3. Perzeptron

3.1 Definition von Perzeptron

Überblick:

- Ein-/Ausgabe, Propagierungs-/Aktivierungsfunktionen
- Anwendungsbeispiel Logik-Funktionen
- System von Ungleichungen für f_{AND}

Ein-/Ausgabe, Propag./Aktiv.funktionen

Rosenblatt (1962)

- Reelle Input-Werte $\boldsymbol{x_i}$, reelle Gewichte $\boldsymbol{w_i}$, reeller Schwellenwert $\boldsymbol{\theta}$
- Linearer Assoziator als Propagierungsfunktion
- Stufenförmige Aktivierungsfunktion
- Binärer Output-Wert $\boldsymbol{y}$
- Perzeptron Entscheidungsregel

$$\boldsymbol{y} := \begin{cases} 1 & : \text{ falls } \sum_{i=1}^{I} \boldsymbol{w_i x_i} \geq \boldsymbol{\theta} \\ 0 & : \text{ sonst} \end{cases}$$

Hinweis auf Kurzformen des linearen Assoziators: $\sum_{i=1}^{I} \boldsymbol{w_i x_i} = \boldsymbol{w}^T \boldsymbol{x} = \boldsymbol{w} \circ \boldsymbol{x}$

Anwendungsbeispiel Logik-Funktionen

Logik-Funktionen (binäre Entscheidungsaufgabe).

Beispiele f_{AND} und f_{OR}:

x_1	x_2	f_{AND}	f_{OR}
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

Frage: Wie erhält man Gewichte?

Antwort: Lösen von System von Ungleichungen.

$$\begin{aligned} w_1 x_1 + w_2 x_2 \geq \theta \quad &\Longleftrightarrow \quad y = 1 \\ w_1 x_1 + w_2 x_2 < \theta \quad &\Longleftrightarrow \quad y = 0 \end{aligned}$$

System von Ungleichungen für f_{AND}

Im Falle von f_{AND} heisst das:

$$
\begin{aligned}
\boldsymbol{w_1} \cdot 0 + \boldsymbol{w_2} \cdot 0 &< \boldsymbol{\theta} \\
\boldsymbol{w_1} \cdot 0 + \boldsymbol{w_2} \cdot 1 &< \boldsymbol{\theta} \\
\boldsymbol{w_1} \cdot 1 + \boldsymbol{w_2} \cdot 0 &< \boldsymbol{\theta} \\
\boldsymbol{w_1} \cdot 1 + \boldsymbol{w_2} \cdot 1 &\geq \boldsymbol{\theta}
\end{aligned}
$$

3.2 Lineare Klassifikatoren

Überblick:

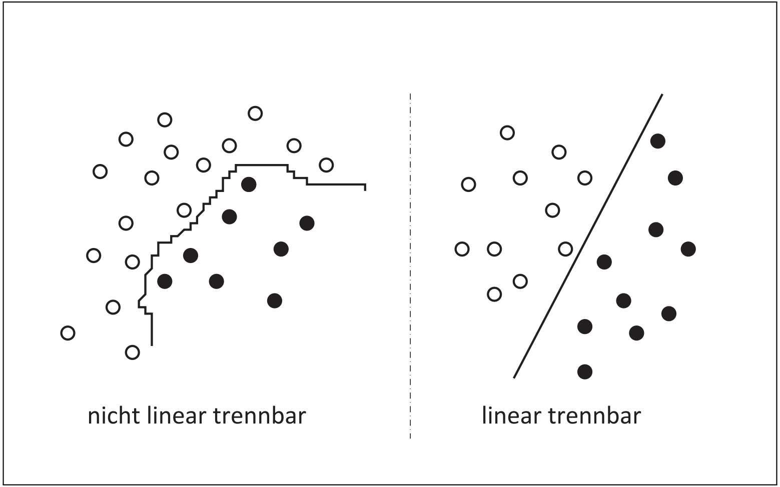
- Lineare Trennbarkeit
- Beispiele zur Trennbarkeit
- Absolute lineare Trennbarkeit
- Trenngerade in Achsenabschnittsform
- Trenngerade orthogonal zu Gewichtsvektor
- Realisierung von f_{AND}
- Realisierung von f_{OR}

Lineare Trennbarkeit

Definition:

Zwei Mengen $\mathcal{P}$ und $\mathcal{N}$ von Punkten einer Trainingsmenge ($\Omega_T := \mathcal{P} \cup \mathcal{N}$) in einem I-dimensionalen Eingaberaum sind **linear trennbar**,
falls $I + 1$ reelle Zahlen $\theta, w_1, \ldots, w_I$ existieren,
so dass für jeden Punkt $x \in \mathcal{P}$ gilt: $\sum_{i=1}^{I} w_i x_i \geq \theta$,
und für jeden Punkt $x \in \mathcal{N}$ gilt: $\sum_{i=1}^{I} w_i x_i < \theta$.

Beispiele zur Trennbarkeit



Absolute lineare Trennbarkeit

Im Falle absoluter linearer Trennbarkeit gilt:

$$\boldsymbol{x} \in \boldsymbol{\mathcal{P}} \iff \boldsymbol{w}^T \boldsymbol{x} > \boldsymbol{\theta}$$
$$\boldsymbol{x} \in \boldsymbol{\mathcal{N}} \iff \boldsymbol{w}^T \boldsymbol{x} < \boldsymbol{\theta}$$

Trenngerade in Achsenabschnittsform

$\boldsymbol{x} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I)^T$, $\boldsymbol{w} := (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_I)^T$

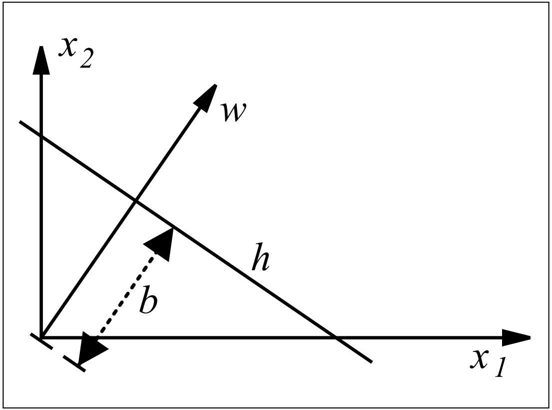
Weiter mit Spezialfall $\boldsymbol{I = 2}$.

Die Trenngerade $\boldsymbol{h : w_1 x_1 + w_2 x_2 = \theta}$ hat die Achsenabschnittsform:

$$\boldsymbol{x_2 := \frac{\theta}{w_2} - \frac{w_1}{w_2} x_1}$$

Trenngerade orthogonal zu Gewichtsvektor

Satz: Die Trenngerade $\boldsymbol{h}$ steht senkrecht zum Gewichtsvektor $\boldsymbol{w}$: $\boldsymbol{h} \perp \boldsymbol{w}$



Trenngerade orthogonal zu Gewichtsvektor

Beweis des Satzes:

$x_2 = 0 \rightarrow x_1 = \frac{\theta}{w_1}, \qquad x_1 = 0 \rightarrow x_2 = \frac{\theta}{w_2}$

Vektor der Trenngerade:

$$\vec{h} := \begin{pmatrix} \frac{\theta}{w_1} \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ \frac{\theta}{w_2} \end{pmatrix} = \begin{pmatrix} \frac{\theta}{w_1} \\ -\frac{\theta}{w_2} \end{pmatrix}$$

Skalarprodukt:

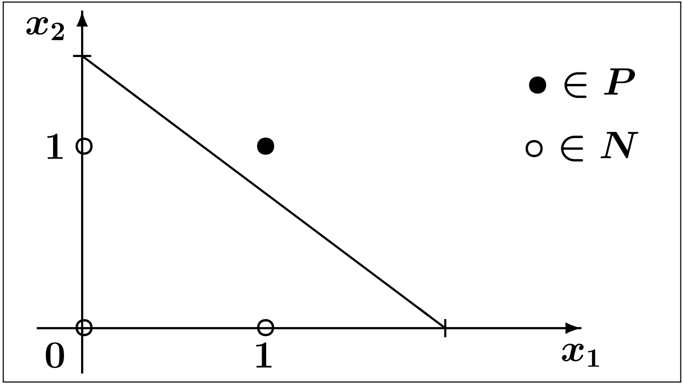
$$\vec{h} \circ w = \frac{\theta}{w_1} \cdot w_1 + (-)\frac{\theta}{w_2} \cdot w_2 = \theta - \theta = 0$$

q.e.d.

Allgemein: w ist immer die Normale zur Hyper-Trennebene!

Realisierung von f_{AND}

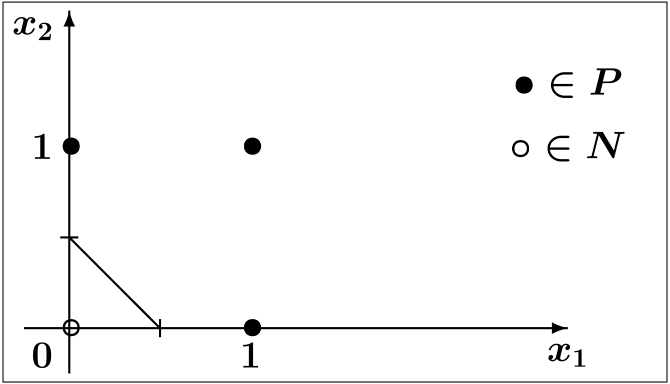
Bsp.: $\boldsymbol{w^e := (w_0, w_1, w_2)^T = (-1, 0.5, 0.7)^T}$



Für $\boldsymbol{x_2 = 0}$ folgt $\boldsymbol{x_1 = 2}$; Für $\boldsymbol{x_1 = 0}$ folgt $\boldsymbol{x_2 \approx 1.4}$

Realisierung von f_{OR}

Bsp.: $\boldsymbol{w^e := (w_0, w_1, w_2)^T = (-1, 2, 2)^T}$



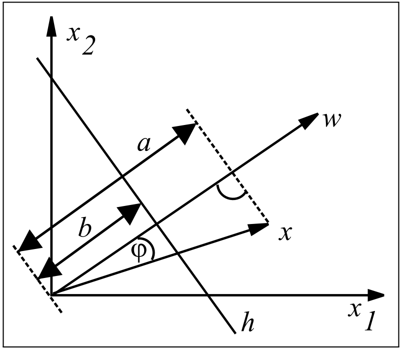
Für $\boldsymbol{x_2 = 0}$ folgt $\boldsymbol{x_1 = 0.5}$; Für $\boldsymbol{x_1 = 0}$ folgt $\boldsymbol{x_2 = 0.5}$

3.3 Interpretation des Skalarprodukts

Überblick:

- Projektion von $\boldsymbol{x}$ auf $\boldsymbol{w}$
- Abstand der Trenngerade vom Ursprung
- Binäre Klassifikation nach 1D-Projektion

Projektion von $\boldsymbol{x}$ auf $\boldsymbol{w}$



Es gilt $\boldsymbol{w}^T\boldsymbol{x} = \boldsymbol{w} \circ \boldsymbol{x} = \|\boldsymbol{w}\| \cdot \underbrace{\|\boldsymbol{x}\| \cdot \cos\varphi}_{\boldsymbol{a}},\quad$ mit $\cos\varphi = \frac{\boldsymbol{a}}{\|\boldsymbol{x}\|}$

Projektion von $\boldsymbol{x}$ auf $\boldsymbol{w}$ und Skalierung mit $\|\boldsymbol{w}\|$.

Abstand der Trenngerade vom Ursprung

Für beliebigen Punkt $\boldsymbol{x}$ auf der Trenngeraden $\boldsymbol{h}$ gilt:

$$\boldsymbol{\theta} = \boldsymbol{w}^T \boldsymbol{x} = \|\boldsymbol{w}\| \cdot \underbrace{\|\boldsymbol{x}\| \cdot \cos \boldsymbol{\varphi}}_{\boldsymbol{a}=\boldsymbol{b}}$$

also

$$\boldsymbol{b} = \frac{\boldsymbol{\theta}}{\|\boldsymbol{w}\|}$$

Binäre Klassifikation nach 1D-Projektion

Äquivalenzen:

$$
\begin{aligned}
& \boldsymbol{w}^T \boldsymbol{x} \geq \boldsymbol{\theta} \;\rightarrow\; \boldsymbol{y} = \mathbf{1} \\
\iff\quad & \|\boldsymbol{w}\| \cdot \boldsymbol{a} \geq \|\boldsymbol{w}\| \cdot \boldsymbol{b} \;\rightarrow\; \boldsymbol{y} = \mathbf{1} \\
\iff\quad & \boldsymbol{a} \geq \boldsymbol{b} \;\rightarrow\; \boldsymbol{y} = \mathbf{1}
\end{aligned}
$$

3.4 Lineare Trennfunktion

Überblick:

- Trennfunktion als Hyperebene
- Lernaufgabe der linearen Trennung
- Äquivalentes System von Ungleichungen
- Visualisierung äquivalentes System
- Äquivalente Repräsentationsarten für f_{AND}
- Äquivalentes Ungleichungssystem für f_{AND}
- Kompakte Notation des Ungleichungssystems
- Äquivalentes Ungleichungssystem für f_{AND}
- Kompakte Notation für f_{AND}

Trennfunktion als Hyperebene

Die lineare Trennfunktion ist $(\boldsymbol{I}-\mathbf{1})$-dimensionale Hyperebene des $\boldsymbol{I}$-dimensionalen Eingaberaumes.

Perzeptron-Definition mit spezieller Eingabekomponente $\boldsymbol{x_0} = \mathbf{1}$ und Gewicht $\boldsymbol{w_0} = -\boldsymbol{\theta}$ bedeutet, dass die Trennfunktion durch $\boldsymbol{w^{e,T} x^e} = \mathbf{0}$ gegeben ist.

Diese geht wegen $\boldsymbol{b} = \mathbf{0}$ durch den Ursprung des um eine Dimension erhöhten Eingaberaumes.

Trennfunktion als Hyperebene

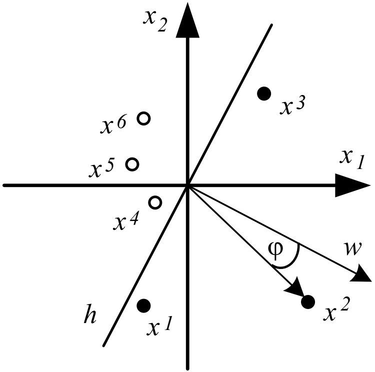
Falls $\boldsymbol{h}$ durch den Ursprung des Koordinatensystems geht, dann ist die lineare Entscheidungsaufgabe des Perzeptrons äquivalent mit folgender Bedingung:

$$\boldsymbol{x} \in \boldsymbol{\mathcal{P}}, \ \text{falls } -\frac{\boldsymbol{\pi}}{\mathbf{2}} \leq \boldsymbol{\varphi} \leq \frac{\boldsymbol{\pi}}{\mathbf{2}}$$

$$\boldsymbol{x} \in \boldsymbol{\mathcal{N}}, \ \text{falls } \frac{\boldsymbol{\pi}}{\mathbf{2}} < \boldsymbol{\varphi} < \frac{\mathbf{3}}{\mathbf{2}}\boldsymbol{\pi}$$

Lernaufgabe der linearen Trennung

Finde für Trainingselemente $\boldsymbol{x} \in \boldsymbol{\Omega}_T$ einen Gewichtsvektor $\boldsymbol{w}$ zur Trennung von $\mathcal{P}$ und $\mathcal{N}$.



$$\mathcal{P} := \{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3\}$$

$$\mathcal{N} := \{\boldsymbol{x}^4, \boldsymbol{x}^5, \boldsymbol{x}^6\}$$

Äquivalentes System von Ungleichungen

$$\boldsymbol{x}^m \in \mathcal{P} \Leftrightarrow \boldsymbol{w}^T \boldsymbol{x}^m > 0$$

$$\Leftrightarrow y^{m,soll} = 1 \Leftrightarrow y^{m,soll} \cdot \boldsymbol{w}^T \cdot \boldsymbol{x}^m > 0$$

$soll$: Soll-Output, gewünschter Output

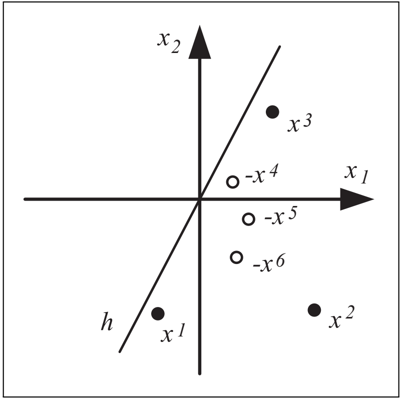
$$\boldsymbol{x}^m \in \mathcal{N} \Leftrightarrow \boldsymbol{w}^T \boldsymbol{x}^m < 0$$

$$\Leftrightarrow y^{m,soll} \overset{*)}{=} -1 \Leftrightarrow y^{m,soll} \cdot \boldsymbol{w}^T \cdot \boldsymbol{x}^m > 0$$

*): alternativ zu Folie 3, aber äquivalent

$$\Longrightarrow \boldsymbol{w}^T \cdot \zeta^m > 0 \text{ mit } \zeta^m := y^{m,soll} \cdot \boldsymbol{x}^m$$

Visualisierung äquivalentes System



$\Omega'_T := \mathcal{P}' \cup \mathcal{N}'$ mit

$$\mathcal{P}' := \{\boldsymbol{\zeta}^m = \boldsymbol{x}^m | \boldsymbol{x}^m \in \mathcal{P}\}$$

und

$$\mathcal{N}' := \{\boldsymbol{\zeta}^m = -\boldsymbol{x}^m | \boldsymbol{x}^m \in \mathcal{N}\}$$

Äquivalente Lernaufgabe:

Finde Gewichtsvektor $\boldsymbol{w}$, so dass $\boldsymbol{w}^T\boldsymbol{\zeta} > \mathbf{0}$, $\forall \boldsymbol{\zeta} \in \Omega'_T$.

Äquivalente Repräsentationsarten für f_{AND}

	$\boldsymbol{x}^1$	$\boldsymbol{x}^2$	$\boldsymbol{x}^3$	$\boldsymbol{x}^4$
$\boldsymbol{x}_0$	1	1	1	1
$\boldsymbol{x}_1$	0	0	1	1
$\boldsymbol{x}_2$	0	1	0	1
$\boldsymbol{y}^{soll}$	-1	-1	-1	1

$\rightsquigarrow$

	ζ^1	ζ^2	ζ^3	ζ^4
ζ_0	-1	-1	-1	1
ζ_1	0	0	-1	1
ζ_2	0	-1	0	1

Äquivalentes Ungleichungssystem für f_{AND}

$$\begin{aligned}
w_0 \cdot \zeta_0^1 + w_1 \cdot \zeta_1^1 + w_2 \cdot \zeta_2^1 > 0 \quad &\rightarrow \quad -w_0 > 0 \\
w_0 \cdot \zeta_0^2 + w_1 \cdot \zeta_1^2 + w_2 \cdot \zeta_2^2 > 0 \quad &\rightarrow \quad -w_0 - w_2 > 0 \\
w_0 \cdot \zeta_0^3 + w_1 \cdot \zeta_1^3 + w_2 \cdot \zeta_2^3 > 0 \quad &\rightarrow \quad -w_0 - w_1 > 0 \\
w_0 \cdot \zeta_0^4 + w_1 \cdot \zeta_1^4 + w_2 \cdot \zeta_2^4 > 0 \quad &\rightarrow \quad w_0 + w_1 + w_2 > 0
\end{aligned}$$

Hinweis: Identische Ungleichungen wie in Folie 5.

Kompakte Notation des Ungleichungssystems

Für die gesamte Trainingsmenge gilt also: $\quad \boldsymbol{Z}^{\boldsymbol{T}} \cdot \boldsymbol{w} > \vec{\mathbf{0}}$

Mit $\boldsymbol{Z} := \left(\boldsymbol{\zeta}^{\mathbf{1}}, \ldots, \boldsymbol{\zeta}^{|\boldsymbol{\Omega}_{\boldsymbol{T}}|}\right)$,

und Spaltenvektor $\vec{\mathbf{0}} := \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$ mit $|\boldsymbol{\Omega}_{\boldsymbol{T}}|$ Komponenten.

Kompakte Notation für f_{AND}

Im Beispiel f_{AND} erhält man die $(\mathbf{3} \times \mathbf{4})$-Matrix $\boldsymbol{Z}$:

$$\boldsymbol{Z} := \begin{pmatrix} -\mathbf{1} & -\mathbf{1} & -\mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & -\mathbf{1} & \mathbf{1} \\ \mathbf{0} & -\mathbf{1} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

Das System der Ungleichungen $\boldsymbol{Z}^{\boldsymbol{T}} \cdot \boldsymbol{w} > \vec{\mathbf{0}}$ ergibt sich zu:

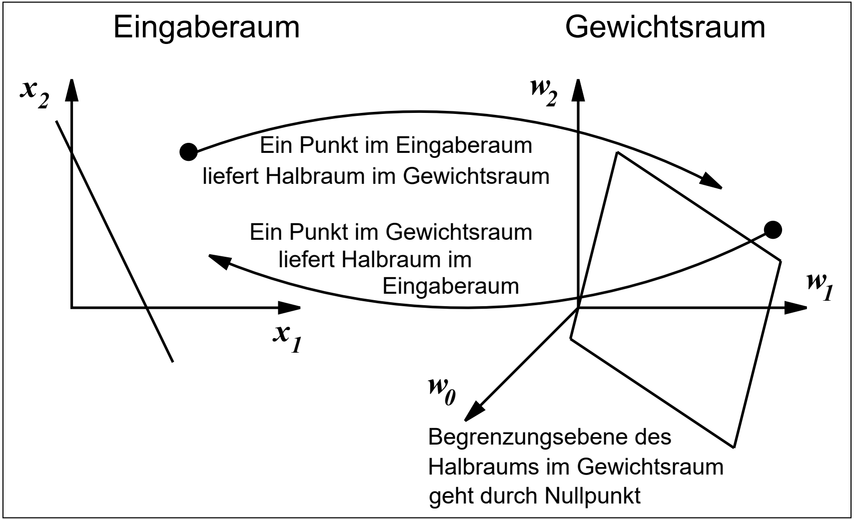
$$\begin{pmatrix} -\mathbf{1} & \mathbf{0} & \mathbf{0} \\ -\mathbf{1} & \mathbf{0} & -\mathbf{1} \\ -\mathbf{1} & -\mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{w}_{\mathbf{0}} \\ \boldsymbol{w}_{\mathbf{1}} \\ \boldsymbol{w}_{\mathbf{2}} \end{pmatrix} > \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$$

3.5 Dualität Eingabe-/Gewichtsraum

Überblick:

- Visualisierung der Dualität
- Definition von Halbräumen
- Fehlerfunktion im Gewichtsraum
- Fehlerfunktion für f_{AND}
- Visualisierung der Fehlerfunktion für f_{AND}

Visualisierung der Dualität



Definition von Halbräumen

Für gegebenes $\boldsymbol{x} \in \mathcal{P}$ ($\boldsymbol{I}$-dimensional)

wird durch die Forderung $\boldsymbol{w_0 + w_1 x_1 + \ldots + w_I x_I > 0}$

im Gewichtsraum (($\boldsymbol{I+1}$)-dimensional) ein positiver Halbraum definiert.

Dual dazu:

Durch gegebenes $\boldsymbol{w}$ (($\boldsymbol{I+1}$)-dimensional)

wird im Eingaberaum ($\boldsymbol{I}$-dimensional) ein positiver Halbraum definiert.

Fehlerfunktion im Gewichtsraum

Basierend auf dem Gewichtsraum kann eine Fehlerfunktion definiert werden.

Die Fehlerfunktion summiert für einen Gewichtsvektor (Punkt im Gewichtsraum) die Anzahl fehlerhafter Resultate (Fehlentscheidungen), wenn die komplette Trainingsmenge klassifiziert wird.

Fehlerfunktion für f_{AND}

$$\boldsymbol{w_0 = -\theta = -1,\ \mathcal{N} := \{x^1, x^2, x^3\},\ \mathcal{P} := \{x^4\}}.$$

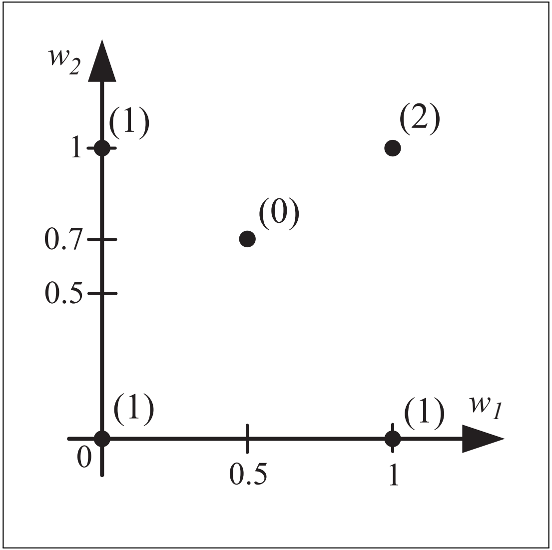
$$\begin{aligned}
x^1 &:= (0,0)^T &:&\quad -1 + w_1 0 + w_2 0 < 0\\
x^2 &:= (0,1)^T &:&\quad -1 + w_1 0 + w_2 1 < 0\\
x^3 &:= (1,0)^T &:&\quad -1 + w_1 1 + w_2 0 < 0\\
x^4 &:= (1,1)^T &:&\quad -1 + w_1 1 + w_2 1 \geq 0
\end{aligned}$$

Relevante Ungleichung erfüllt ?

Ja $\rightarrow$ J; Nein $\rightarrow$ N

$(w_1, w_2) \setminus x$	x^1	x^2	x^3	x^4	D
$(0,0)^T$	J	J	J	N	1
$(0,1)^T$	J	N	J	J	1
$(1,0)^T$	J	J	N	J	1
$(1,1)^T$	J	N	N	J	2

Visualisierung der Fehlerfunktion für f_{AND}



3.6 Perzeptron-Lernen

Überblick:

- Lernaufgabe
- Lernalgorithmus (Lernregel)
- Graphische Illustration Lernzyklus
- Beispiel einer Gewichtsadaption
- Äquivalente Lernregel
- Ablauf des Lernens
- Konvergenzsatz
- Beweis des Konvergenzsatzes

Lernaufgabe

Gegeben:

$\Omega_T := \mathcal{N} \cup \mathcal{P}$, Trainingselemente (x, y^{soll}), Zugehörigkeit $x \in \mathcal{P}$ oder $x \in \mathcal{N}$ ist bekannt ($\Rightarrow$ *überwachtes Lernen*).

Output $y^{t,ist}$ von rudimentärem Perzeptron (Zeitpunkt bzw. Adaptionsindex t) kann mit gefordertem Output y^{soll} verglichen werden. Es gilt $y^{t,ist}, y^{soll} \in \{-1, 1\}$.

Gesucht:

Gewichtsvektor w, der beide Mengen $\mathcal{P}, \mathcal{N}$ trennt.

Lernalgorithmus (Lernregel)

Gemäß Rosenblatt:

a. Falls $\boldsymbol{x} \in \boldsymbol{\mathcal{P}}$ und $\boldsymbol{y^{t,ist} = y^{soll}}$ $(= \mathbf{1})$ $\rightarrow$ Trennebene richtig
   $\rightarrow$ OK

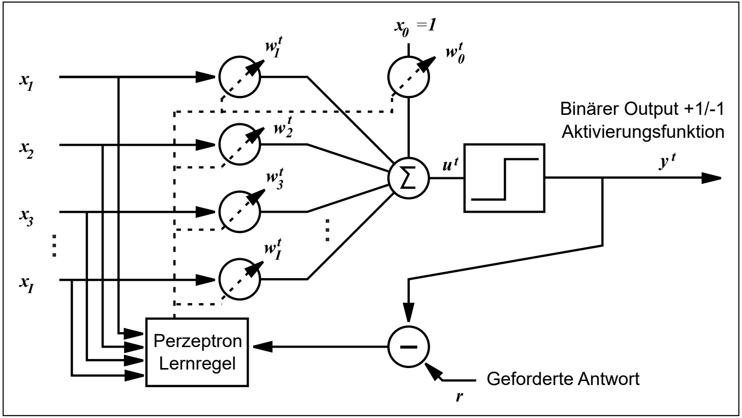
b. Falls $\boldsymbol{x} \in \boldsymbol{\mathcal{P}}$ und $\boldsymbol{y^{t,ist} = -y^{soll}}$ $(= \mathbf{-1})$ $\rightarrow$ Trennebene falsch
   $\rightarrow \boldsymbol{w^{t+1} := w^t + x}$

c. Falls $\boldsymbol{x} \in \boldsymbol{\mathcal{N}}$ und $\boldsymbol{y^{t,ist} = y^{soll}}$ $(= \mathbf{-1})$ $\rightarrow$ Trennebene richtig
   $\rightarrow$ OK

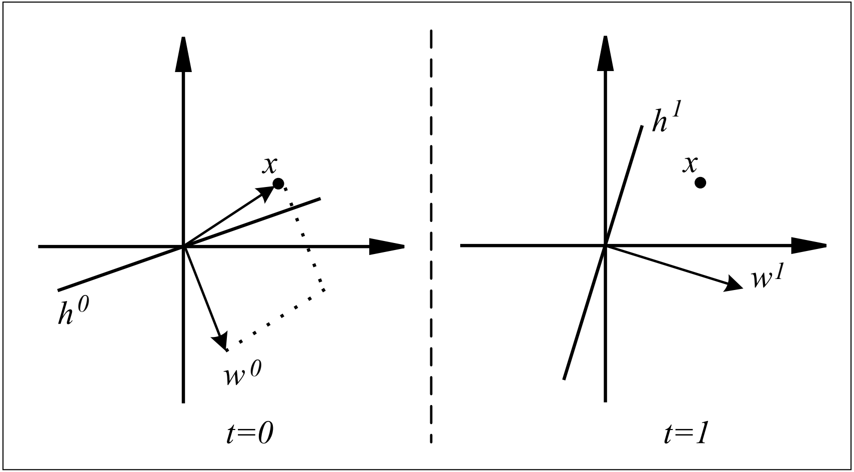
d. Falls $\boldsymbol{x} \in \boldsymbol{\mathcal{N}}$ und $\boldsymbol{y^{t,ist} = -y^{soll}}$ $(= \mathbf{1})$ $\rightarrow$ Trennebene falsch
   $\rightarrow \boldsymbol{w^{t+1} := w^t - x}$

Graphische Illustration Lernzyklus

Hinweis: Im Folgenden wird oft das Symbol $\boldsymbol{r}$ (für „required“) statt $\boldsymbol{y}^{soll}$ sowie das Symbol $\boldsymbol{y}^t$ statt $\boldsymbol{y}^{t,ist}$ verwendet, zur kompakteren Notation.



Beispiel einer Gewichtsadaption



Aquivalente Lernregel

Kompakte Notation der Lernregel:

$$\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + d^t \cdot y^{t,ist} \cdot \boldsymbol{x}\,,\ \text{mit}$$

$$d^t := \begin{cases} -1 & : \text{ falls } y^{t,ist} \neq y^{soll} \\ 0 & : \text{ sonst} \end{cases}$$

Alternative Formulierung der Lernregel:

$$\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + d^t \cdot \boldsymbol{\zeta}\,,\ \text{mit}$$

$$\boldsymbol{\zeta} := y^{soll} \cdot \boldsymbol{x}$$

$$d^t := \begin{cases} 1 & : \text{ falls } y^{t,ist} \neq y^{soll} \\ 0 & : \text{ sonst} \end{cases}$$

Ablauf des Lernens

a. Start mit zufälligem bzw. intelligent gewähltem Gewichtsvektor $\boldsymbol{w}^{\mathbf{0}}$.

b. Wiederholte Anwendung der Lernregel für Elemente aus $\boldsymbol{\Omega_T}$, bis keine Änderung in der Klassenzuordnung mehr feststellbar ist.

Konvergenzsatz

Konvergenzsatz:
Falls eine absolute lineare Trennung der Trainingsmenge $\boldsymbol{\Omega_T}$ möglich ist, dann wird nach *endlich* vielen Schritten ein Lösungsvektor gefunden.

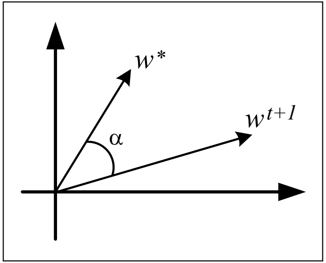
Beweis durch gegenteilige Annahme und Herleitung eines Widerspruchs:
Es tritt niemals der Fall ein, dass durch ein $\boldsymbol{w}^t$ alle Trainingselemente richtig klassifiziert werden, bzw. man benötigt *unendlich* viele Schritte.

Vereinbarungen:
$\boldsymbol{\Omega}'_T := \mathcal{P} \cup \mathcal{N}'$; $\boldsymbol{w}^*$ Lösungsvektor der Länge ℓ.

Beweis des Konvergenzsatzes

Falls $\boldsymbol{x}$ durch $\boldsymbol{w}^t$ nicht richtig klassifiziert wurde, dann Adaption
$\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + \zeta$.

Geometrische Relation zwischen $\boldsymbol{w}^{t+1}$ und $\boldsymbol{w}^*$.



$$\boldsymbol{w}^* \circ \boldsymbol{w}^{t+1} = \|\boldsymbol{w}^*\| \cdot \|\boldsymbol{w}^{t+1}\| \cdot \cos \boldsymbol{\alpha}$$

$$\cos \boldsymbol{\alpha} = \frac{\boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{t+1}}{\ell \cdot \|\boldsymbol{w}^{t+1}\|} \qquad (1)$$

Beweis des Konvergenzsatzes

Wir zeigen: Falls $\boldsymbol{w}^t$ unendlich oft adaptiert werden muss, und somit $\boldsymbol{t} \to \infty$, dann strebt der Quotient in Gleichung (1) gegen ∞.

Widerspruch!

Beweis des Konvergenzsatzes

Betrachte Zähler von (1):

$$\begin{aligned} \boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{t+1} &= \boldsymbol{w}^{*,T} \cdot (\boldsymbol{w}^{t} + \boldsymbol{\zeta}) \\ &= \boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{t} + \boldsymbol{w}^{*,T} \cdot \boldsymbol{\zeta} \\ &\geq \boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{t} + \boldsymbol{\delta} \end{aligned}$$

mit $\quad \boldsymbol{\delta} := \mathbf{min}_{\forall \zeta^m \in \Omega'_T} \{\boldsymbol{w}^{*,T} \cdot \boldsymbol{\zeta}^{m}\}$

Vollständige Induktion:

$$\boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{t+1} \geq \boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^{0} + (\boldsymbol{t} + \mathbf{1}) \cdot \boldsymbol{\delta} \tag{2}$$

Beweis des Konvergenzsatzes

Betrachte Nenner von (1):

$$\|\boldsymbol{w}^{t+1}\|^2 = (\boldsymbol{w}^t + \zeta)^T \cdot (\boldsymbol{w}^t + \zeta) = \|\boldsymbol{w}^t\|^2 + \mathbf{2} \cdot \boldsymbol{w}^{t,T} \cdot \zeta + \|\zeta\|^2$$

Hierbei gilt:

$\boldsymbol{w}^{t,T} \cdot \zeta \leq \mathbf{0}$, da andernfalls die Korrektur nicht nötig gewesen wäre.

Somit folgt:

$$\|\boldsymbol{w}^{t+1}\|^2 \leq \|\boldsymbol{w}^t\|^2 + \|\zeta\|^2 \leq \|\boldsymbol{w}^t\|^2 + \epsilon$$

mit $\quad \epsilon := \max_{\forall \zeta^m \in \Omega'_T}\{\|\zeta^m\|^2\}$

Beweis des Konvergenzsatzes

Vollständige Induktion:

$$\|\boldsymbol{w}^{t+1}\|^2 \leq \|\boldsymbol{w}^0\|^2 + (\boldsymbol{t}+1) \cdot \boldsymbol{\epsilon} \tag{3}$$

Aus (2) und (3) folgt:

$$\cos \alpha \geq \frac{\boldsymbol{w}^{*,T} \cdot \boldsymbol{w}^0 + (\boldsymbol{t}+1) \cdot \boldsymbol{\delta}}{\boldsymbol{\ell} \cdot \sqrt{\|\boldsymbol{w}^0\|^2 + (\boldsymbol{t}+1) \cdot \boldsymbol{\epsilon}}}$$

Für $\boldsymbol{t} \to \infty$ strebt der Quotient in obiger Ungleichung gegen ∞, also Widerspruch!

3.7 Verbesserungen beim Perzeptron-Lernen

Überblick:

- Wahl von Startgewichtsvektor
- Normierung Eingabe-/Gewichtsvektoren
- Lernregel mit zusätzlichem Adaptionsparameter
- Fehlerkorrigierende Lernregel mit zusätzl. Adaptionsparameter

Wahl von Startgewichtsvektor

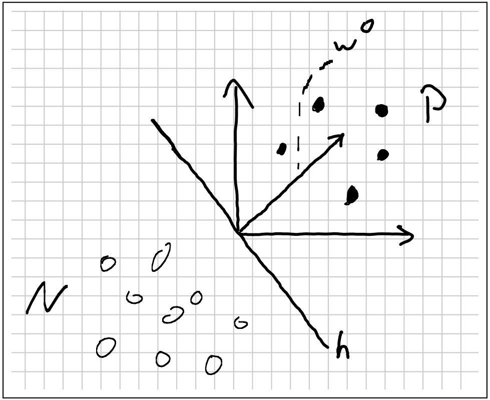
Beispiel: Schwerpunktvektor der Menge $\mathcal{P}$ wird Startgewichtsvektor.

$$\boldsymbol{w}^0 := \frac{1}{|\mathcal{P}|} \sum_{m=1}^{|\mathcal{P}|} \boldsymbol{x}^m \; ; \quad \text{für} \quad \boldsymbol{x}^m \in \mathcal{P}$$

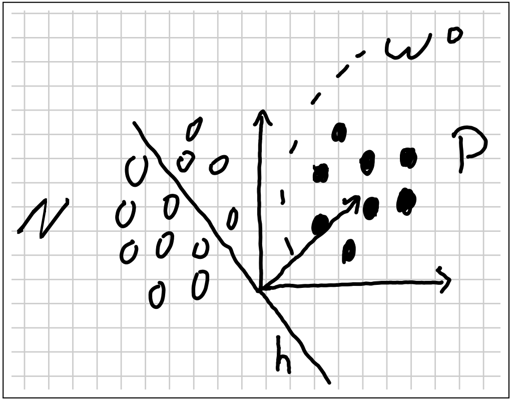
Wahl von Startgewichtsvektor

Startgewichtsvektor

optimal



nicht optimal



Wahl von Startgewichtsvektor

Beispiel: Differenzvektor zwischen den Schwerpunktvektoren von $\mathcal{P}$ und von $\mathcal{N}$ wird Startgewichtsvektor.

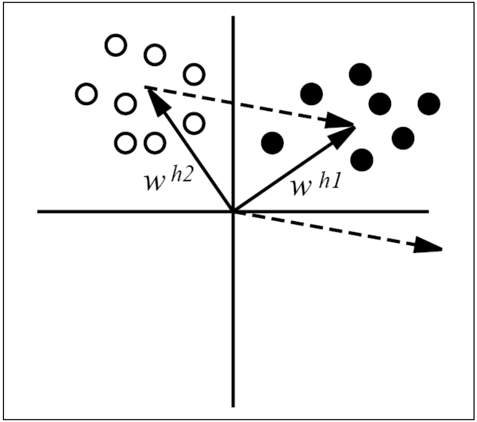
$$\boldsymbol{w}^0 := \underbrace{\frac{1}{|\mathcal{P}|} \sum_{m=1}^{|\mathcal{P}|} \boldsymbol{x}^m}_{\boldsymbol{w}^{h1}} - \underbrace{\frac{1}{|\mathcal{N}|} \sum_{n=1}^{|\mathcal{N}|} \boldsymbol{x}^{|\mathcal{P}|+n}}_{\boldsymbol{w}^{h2}};$$

für $\quad \boldsymbol{x}^m \in \mathcal{P},\ 1 \leq m \leq |\mathcal{P}|$

und $\quad \boldsymbol{x}^{|\mathcal{P}|+n} \in \mathcal{N},\ 1 \leq n \leq |\mathcal{N}|$

Wahl von Startgewichtsvektor

Illustration des Differenzvektors zwischen den Schwerpunktvektoren von $\mathcal{P}$ und von $\mathcal{N}$.



Normierung Eingabe-/Gewichtsvektoren

Problem: Falls $\boldsymbol{x} \in \boldsymbol{\mathcal{P}}$ und $\boldsymbol{y}^{t,ist} = -\mathbf{1}$,

- und falls speziell $||\boldsymbol{x}|| \gg ||\boldsymbol{w}^t||$, dann wäre

  $\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + \boldsymbol{x} \approx \boldsymbol{x}$,

  d.h. bisheriger Gewichtsvektor wird fast ignoriert;

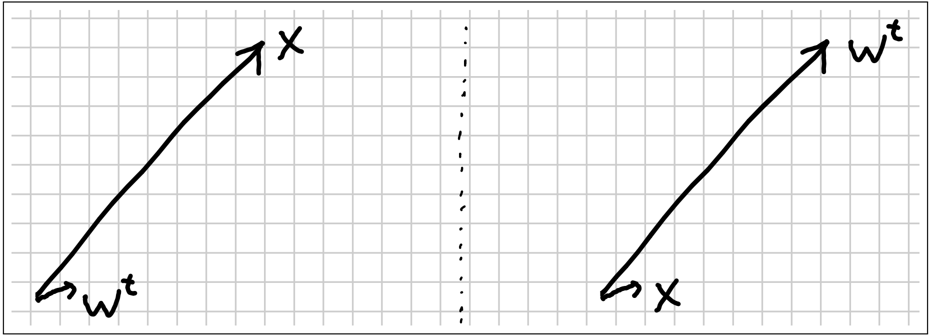
- oder falls speziell $||\boldsymbol{x}|| \ll ||\boldsymbol{w}^t||$, dann wäre

  $\boldsymbol{w}^{t+1} := \boldsymbol{w}^t + \boldsymbol{x} \approx \boldsymbol{w}^t$,

  d.h. kurzer Eingabevektor spielt geringe Rolle.

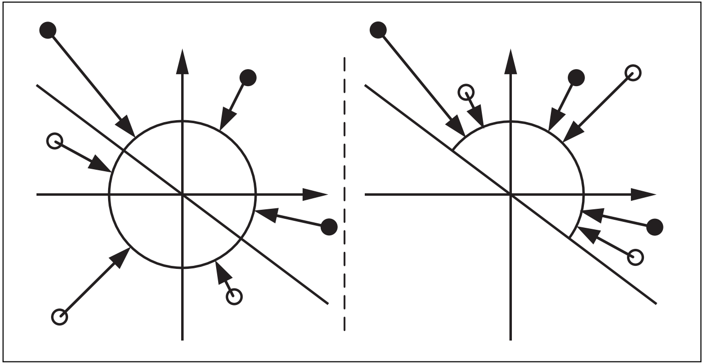
Normierung Eingabe-/Gewichtsvektoren

Bei extremen Längenunterschieden zwischen Eingabe- und Gewichtsvektoren spielt der kürzere Vektor bei der Vektoraddition eventuell eine zu geringe Rolle.



Normierung Eingabe-/Gewichtsvektoren

Durch Normierung der Eingabevektoren wird Klassifikationsaufgabe nicht verändert.



$\Omega_T := \mathcal{P} \cup \mathcal{N}$ $\qquad$ $\Omega'_T := \mathcal{P} \cup \mathcal{N}'$

Lernregel mit zusätzlichem Adaptionsparameter

Ein zusätzlicher Adaptionsparameter γ beeinflusst die Lerngeschwindigkeit (d.h. die Anzahl der nötigen Adaptionen).

Der Wert für diesen Parameter (ein sog. Hyperparameter) wird vom Benutzer vorab konstant festgelegt, und heisst Lernkonstante.

$$\boldsymbol{w}^{t+1} := \boldsymbol{w}^{t} + \gamma \cdot \boldsymbol{d}^{t} \cdot \zeta\ ; \qquad \mathbf{0} < \gamma \leq \mathbf{1}$$

Fehlerkorrigierende Lernregel mit zusätzlichem Adaptionsparameter

Mit der grundlegenden Perzeptron Lernregel ist eine einmalige Gewichtsadaption i.A. *nicht* fehlerkorrigierend.

Vorteilhaft wäre eine fehlerkorrigierende Lernregel!

Sei $\boldsymbol{x} \in \boldsymbol{\mathcal{P}}$ als normierter Vektor der Länge $\mathbf{1}$ angenommen.

Mit der aktuellen Gewichtung $\boldsymbol{w}^t$ gelte $\boldsymbol{w}^{t,T}\boldsymbol{x} \leq \mathbf{0}$, d.h. $\boldsymbol{y}^{t,ist} = -\mathbf{1}$.

Also wurde $\boldsymbol{x}$ falsch klassifiziert mit Abstand $\boldsymbol{\rho}^t := -\boldsymbol{w}^{t,T}\boldsymbol{x}$ von der Trennebene.

Fehlerkorrigierende Lernregel mit zusätzlichem Adaptionsparameter

Neue, fehlerkorrigierende Lernregel:

$$\boldsymbol{w}^{t+1} := \boldsymbol{w}^{t} + (\beta + \boldsymbol{\rho}^{t})\boldsymbol{x}\,; \text{ kleine positive Zahl } \beta$$

Nun wird $\boldsymbol{x}$ richtig klassifiziert, weil:

$$\begin{aligned} \boldsymbol{w}^{t+1,T}\boldsymbol{x} &= (\boldsymbol{w}^{t} + (\beta + \boldsymbol{\rho}^{t})\boldsymbol{x})^{T}\boldsymbol{x} \\ &= \boldsymbol{w}^{t,T}\boldsymbol{x} + (\beta + \boldsymbol{\rho}^{t})\underbrace{\|\boldsymbol{x}\|^{2}}_{=1} \\ &= \beta > 0 \end{aligned}$$

Fehlerkorrigierende Lernregel mit zusätzlichem Adaptionsparameter

Die Werte von β und $\boldsymbol{\rho}^t$ garantieren, dass im Gewichtsraum durch den Übergang von $\boldsymbol{w}^t$ nach $\boldsymbol{w}^{t+1}$ die Fehlerfunktion sinkt.

$\Rightarrow$ Gewichtsadaption mit unmittelbarer Fehlerbehebung.

Falls, abweichend von vorheriger Annahme, $\boldsymbol{x} \in \boldsymbol{\mathcal{N}}$, und jedoch $\boldsymbol{w}^{t,T}\boldsymbol{x} > \mathbf{0}$, d.h. $\boldsymbol{y}^{t,ist} = \mathbf{1}$. Dann wird alternativ definiert:

$$\boldsymbol{\rho}^t := \boldsymbol{w}^{t,T}\boldsymbol{x}; \quad \boldsymbol{w}^{t+1} := \boldsymbol{w}^t - (\boldsymbol{\beta} + \boldsymbol{\rho}^t)\boldsymbol{x}$$

Damit folgt $\boldsymbol{w}^{t+1,T}\boldsymbol{x} = -\boldsymbol{\beta} < \mathbf{0}$, und somit wird $\boldsymbol{x}$ nun richtig klassifiziert.

Fehlerkorrigierende Lernregel mit zusätzlichem Adaptionsparameter

Die beiden Ausprägungen der fehlerkorrigierenden Lernregel haben jeweils einen Adaptionsparameter β, und der Wert dieses Hyperparameters wird vom Benutzer vorab konstant festgelegt.

Der Parameter $\boldsymbol{\rho}^t$ ist dagegen variabel. Der Wert wird aufgrund der fehlerhaften Entscheidung des Perzeptrons für ein Trainingselement $\boldsymbol{x}$ automatisch durch dessen Abstand von der Trennebene festgelegt.