

Lösungsblatt 2

Maschinelles Lernen und Klassifikation

Aufgabe 1: Zufallsexperiment

- a) Nehmen Sie eine Stoppuhr und lassen Sie eine andere Person diese starten. Versuchen Sie, ohne die Stoppuhr zu sehen, nach Ablauf von 5 Sekunden „Stopp“ zu sagen, woraufhin die andere Person die Uhr stoppt und den darauf angezeigten Wert notiert. Wiederholen Sie den Versuch n mal, ohne die bereits notierten Werte zu betrachten. Wählen Sie einen sinnvollen Wert für n , um einen Naiven Bayes-Klassifikator zu trainieren.
- b) Nehmen Sie an, das Experiment in Teilaufgabe a) unterliegt einem Zufallsprozess, der mit einer Gauß'schen Normalverteilung beschrieben werden kann. Somit lässt sich für jede Person, die den beschriebenen Versuch durchführt, eine spezifische Wahrscheinlichkeitsdichtefunktion beschreiben. Schätzen Sie die Parameter Ihrer persönlichen Verteilung mit den gemessenen Ergebnissen mit der *Maximum Likelihood*-Methode.
- c) Die in Teilaufgabe b) beschriebene Verteilungsfunktion soll die Wahrscheinlichkeitsdichte für eine Klasse ω_1 beschreiben. Nehmen Sie an, eine Kommilitonin hat einen ähnlichen Versuch durchgeführt. Aus ihren Messungen ergibt sich eine Klasse ω_2 , die ebenfalls einer Normalverteilung unterliegt mit dem Erwartungswert $\mu_2 = 5.5$ und der Varianz $\sigma_2^2 = 1$.

Führen Sie einen einzelnen weiteren Versuch wie beim Zufallsexperiment in Teilaufgabe a) durch und versuchen Sie, mit einem Naiven Bayes-Klassifikator zu ermitteln, ob das Ergebnis zu Ihnen selbst oder Ihrer Kommilitonin passt (ω_1 oder ω_2). Sie haben dabei die Handschrift analysiert, mit der das Ergebnis notiert wurde; Sie vermuten deshalb von vornherein, dass es mit einer Wahrscheinlichkeit von 0.3 zu ω_1 gehört und mit einer Wahrscheinlichkeit von 0.7 zu ω_2 .

Lösung:

- a) when $n = 10$; 10 random numbers:
5.951; 6.196; 6.409; 4.908; 4.876; 5.590; 6.102; 4.915; 4.818; 6.136;
- b) Expected value and variance:

$$\begin{aligned}\mu &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \frac{1}{10} \cdot (5.951 + 6.196 + 6.409 + 4.908 + 4.876 + 5.590 + 6.102 \\ &\quad + 4.915 + 4.818 + 6.136) \\ &\approx 5.590\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^n (x_i - \mu)^2 \cdot \frac{1}{n} \\
&= (5.951 - 5.590)^2 \cdot \frac{1}{10} + (6.196 - 5.590)^2 \cdot \frac{1}{10} + (6.409 - 5.590)^2 \cdot \frac{1}{10} + \\
&\quad (4.908 - 5.590)^2 \cdot \frac{1}{10} + (4.876 - 5.590)^2 \cdot \frac{1}{10} + (5.590 - 5.590)^2 \cdot \frac{1}{10} + \\
&\quad (6.102 - 5.590)^2 \cdot \frac{1}{10} + (4.915 - 5.590)^2 \cdot \frac{1}{10} + (4.818 - 5.590)^2 \cdot \frac{1}{10} + \\
&\quad (6.136 - 5.590)^2 \cdot \frac{1}{10} \\
&\approx 0.645 \\
\sigma &\approx 0.803
\end{aligned}$$

Therefore, the class 1 follows following distribution: $p(x|\omega_1) \sim N(5.590; 0.645)$ for chosen random numbers.

c) Posterior probability in Bayes theorem is defined:

$$P(\omega_c|x) = \frac{p(x|\omega_c)P(\omega_c)}{p(x)}$$

where $p(x)$ is same for $C = \{1; 2\}$. Therefore, $\frac{p(x|\omega_c)P(\omega_c)}{p(x)}$ is proportional to $p(x|\omega_c)P(\omega_c)$. This implies that we need to find the ω_c which maximizes $p(x|\omega_c)P(\omega_c)$. $\implies \arg \max_c p(x|\omega_c)P(\omega_c)$

Generate another random number, for example, $x = 5.805$. Then when $C = 1$,

$$\begin{aligned}
P(\omega_1|x) &= \frac{p(x|\omega_1)P(\omega_1)}{p(x)} \\
p(x|\omega_1)P(\omega_1) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2)} P(\omega_1) \\
&= \frac{1}{\sqrt{2\pi} \cdot 0.803} e^{(-\frac{1}{2}(\frac{5.805-5.590}{0.803})^2)} \cdot 0.3 \\
&\approx 0.497 \cdot e^{-0.036} \cdot 0.3 \approx 0.144
\end{aligned}$$

When $C = 2$,

$$\begin{aligned}
 P(\omega_2|x) &= \frac{p(x|\omega_2)P(\omega_2)}{p(x)} \\
 p(x|\omega_2)P(\omega_2) &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_2})^2)} P(\omega_2) \\
 &= \frac{1}{\sqrt{2\pi} \cdot 1} e^{(-\frac{1}{2}(\frac{5.805-5.5}{1})^2)} \cdot 0.7 \\
 &\approx 0.399 \cdot e^{-0.047} \cdot 0.7 \approx 0.293
 \end{aligned}$$

Therefore from the pervious calculation we have following probabilities.

$$\begin{aligned}
 p(x|\omega_1)P(\omega_1) &\approx 0.144 \\
 p(x|\omega_2)P(\omega_2) &\approx 0.293
 \end{aligned}$$

where $p(x|\omega_2)p(\omega_2)$ is higher than $p(x|\omega_1)p(\omega_1)$. As $P(\omega_c|x) = \frac{p(x|\omega_c)P(\omega_c)}{p(x)}$ is proportional to $p(x|\omega_c)P(\omega_c)$, we can conclude that $P(\omega_1|x) < P(\omega_2|x)$ implies $\arg \max_c p(x|\omega_c)P(\omega_c)$ for $C = \{1, 2\}$ yields $C = 2$.

Therefore, when $x = 5.805$, it is assigned to the class ω_2 according to the MAP-criteria.

Aufgabe 2: Fehlerwahrscheinlichkeit

Für zwei Klassen ω_1 und ω_2 mit den A-Priori-Wahrscheinlichkeiten $P(\omega_1) = \frac{3}{5}$ und $P(\omega_2) = \frac{2}{5}$ sind die folgenden Wahrscheinlichkeitsdichtefunktionen definiert:

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} \quad \text{und} \quad p(x|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}$$

- a) Wie hoch ist die Fehlerwahrscheinlichkeit in Abhängigkeit eines Schwellwerts θ , wenn sich eine Entscheidungsfunktion für ω_1 entscheidet wenn $x < \theta$ und ansonsten für ω_2 ?
- b) **Online-Frage Nr. 1:** Wie lautet der optimale Schwellwert θ_{opt} ?
- (i) 0.217
 - (ii) 1.101
 - (iii) 2.524
 - (iv) 0.251
 - (v) 1.058
- c) Berechnen Sie den kleinsten möglichen Klassifikationsfehler.
Hinweis: Formen Sie die Integrale aus der Funktion zur Berechnung der Fehlerwahrscheinlichkeit aus Teilaufgabe a) so um, dass sich deren Lösungen aus einer Tabelle zur Standardnormalverteilung $\Phi(z)$ ablesen lassen (z.B. unter http://de.wikipedia.org/wiki/Tabelle_Standardnormalverteilung).
- d) Sie haben eine Reihe von Messungen durchgeführt und die Messdaten mit einem Naiven Bayes-Klassifikator, bei dem Sie den Schwellwert auf das zuvor errechnete θ_{opt} gesetzt haben, klassifiziert. Sie stellen dabei fest, dass die tatsächliche Fehlerrate höher liegt als die in Teilaufgabe c) errechnete Fehlerwahrscheinlichkeit. Warum ist dies so?

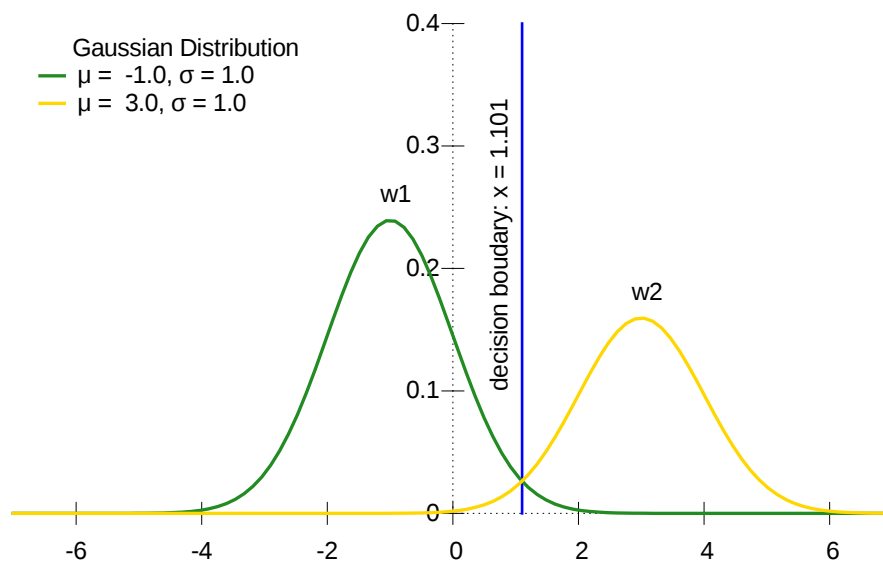
Lösung:

a)

$$\begin{aligned} P_{Fehler}(\theta) &= \int_{\theta}^{\infty} p(x|\omega_1)P(\omega_1)dx + \int_{-\infty}^{\theta} p(x|\omega_2)P(\omega_2)dx \\ &= \frac{3}{5} \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} dx + \frac{2}{5} \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} dx \end{aligned}$$

- b) Der optimale Schwellwert θ_{opt} liegt an der Stelle an der sich die Fehlerwahrscheinlichkeitsfunktionen kreuzen. Der richtige Wert kann auch über eine grafische Darstellung am Schnittpunkt der beiden Funktionen abgelesen werden.

$$\begin{aligned}
 P(\omega_1)p(x|\omega_1) &\stackrel{!}{=} P(\omega_2)p(x|\omega_2) \\
 \frac{3}{5} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\theta_{opt}+1)^2}{2}} &= \frac{2}{5} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\theta_{opt}-3)^2}{2}} \\
 3e^{-\frac{(\theta_{opt}+1)^2}{2}} &= 2e^{-\frac{(\theta_{opt}-3)^2}{2}} \\
 \frac{3}{2} &= e^{\frac{(\theta_{opt}+1)^2 - (\theta_{opt}-3)^2}{2}} \\
 1.5 &= e^{4\theta_{opt}-4} \\
 \ln 1.5 &= 4\theta_{opt} - 4 \\
 \theta_{opt} &= 1 + \frac{\ln 1.5}{4} \approx 1.101
 \end{aligned}$$



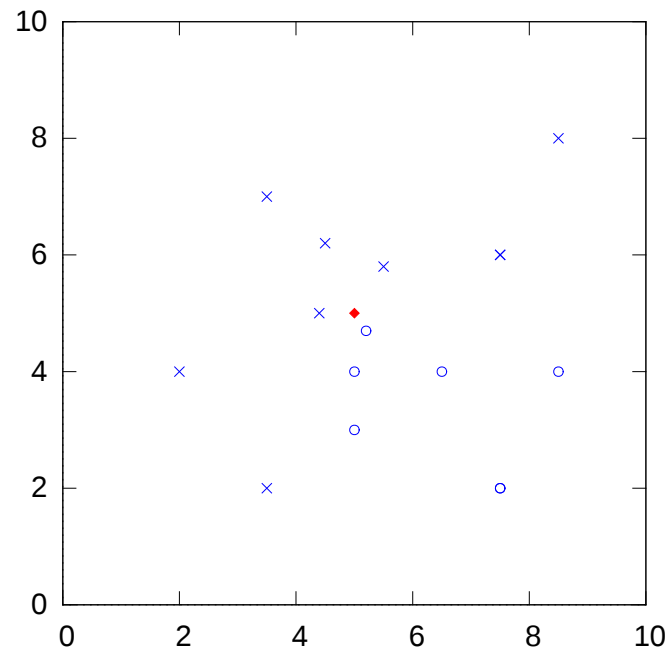
- c) The minimum error probability is when the threshold θ is the optimum value. Therefore, we need to calculate $P_{error}(\theta_{opt})$. Standard normal distribution is: $\Phi_{0;1}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt$.

$$\begin{aligned}
 P_{error}(\theta) &= \int_{\theta}^{\infty} p(x|\omega_1)P(\omega_1)dx + \int_{-\infty}^{\theta} p(x|\omega_2)P(\omega_2)dx \\
 &= \frac{3}{5} \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} dx + \frac{2}{5} \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} dx \\
 &= \frac{3}{5} \int_{1.1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}} dx + \frac{2}{5} \int_{-\infty}^{1.1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} dx \\
 &= \frac{3}{5} \cdot \frac{1}{\sqrt{2\pi}} \int_{1.1+1}^{\infty} e^{-\frac{x^2}{2}} dx + \frac{2}{5} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1.1-3} e^{-\frac{x^2}{2}} dx \\
 &= \frac{3}{5} \cdot \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{2.1} e^{-\frac{x^2}{2}} dx\right) + \frac{2}{5} \cdot \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1.9} e^{-\frac{x^2}{2}} dx\right) \\
 &= \frac{3}{5} \cdot (1 - \Phi(2.1)) + \frac{2}{5} \cdot (1 - \Phi(1.9)) \\
 &\approx \frac{1}{5} \cdot (1 - 0.98) + \frac{4}{5} \cdot (1 - 0.97) \\
 &= \frac{3}{5} \cdot 0.02 + \frac{2}{5} \cdot 0.03 \\
 &\approx 0.022
 \end{aligned}$$

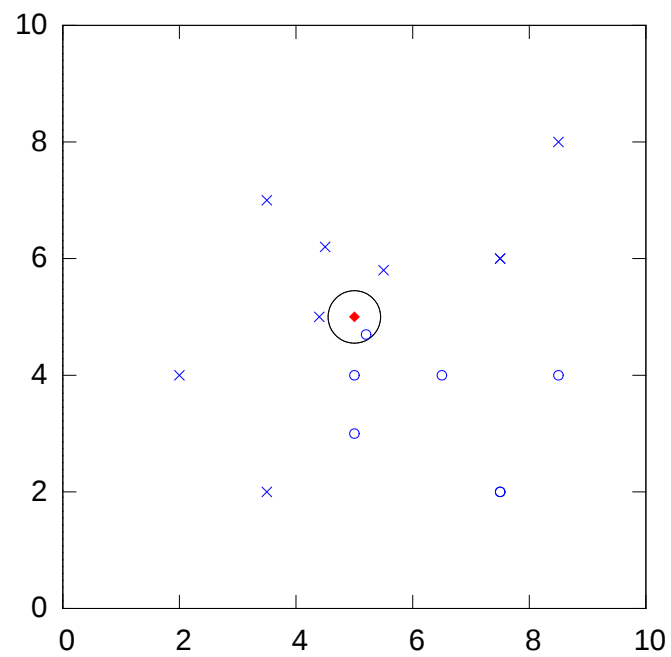
- d)
 - Die Wahrscheinlichkeitsdichtefunktion ist immer nur eine Schätzung, die der realen Verteilung möglichst nahe kommen soll, sie aber nicht perfekt abbildet.
 - Zu wenige oder verzerrte Messdaten.

Aufgabe 3: *k*-Nearest-Neighbors

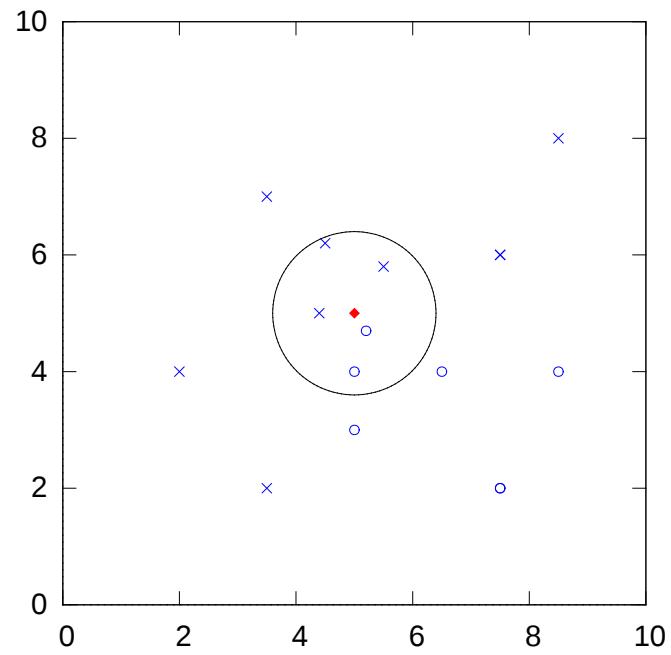
Gegeben sei der folgende Datensatz mit zwei Klassen (Kreise und Kreuze). Klassifizieren Sie das mit einer Raute markierte Merkmal an den Koordinaten (5, 5) mit dem *k*-Nearest-Neighbors Algorithmus für $k = 1$ und $k = 5$. Benutzen Sie dabei zur Gewichtung die euklidische Distanz.



Lösung für $k = 1$: Die Raute wird als Kreis klassifiziert.



Lösung für $k = 5$: Die Raute wird als Kreuz klassifiziert.

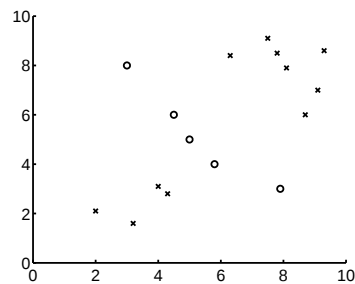


Online-Frage Nr. 2: Wie wird die Raute für $k = 5$ klassifiziert?

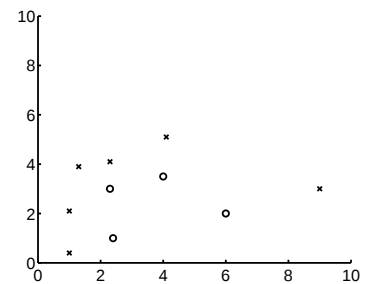
- (i) keine Klasse
- (ii) Kreis
- (iii) Kreuz

Aufgabe 4: Perzeptronen

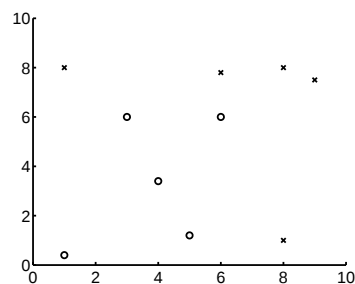
In den Grafiken (a) bis (d) sind jeweils Datenpunkte zweier Klassen eingezeichnet.



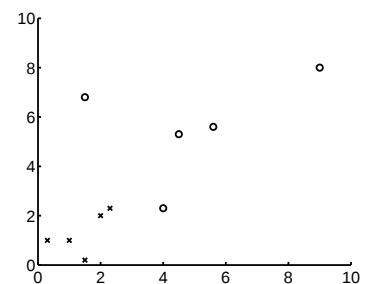
(a)



b)



(c)



d)

Online-Frage Nr. 3: Welche der folgenden Aussagen sind für die jeweils angegebene Abbildung wahr? Bitte alle richtigen Antworten im Online-Formular ankreuzen.

- (i) Abbildung a) Die beiden Klassen sind linear separierbar.
- (ii) Abbildung b) Die beiden Klassen lassen sich mit einem einfachen Perzeptron trennen.
- (iii) Abbildung b) Die beiden Klassen lassen sich mit einem *Multi Layer Perceptron* mit einer versteckten Schicht, die aus zwei Neuronen besteht, trennen.
- (iv) Abbildung c) Die beiden Klassen sind linear separierbar.
- (v) Abbildung c) Nach einer Umrechnung in Polarkoordinaten (ohne Verschiebung) sind die beiden Klassen linear separierbar.
- (vi) Abbildung d) Die beiden Klassen lassen sich mit einem einfachen Perzeptron trennen.

Lösung:

- (i) Falsch.
- (ii) Falsch.
- (iii) Richtig.
- (iv) Falsch.
- (v) Falsch.

(vi) Richtig.

Aufgabe 5: Neuronale Netze

Gegeben seien die folgenden Punktwolken mit zweidimensionalen Merkmalsvektoren, die der positiven Klasse A und negativen Klasse B zugeordnet sind:

$$A = \{\vec{a}_1, \vec{a}_2\} = \left\{ \begin{pmatrix} -0.4 \\ 0.8 \end{pmatrix}; \begin{pmatrix} -0.8 \\ -0.2 \end{pmatrix} \right\} \text{ und } B = \{\vec{b}_1, \vec{b}_2\} = \left\{ \begin{pmatrix} 0.3 \\ 1.6 \end{pmatrix}; \begin{pmatrix} -0.4 \\ 1.5 \end{pmatrix} \right\}.$$

a) Berechnen Sie ein einfaches Perzeptron mit binärer Schwellwertfunktion, das die Klassen A und B trennt. Verwenden Sie den folgenden iterativen Algorithmus:

$$\begin{pmatrix} \vec{w}_{i+1} \\ b_{i+1} \end{pmatrix} = \begin{pmatrix} \vec{w}_i \\ b_i \end{pmatrix} - \sum_{\vec{e} \in E} \left[\delta_{\vec{e}} \cdot \begin{pmatrix} \vec{e} \\ 1 \end{pmatrix} \right]$$

- b_i entspricht dem Schwellenwert in Iterationsschritt i .
- $\vec{w}_i = \begin{pmatrix} w_{i1} \\ w_{i2} \end{pmatrix}$ ist der Gewichtsvektor im Iterationsschritt i .
- $E_i = \{\vec{e}_{i1}, \vec{e}_{i2}, \dots\} \subseteq \{\vec{a}_1, \vec{a}_2, \vec{b}_1, \vec{b}_2\}$ ist die Menge aller fehlerhaft klassifizierten Merkmalsvektoren in Iteration i .
- $\delta_{\vec{e}}$ ist das Fehlersignal: $\delta_{\vec{e}} = -1$ falls $\vec{e} \in a$ und $\delta_{\vec{e}} = 1$ falls $\vec{e} \in b$.
- Beginnen Sie die Iteration mit den Werten $\vec{w}_0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ und $b_0 = 0$ und skizzieren Sie die Trenngerade nach jeder Iteration i . Der Algorithmus terminiert in Iteration n , sobald alle Punkte richtig klassifiziert werden, d.h. $E_n = \emptyset$.

b) **Onlinefrage Nr. 4:** Welchen Wert erhält man für $\{w_{n1}, w_{n2}, b_n\}$ wenn der Algorithmus terminiert?

- i) 3.2, -0.4, -1
- ii) -0.5, -0.9, 1
- iii) -0.9, -1.9, 2
- iv) -1.0, -2.0, 2
- v) 0.5, 2.8, 0

Lösung:

a) Die lineare Diskriminierungsfunktion eines einfachen Perzeptrons mit zwei Neuronen hat allgemein die Form: $g(x) = w_1 x_1 + w_2 x_2 + b$.

Die Klassenzuordnung erfolgt über den Wert von $g(x)$: Falls $g(x) > 0 \implies$ Klasse A und falls $g(x) < 0 \implies$ Klasse B. Die Trenngerade zwischen den beiden Klassen verläuft entsprechend bei $g(x) = 0$.

Zur Zeichnung in ein Achsenkreuz mit den beiden Achsen x_1 und x_2 wird die Trenngerade bei $g(x) = 0$ nach x_2 aufgelöst:

$$\begin{aligned}
w_1x_1 + w_2x_2 + b &= 0 \\
w_2x_2 &= -w_1x_1 - b \\
x_2 &= -\frac{w_1}{w_2}x_1 - \frac{1}{w_2}b
\end{aligned}$$

Als Trenngerade für die initialen Werte ergibt sich aus $g_0(x) = x_1 - x_2$ demnach die Geradengleichung $x_2 = x_1$ für ein zweidimensionales Achsenkreuz.

Der Algorithmus durchläuft die folgenden Iterationsschritte:

$$1. \vec{w}_0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}; b_0 = 0 \implies g_0(x) = x_1 - x_2$$

$$g_0(\vec{a}_1) = -1.2 \xRightarrow{!} B \text{ (Fehlklassifikation)}$$

$$g_0(\vec{a}_2) = -0.6 \xRightarrow{!} B \text{ (Fehlklassifikation)}$$

$$g_0(\vec{b}_1) = -1.3 \implies B$$

$$g_0(\vec{b}_2) = -1.9 \implies B$$

Damit sind zwei Punkte aus Klasse A falsch klassifiziert, $E_0 = \{\vec{a}_1, \vec{a}_2\}$.

Daraus ergibt sich nach dem ersten Iterationsschritt:

$$\begin{aligned}
\begin{pmatrix} \vec{w}_1 \\ b_1 \end{pmatrix} &= \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - (-1)\vec{a}_1 - (-1)\vec{a}_2 \\
&= \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - (-1) \begin{pmatrix} -0.4 \\ 0.8 \\ 1 \end{pmatrix} - (-1) \begin{pmatrix} -0.8 \\ -0.2 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix} \\
&\implies g_1(x) = -0.2x_1 - 0.4x_2 + 2
\end{aligned}$$

Geradengleichung für die Zeichnung: $x_2 = -0.5x_1 + 5$

$$2. g_1(\vec{a}_1) = 1.76 \implies A$$

$$g_1(\vec{a}_2) = 2.24 \implies A$$

$$g_1(\vec{b}_1) = 1.3 \xRightarrow{!} A \text{ (Fehlklassifikation)}$$

$$g_1(\vec{b}_2) = 1.48 \xRightarrow{!} A \text{ (Fehlklassifikation)}$$

Damit sind zwei Punkte aus Klasse B falsch klassifiziert: $E_1 = \{\vec{b}_1, \vec{b}_2\}$

$$\begin{aligned}
\begin{pmatrix} \vec{w}_2 \\ b_2 \end{pmatrix} &= \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix} - (1)\vec{b}_1 - (1)\vec{b}_2 = \begin{pmatrix} -0.2 \\ -0.4 \\ 2 \end{pmatrix} - (1) \begin{pmatrix} 0.4 \\ 1.6 \\ 1 \end{pmatrix} - (1) \begin{pmatrix} -0.4 \\ 1.5 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.1 \\ -3.5 \\ 0 \end{pmatrix} \\
&\implies g_2(x) = -0.1x_1 - 3.5x_2
\end{aligned}$$

Geradengleichung für die Zeichnung: $x_2 = \frac{-x_1}{35}$

$$\begin{aligned}
3. \quad g_2(\vec{a}_1) &= -2.76 \stackrel{!}{\Rightarrow} A \\
g_2(\vec{a}_2) &= 0.78 \Rightarrow A \\
g_2(\vec{b}_1) &= -5.63 \Rightarrow B \\
g_2(\vec{b}_2) &= -5.21 \Rightarrow B
\end{aligned}$$

$$\begin{pmatrix} \vec{w}_3 \\ b_3 \end{pmatrix} = \begin{pmatrix} -0.5 \\ -2.7 \\ 1 \end{pmatrix}$$

Geradengleichung für die Zeichnung: $x_2 = \frac{-5x_1+10}{27}$

$$\begin{aligned}
4. \quad g_3(\vec{a}_1) &= -0.96 \stackrel{!}{\Rightarrow} A \\
g_3(\vec{a}_2) &= 1.94 \Rightarrow A \\
g_3(\vec{b}_1) &= -4.47 \Rightarrow B \\
g_3(\vec{b}_2) &= -2.85 \Rightarrow B
\end{aligned}$$

$$\begin{pmatrix} \vec{w}_4 \\ b_4 \end{pmatrix} = \begin{pmatrix} -0.9 \\ -1.9 \\ 2 \end{pmatrix}$$

Geradengleichung für die Zeichnung: $x_2 = \frac{-9x_1+20}{19}$

$$\begin{aligned}
5. \quad g_4(\vec{a}_1) &= 0.84 \Rightarrow A \\
g_4(\vec{a}_2) &= 3.1 \Rightarrow A \\
g_4(\vec{b}_1) &= -1.31 \Rightarrow B \\
g_4(\vec{b}_2) &= -0.49 \Rightarrow B
\end{aligned}$$

Alle Trainingsdatenpunkte werden damit korrekt klassifiziert: $E_2 = \emptyset \Rightarrow$ der Algorithmus terminiert mit den Werten $(\vec{w}_4, b_4) = (-0.9, -1.9, 2.0)^T$.

