\text{innere Ableitung}}$$

$$= - \frac{1}{e^{f_{y_i,i}}} \cdot \left[\frac{e^{f_{a,i}} \mathbb{1}(y_i = a) \cdot \sum e^{f_{j,i}} - e^{f_{y_i,i}} e^{f_{a,i}}}{\sum_j e^{f_{j,i}}}\right]$$

$$= - \frac{e^{f_{a,i}} \mathbb{1}(y_i = a)}{e^{f_{y_i,i}}} + \frac{e^{f_{y_i,i}} e^{f_{a,i}}}{e^{f_{y_i,i}} \sum_j e^{f_{j,i}}} = \frac{e^{f_{a,i}}}{\sum_j e^{f_{j,i}}} - \mathbb{1}(y_i = a)$$

$$\Rightarrow \nabla_{f_a} C(f) = \frac{1}{m} \sum_{i=1}^{m} \left[\frac{e^{f_{a,i}}}{\sum_j e^{f_{j,i}}} - \mathbb{1}(y_i = a)\right] \quad \text{qed.}$$

c) $$\frac{\partial f_{ki}}{\partial W} = \frac{\partial(W_k x_i + b_k)}{\partial W} = \underbrace{\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ x_1 & \cdots & x_i \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}}_{M} \leftarrow \text{k-te Zeile}$$

$$\frac{\partial f_{ki}}{\partial b} = \frac{\partial(W_k x_i + b_k)}{\partial b} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow \text{k-te Zeile}$$