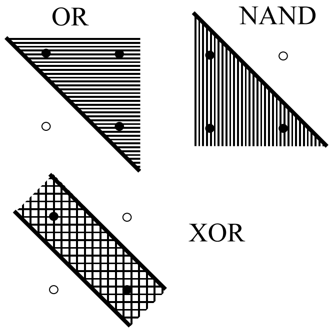
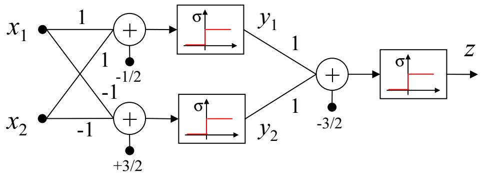
Lösung des XOR-Problems mit einem Zwei-Lagen-Perceptron

Das Exklusiv-Oder-Problem lässt sich hingegen im Gegensatz zu AND und OR *nicht* durch einen linearen Klassifikator lösen.

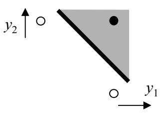
Eine Lösung ergibt sich jedoch durch Kombination zweier Trennlinien. Dies führt auf ein zweilagiges Perceptron!





		erste Schicht		zweite Schicht	
x_1	x_2	y_1	y_2	XOR	Klasse
0	0	0	1	0	ω_1
0	1	1	1	1	ω_2
1	0	1	1	1	ω_2
1	1	1	0	0	ω_1

$z = (x_1 \vee x_2) \wedge \overline{(x_1 \wedge x_2)} = y_1 \wedge \overline{y_2}$ XOR



$y_1 = \sigma(x_1 + x_2 - \frac{1}{2})$ OR

$y_2 = \sigma(-x_1 - x_2 + \frac{3}{2})$ NAND

$z = \sigma(y_1 + y_2 - \frac{3}{2})$ AND

Nach der Abbildung der ersten Schicht ergibt sich ein linear trennbares Problem!!

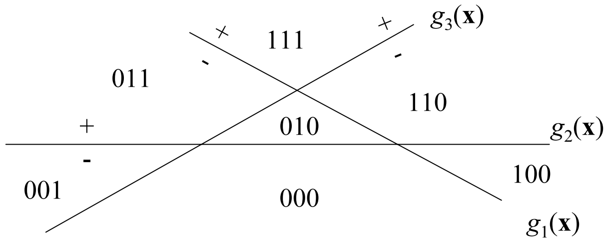
Funktionsweise der ersten Schicht: Abbildung des rellwertigen Eingangsraums auf die Eckpunkte eines Hyperwürfels

In der ersten Schicht eines Perceptrons reagiert ein Neuron mit 0 oder 1, in Abhängigkeit davon, in welcher Hälfte der von der Hyperfläche erzeugten beiden Halbräume sich der Eingangsvektor befindet.

Jedes weitere Neuron erzeugt eine zusätzliche Trennebene. Die erste Schicht eines Perceptrons bildet somit den reellwertigen Eingangsraum auf die Eckpunkte eines Hyperwürfels ab.

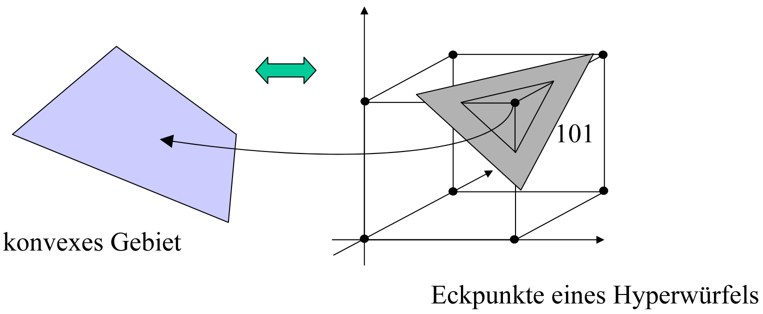
Die sich schneidenden Trennebenen bilden linear begrenzte (konvexe) ***Gebiete***, sogenannte ***Polyeder***. Jedes Gebiet korrespondiert zu einem Eckpunkt des Hyperwürfels.

Nachfolgend ist eine Grafik zu sehen für 3 Neuronen und einen zweidimensionalen Eingangsraum:



Funktionsweise der zweiten Schicht: Abtrennung eines Eckpunktes des Hyperwürfels

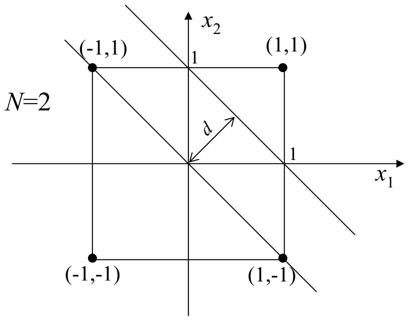
In der zweiten Schicht kann durch Abtrennung eines Eckpunktes des Hyperwürfels mit einer neuen Trennfläche ein eindeutiges Merkmal für ein konvexes Gebiet erzeugt werden:



Optimale Trennebene zur Abtrennung eines Eckpunktes von einem Hyperwürfel

Wird als nichtlineare Funktion im Neuron die Signum-Funktion verwendet, so liegt der daraus resultierende Hyperwürfel zentriert im Ursprung. Die optimale Trennebene liegt in der Mitte zwischen einem Eckpunkt und einer Ebene die von den nächsten Nachbarn aufgespannt wird.

Ohne Einschränkung der Allgemeinheit, können wir den Eckpunkt entlang der ersten Raumdiagonale **u** betrachten. Dafür gilt:



$$\mathbf{u}^T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \end{bmatrix} \quad \text{mit:} \quad \dim(\mathbf{x}) = N$$

$$\Rightarrow \mathbf{e}_u = \mathbf{u} / \|\mathbf{u}\| = \mathbf{u} / \sqrt{N}$$

Eigentlich genügt es, die Trennebene in einem Abstand $d=N^{1/2}$-ε mit einem hinreichend kleinen Abstand ε zu platzieren. Problem: über ε kann keine genaue Aussage gemacht werden.

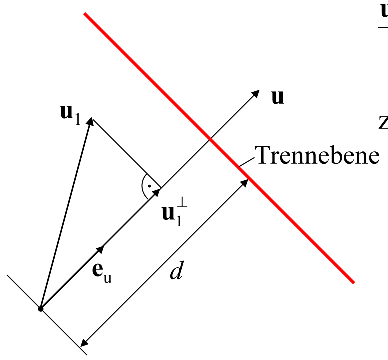
Die Trennebene liegt genau zwischen **u** und der orthogonalen Projektion eines der Nachbarecken $\mathbf{u}_1$ auf **u**:

$$\mathbf{u}_1^T = [\underbrace{1 \quad 1 \quad \dots \quad 1}_{N-1} \quad -1]$$

$$\mathbf{u}_1^{\perp} = \underbrace{\langle \mathbf{u}_1, \mathbf{u} \rangle}_{N-2} \, 1/\sqrt{N}\, \mathbf{e}_u^T = \frac{N-2}{N}\mathbf{u}$$

und damit der Fußpunkt der Trennebene bei:

$$\frac{\mathbf{u}_1^{\perp} + \mathbf{u}}{2} = \tfrac{1}{2}\left(\frac{N-2}{N} + 1\right)\mathbf{u} = \left(1 - \frac{1}{N}\right)\mathbf{u} = \frac{N-1}{N}\mathbf{u} = \underbrace{\frac{N-1}{\sqrt{N}}}_{d}\mathbf{e}_u$$



z. Bsp. $N = 3 \Rightarrow 2/3\,\mathbf{u}$

$N = 4 \Rightarrow 3/4\,\mathbf{u}$

$N = 100 \Rightarrow 0{,}99\,\mathbf{u}$

Ein zu **u** benachbarter Eckpunkt des Würfels produziert im Skalarprodukt $<\mathbf{u}_1,\mathbf{u}>$ genau den Wert (N-2) !

Die Gleichung der Trennebene ergibt sich zu:

$$\langle \mathbf{x}, \mathbf{e}_u \rangle = d = \frac{N-1}{\sqrt{N}}$$

$$\Rightarrow \langle \mathbf{x}, \mathbf{u} \rangle \frac{1}{\sqrt{N}} = \frac{N-1}{\sqrt{N}}$$

$$\Rightarrow \boxed{\langle \mathbf{x}, \mathbf{u} \rangle = (N-1)}$$

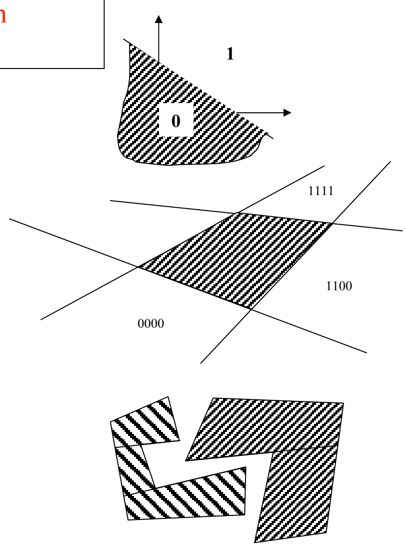
Das 3-Lagen-Perceptron mit Schwellwertfunktion

Mit einem 3-L-S-MLPC lassen sich beliebige linear begrenzte Cluster in Merkmalsräumen klassifizieren!!

1. Schicht: sich schneidende Hyperebenen bilden ***konvexe Polyeder*.** Diese werden auf die ***Eckpunkte eines Hyperwürfels*** abgebildet.

2. Schicht: Durch abtrennen von Eckpunkten des Hyperwürfels werden ***konvexe Polyeder*** durch Schnittmengenbildung von Halbräumen selektiert (AND).

3. Schicht: ***beliebige linear begrenzte Gebiete*** entstehen durch Vereinigung von konvexen Gebieten (OR).

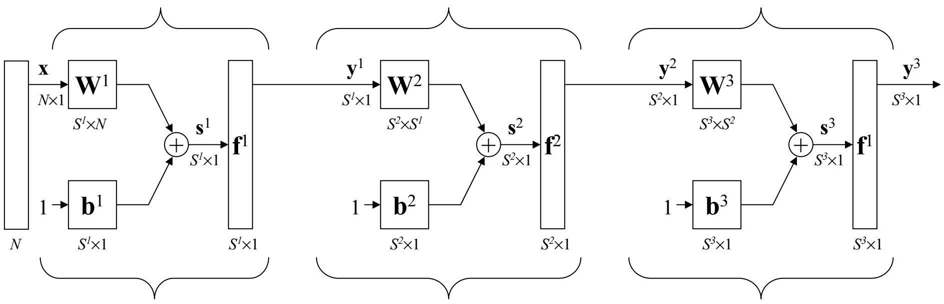


Perceptron mit 3 Lagen zur Realisierung eines Klassifikators für beliebig linear begrenzte Gebiete (Polyeder)

erste Schicht: linear begrenzte Gebiete werden auf Hyperwürfel abgebildet

zweite Schicht: Schnittmengen von Halbräumen (Polyeder) werden Booleschen Vektoren zugeordnet (AND)

dritte Schicht Vereinigung von konvexen Gebieten (OR)



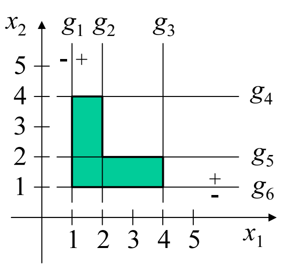
$\mathbf{y}^1=\mathbf{f}^1(\mathbf{W}^1\mathbf{x}+\mathbf{b}^1)$ $\quad$ $\mathbf{y}^2=\mathbf{f}^2(\mathbf{W}^2\mathbf{y}^1+\mathbf{b}^2)$ $\quad$ $\mathbf{y}^3=\mathbf{f}^3(\mathbf{W}^3\mathbf{y}^2+\mathbf{b}^3)$

$$\mathbf{y}^3=\mathbf{f}^3(\mathrm{W}^3\mathrm{f}^2(\mathrm{W}^2\mathrm{f}^1(\mathrm{W}^1\mathrm{x}+\mathbf{b}^1)+\mathbf{b}^2)+\mathbf{b}^3)$$

Beispiel: Selektion eines nichtkonvexen Gebietes mit einem dreilagigen Perceptron mit $f(s)$=sign(s)

$$\boxed{\mathbf{y}^3=\mathbf{f}^3(\mathrm{W}^3\mathrm{f}^2(\mathrm{W}^2\mathrm{f}^1(\mathrm{W}^1\mathrm{x}+\mathbf{b}^1)+\mathbf{b}^2)+\mathbf{b}^3)}$$

Aufgabe:



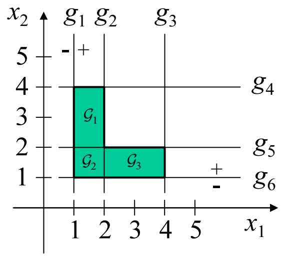
Entwurf der ersten Schicht: Definition der Hyperebenen

$$\mathbf{W}^1=\begin{bmatrix} w_{1,1}=1 & w_{1,2}=0 \\ w_{2,1}=1 & w_{2,2}=0 \\ w_{3,1}=1 & w_{3,2}=0 \\ w_{4,1}=0 & w_{4,2}=1 \\ w_{5,1}=0 & w_{5,2}=1 \\ w_{6,1}=0 & w_{6,2}=1 \end{bmatrix}\begin{array}{l} \leftarrow g_1 \\ \leftarrow g_2 \\ \leftarrow g_3 \\ \leftarrow g_4 \\ \leftarrow g_5 \\ \leftarrow g_6 \end{array} \qquad \mathbf{b}^1=\begin{bmatrix} -1 \\ -2 \\ -4 \\ -4 \\ -2 \\ -1 \end{bmatrix}$$

z.Bsp.: $g_1: \ \langle \mathbf{w}^1, \mathbf{x} \rangle + b^1 = 0$ mit: $\mathbf{w}^1=\begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad b^1=-1$

Entwurf der zweiten Schicht:

Aufgabe:



Entwurf der zweiten Schicht:
Abtrennen von Eckpunkten des Hyperwürfels

Kennzeichnung der Gebiete:

	g_1	g_2	g_3	g_4	g_5	g_6
$\mathcal{G}_1$	+	−	−	−	+	+
$\mathcal{G}_2$	+	−	−	−	−	+
$\mathcal{G}_3$	+	+	−	−	−	+

$$\mathbf{W}^2 = \begin{array}{c} \begin{array}{cccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 \end{bmatrix} \end{array} \begin{array}{l} \leftarrow \mathcal{G}_1 \\ \leftarrow \mathcal{G}_2 \\ \leftarrow \mathcal{G}_3 \end{array} \quad \mathbf{b}^2 = \begin{bmatrix} -5 \\ -5 \\ -5 \end{bmatrix}$$

Die Zeilenvektoren von $\mathbf{W}^2$ zeigen genau auf die abzutrennenden Eckpunkte; der Schwellwert hat den Wert N-1

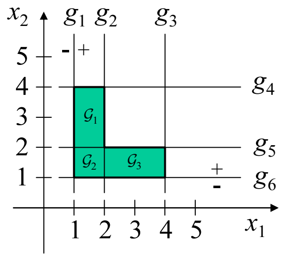
Entwurf der dritten Schicht:

In der dritten Schicht ist die Vereinigungsmenge der Gebiete zu realisieren. Dies geschieht einfach mit einer OR-Verknüpfung. Das bedeutet, dass der Eckpunkt [–1 –1 –1] abgetrennt werden muss. Dies wiederum geschieht mit Hilfe von:

$$\mathbf{W}^3 = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix} \qquad b^3 = -2$$

Vereinfachter Entwurf der zweiten und dritten Schicht:

Aufgabe:



Entwurf der zweiten Schicht:
Abtrennen von Eckpunkten des Hyperwürfels

Kennzeichnung der Gebiete: die Tabelle enthält „don‘t care“-Elemente ($*$). Die dazugehörigen Neuronen können unberücksichtigt bleiben (Verbindung auftrennen)

	g_1	g_2	g_3	g_4	g_5	g_6
$\mathcal{F}_1 = \mathcal{G}_1 \cup \mathcal{G}_2$	$+$	$-$	$*$	$-$	$*$	$+$
$\mathcal{F}_2 = \mathcal{G}_2 \cup \mathcal{G}_3$	$+$	$*$	$-$	$*$	$-$	$+$

$$\mathbf{W}^2 = \begin{array}{c} \begin{array}{cccccc} g_1 & g_2 & g_3 & g_4 & g_5 & g_6 \end{array} \\ \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & -1 & 1 \end{bmatrix} \end{array} \begin{array}{l} \leftarrow \mathcal{F}_1 \\ \leftarrow \mathcal{F}_2 \end{array} \qquad \mathbf{b}^2 = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$$

Die Dimension des reduzierten Würfels beträgt nur N=4 (ohne „don‘t care Elemente!)

Die Zeilenvektoren von $\mathbf{W}^2$ zeigen genau auf die abzutrennenden Eckpunkte; der Schwellwert hat den Wert N-1 des reduzierten Würfels!

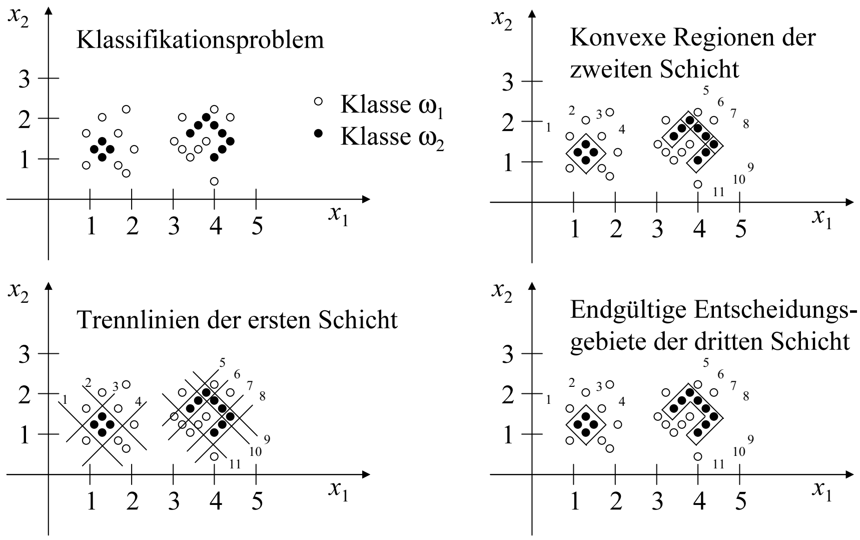
Vereinfachter Entwurf der dritten Schicht:

In der dritten Schicht ist die Vereinigungsmenge der Gebiete zu realisieren. Dies geschieht einfach mit einer OR-Verknüpfung. Das bedeutet, dass der Eckpunkt [–1 –1] abgetrennt werden muss. Dies wiederum geschieht mit Hilfe von:

$$\mathbf{W}^3 = \begin{bmatrix} -1 & -1 \end{bmatrix} \qquad b^3 = -1$$

Beispiel für ein nicht linear separierbares Zwei-Klassenproblem

(M.T. Hagan, H.B. Demuth, M. Beale „Neural Network Design“, PWS Publishing Company, Boston, 1995)



Beispiel für ein nicht linear separierbares Zwei-Klassenproblem

Erste Schicht: es werden 11 Geraden zur Abtrennung der Gebiete als Funktion von zwei Eingangsvariablen benötigt!

$$\mathbf{W}^{1^T} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{b}^{1^T} = \begin{bmatrix} -2 & 3 & 0.5 & 0.5 & -1.75 & 2.25 & -3.25 & 3.75 & 6.25 & -5.75 & -4.75 \end{bmatrix}$$

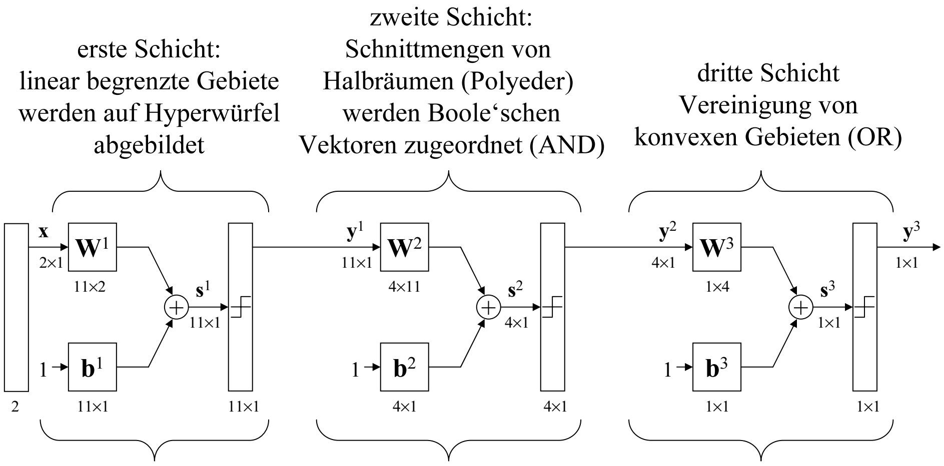
Zweite Schicht: es werden vier konvexe Gebiete selektiert!

$$\mathbf{W}^2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \quad \mathbf{b}^2 = \begin{bmatrix} -3 \\ -3 \\ -3 \\ -3 \end{bmatrix}$$

Dritte Schicht:

$$\mathbf{W}^3 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \quad b^3 = 3$$

Beispiel für ein nicht linear separierbares Zwei-Klassenproblem



$\mathbf{y}^1=\mathbf{sign}(\mathbf{W}^1\mathbf{x}+\mathbf{b}^1)$ $\quad$ $\mathbf{y}^2=\mathrm{sign}(\mathbf{W}^2\mathbf{y}^1+\mathbf{b}^2)$ $\quad$ $\mathbf{y}^3=\mathrm{sign}(\mathbf{W}^3\mathbf{y}^2+\mathbf{b}^3)$

$$\mathbf{y}^3=\mathrm{sign}(W^3\mathrm{sign}(W^2\mathrm{sign}(W^1x+\mathbf{b}^1)+\mathbf{b}^2)+\mathbf{b}^3)$$

Auf dem Weg zum automatischen Entwurf eines Neuronalen Netzes

Die bisher skizzierte Vorgehensweise ist sehr anschaulich und im zweidimensionalen Fall noch intuitiv handhabbar, aber für höherdimensionale Fälle ist sie nicht brauchbar. Alles was wir in der Praxis zur Verfügung haben, sind die Lernstichproben. Benötigt wird ein automatischer Algorithmus, welcher davon ausgehend in einem Lernprozess die optimalen Gewichte des NN automatisch ermittelt.

Will man für diese nichtlineare Optimierungsaufgabe iterative Algorithmen einsetzen, so benötigt man differenzierbare Nichtlinearitäten in den Neuronen (die Schwellwertfunktionen sind dafür nur begrenzt geeignet!).

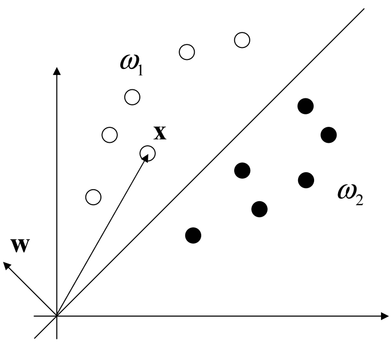
Linear separierbare Klassen – der Perceptron-Algorithmus

Ziel ist, ein Algorithmus zur *automatischen* Anpassung der unbekannten Gewichtsmatrizen $\mathbf{W}^i$ eines Perceptrons an ein Klassifikationsproblem, gegeben eine endliche Stichprobe $\{(\mathbf{x}_i,\omega_k)\}$.
Für den Fall linear separierbarer Klassen gab Rosenblatt einen iterativen Algorithmus zur Lösung dieses Problems an. Es wird angenommen, dass eine lineare Trennfläche, definiert durch die Gleichung $\mathbf{w}^{*T}\mathbf{x}=0$ für das Zweiklassenproblem existiert, derart dass:

$$\mathbf{w}^{*T}\,\mathbf{x} > 0 \quad \forall \mathbf{x} \in \omega_1$$
$$\mathbf{w}^{*T}\,\mathbf{x} < 0 \quad \forall \mathbf{x} \in \omega_2$$

Dies schließt auch den Fall ein, dass die Trennfläche nicht wie hier durch den Ursprung läuft, nämlich $\mathbf{w}^{*T}\mathbf{x}+b^*=0$, da dieser durch Erweiterung von $\mathbf{x}$ auf die obige Formulierung zurückgeführt werden kann:

$$\mathbf{x}'^T = \left[1, \mathbf{x}^T\right] \quad \mathbf{w}'^{*T} = \left[b^*, \mathbf{w}^{*T}\right]$$
$$\Rightarrow \mathbf{w}'^{*T}\mathbf{x}' = \mathbf{w}^{*T}\mathbf{x} + b^*$$



Linear separierbares Problem