UNIVERSITÄT DORTMUND

Wintersemester 2006/07

Fundamente der Computational Intelligence

(Vorlesung)

Prof. Dr. Günter Rudolph

Fachbereich Informatik

Lehrstuhl für Algorithm Engineering

Inhalt

- Biologisches Vorbild
- McCulloch-Pitts Netze
- Perceptron (Minsky/Papert)

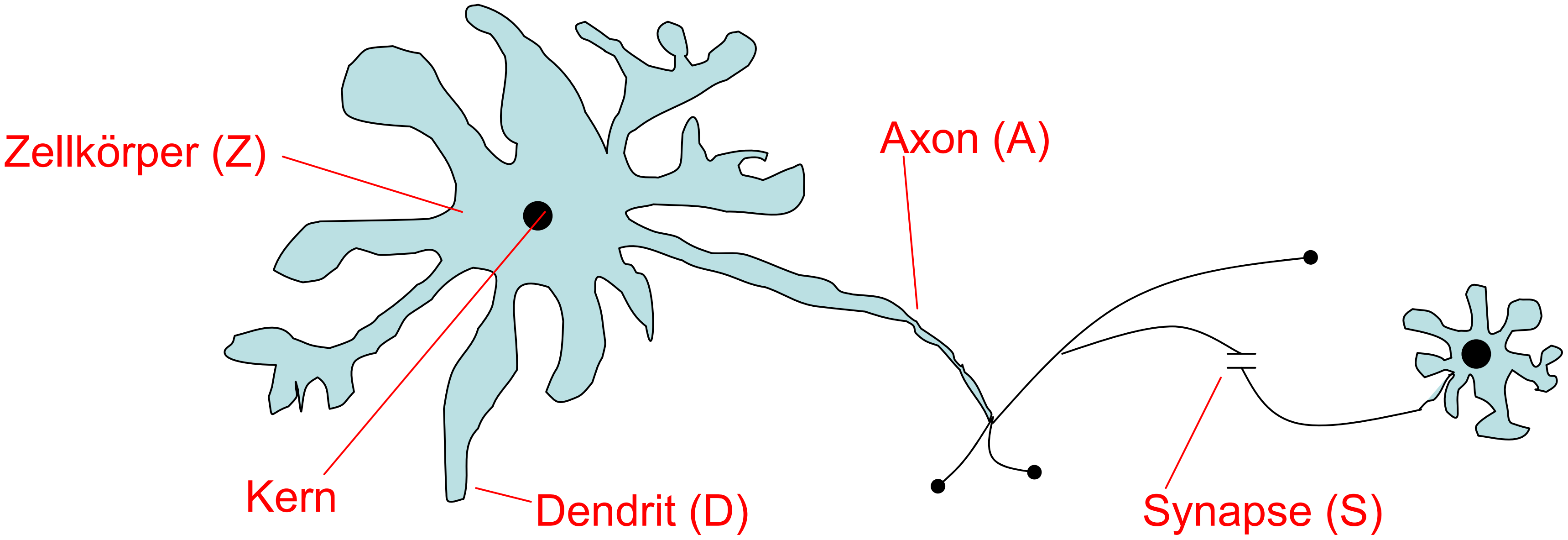
Biologisches Vorbild

- Neuronen
  - Information aufnehmen (D)
  - Information verarbeiten (Z)
  - Information weiterleiten (A / S)

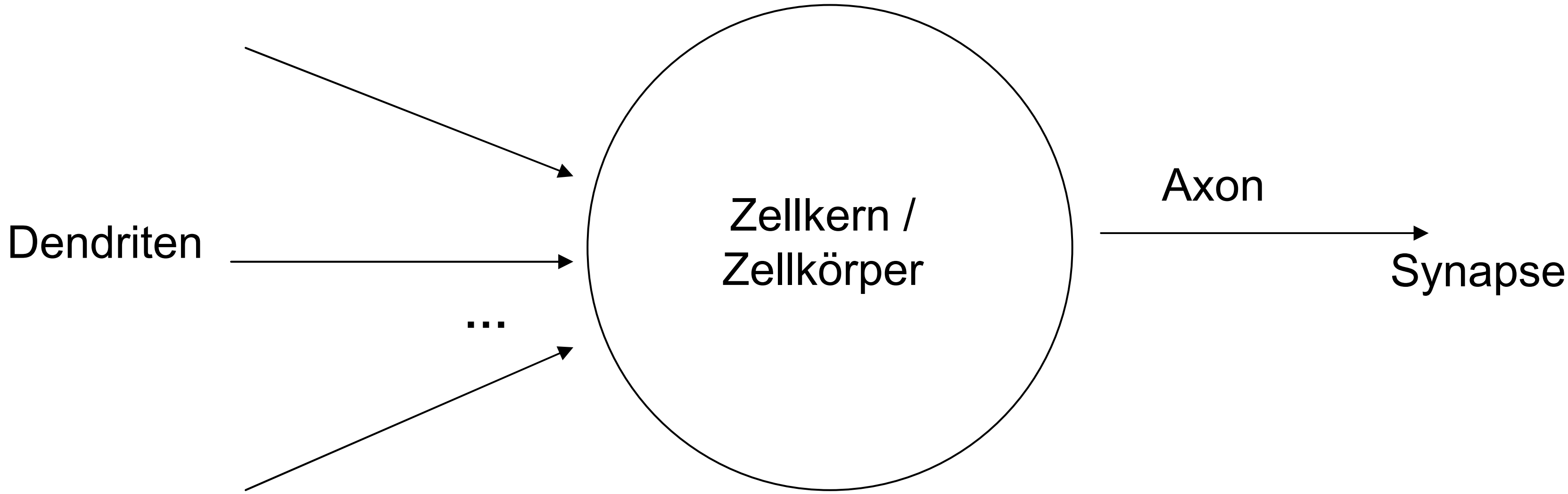
Mensch: 10^{12} Neuronen

Strom im mV-Bereich

120 m / s



Abstraktion

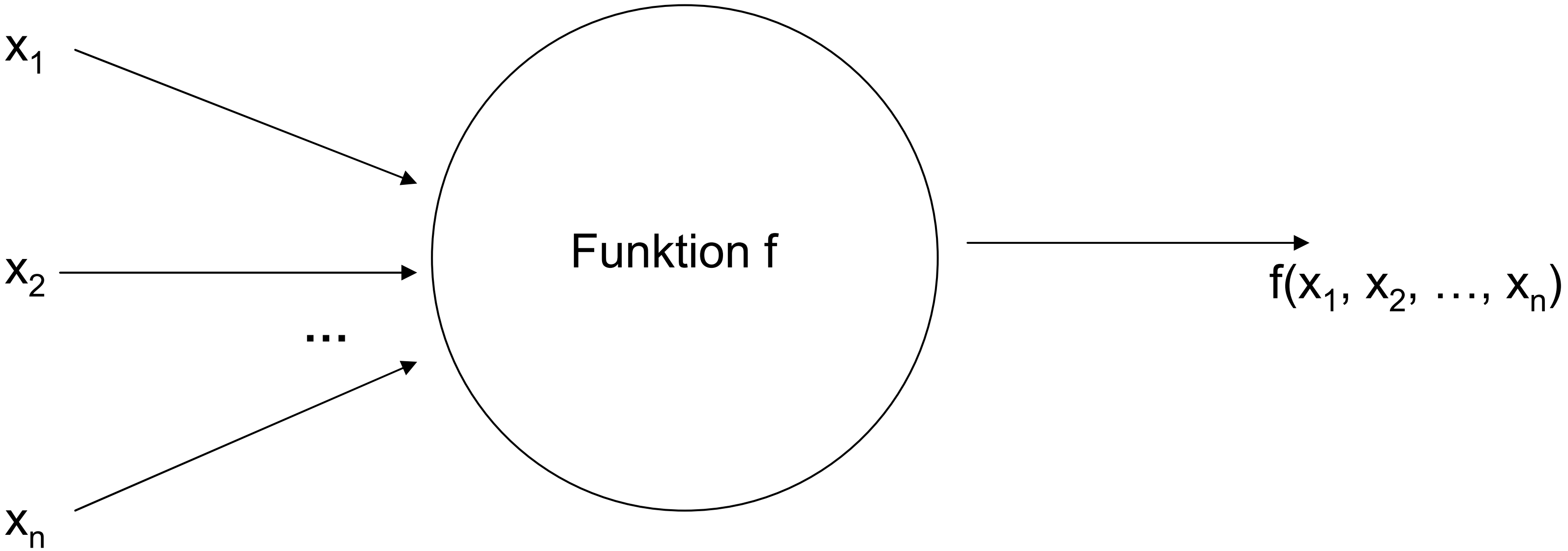


Signal-Eingabe

Signal-Verarbeitung

Signal-Ausgabe

Modell



McCulloch-Pitts-Neuron 1943:

$x_i \in \{ 0, 1 \} =: \mathbb{B}$

$f: \mathbb{B}^n \rightarrow \mathbb{B}$

1943: Warren McCulloch / Walter Pitts

- Beschreibung neurologischer Netzwerke
  → Modell: McCulloch-Pitts-Neuron

- Grundidee:

  - Neuron ist entweder aktiv oder inaktiv

  - Fähigkeiten entstehen durch Vernetzung der Neuronen

- Betrachtung statischer Netzwerke
  (d.h. Verbindungen werden konstruiert und nicht gelernt)

McCulloch-Pitts-Neuron

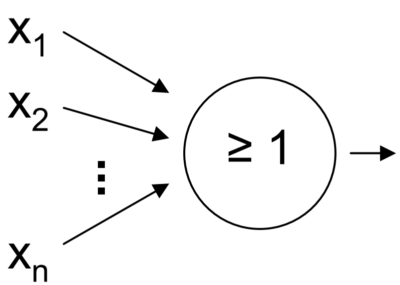
n binäre Eingangssignale $x_1, \ldots, x_n$

Schwellwert $\theta > 0$

$$f(x_1, \ldots, x_n) = \begin{cases} 1 & \text{falls } \sum_{i=1}^{n} x_i \geq \theta \\ 0 & \text{sonst} \end{cases}$$

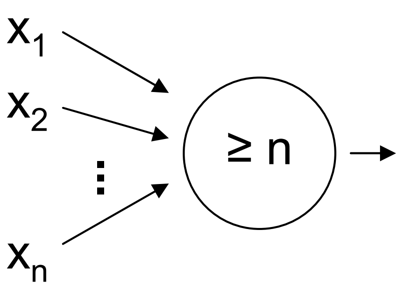
⇒ realisierbar:

Boolsches OR



$\theta = 1$

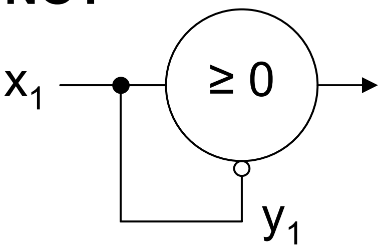
Boolsches AND



$\theta = n$

McCulloch-Pitts-Neuron

NOT



n binäre Eingangssignale $x_1, \ldots, x_n$

Schwellwert $\theta > 0$

<u>zusätzlich:</u> m binäre hemmende Signale $y_1, \ldots, y_m$

$$\tilde{f}(x_1, \ldots, x_n; y_1, \ldots, y_m) = f(x_1, \ldots, x_n) \cdot \prod_{j=1}^{m} (1 - y_j)$$

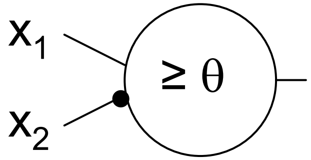
- ist ein $y_j = 1$, dann ist Ausgabe = 0
- sonst:
  - Summe der Eingänge ≥ Schwellwert, dann Ausgabe = 1
    sonst Ausgabe = 0

Analogien

Neuronen	einfache MISO-Prozessoren (mit Parametern: z.B. Schwellwert)
Synapsen	Verbindungen zwischen Neuronen (mit Parametern: Synapsengewicht)
Topologie	Verbindungsstruktur des Netzes
Propagierung	Arbeitsphase des KNN → wandelt Input zu Output
Training / Lernen	Anpassung des KNN an bestimmte Daten

Annahme:

Eingaben liegen auch invertiert vor bzw. ∃ Invertierungseingänge.

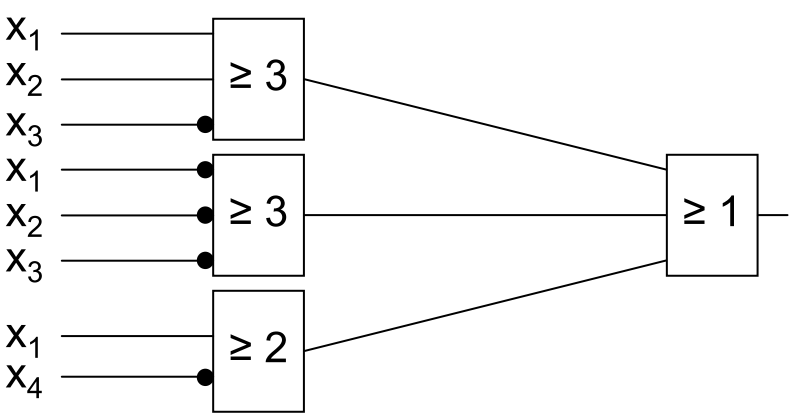


$\Rightarrow x_1 + \bar{x}_2 \geq \theta$

Satz:

Jede logische Funktion F: $\mathbb{B}^n \rightarrow \mathbb{B}$ kann durch ein zweischichtiges McCulloch/Pitts-Netz berechnet werden.

Beispiel: $$F(x) = x_1 x_2 \bar{x}_3 \vee \bar{x}_1 \bar{x}_2 \bar{x}_3 \vee x_1 \bar{x}_4$$



Beweis: (konstruktiv)

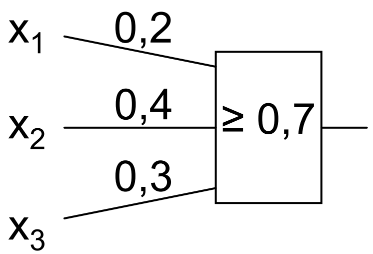
Jede boolsche Funktion F lässt sich in disjunktive Normalform bringen

$\Rightarrow$ 2 Schichten (AND - OR)

1. Jede Klausel erhält ein Dekodierungsneuron mit $\theta = n$
   $\Rightarrow$ Ausgabe = 1 nur falls Klausel erfüllt (AND-Verknüpfung)
2. Alle Ausgänge der Dekodierungsneuronen
   sind Eingänge eines Neurons mit $\theta = 1$ (OR-Verknüpfung)

q.e.d.

Verallgemeinerung: Eingänge mit Gewichten



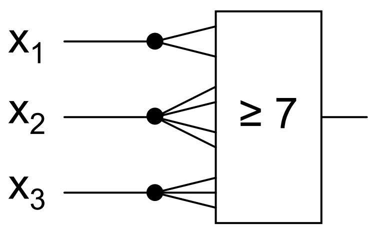
feuert 1 falls

$$0{,}2\,x_1 + 0{,}4\,x_2 + 0{,}3\,x_3 \geq 0{,}7 \quad \Big|\cdot 10$$

$$2\,x_1 + 4\,x_2 + 3\,x_3 \geq 7$$

$$\Downarrow$$

dupliziere Eingänge!



$\Rightarrow$ äquivalent!

Satz:

Gewichtete und ungewichtete MCP-Netze sind äquivalent für Gewichte $\in \mathbb{Q}^+$.

Beweis:

„⇒“ Sei $$\sum_{i=1}^{n} \frac{a_i}{b_i} x_i \geq \frac{a_0}{b_0} \quad \text{mit } a_i, b_i \in \mathbb{N}$$

Multiplikation mit $\prod_{i=0}^{n} b_i$ ergibt Ungleichung mit Koeffizienten in $\mathbb{N}$

Dupliziere Eingang x_i, so dass $a_i \, b_1 \, b_2 \cdots b_{i-1} \, b_{i+1} \cdots b_n$ viele Eingänge vorhanden.

Schwellwert $\theta = a_0 \, b_1 \cdots b_n$

„⇐“

Setze alle Gewichte auf 1. q.e.d.

Kapitel 5: Künstliche neuronale Netze

Perzeptron (Rosenblatt 1958)

→ komplexes Modell → reduziert von Minsky & Papert auf das „Notwendigste“

→ Minsky-Papert-Perzeptron (MPP), 1969

Was leistet ein MPP?

$$w_1 x_1 + w_2 x_2 \geq \theta \quad \begin{cases} \xrightarrow{J} 1 \\ \xrightarrow{N} 0 \end{cases}$$

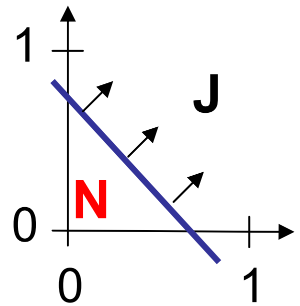
umstellen nach x_2 liefert:

$$x_2 \geq \frac{\theta}{w_2} - \frac{w_1}{w_2} x_1 \quad \begin{cases} \xrightarrow{J} 1 \\ \xrightarrow{N} 0 \end{cases}$$

Bsp:

$$0,9\, x_1 + 0,8\, x_2 \geq 0,6$$

$$\Leftrightarrow \quad x_2 \geq \frac{3}{4} - \frac{9}{8} x_1$$

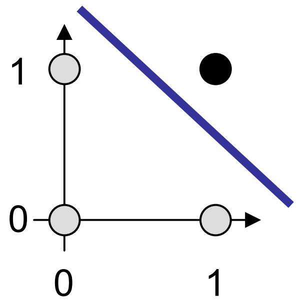


Trenngerade

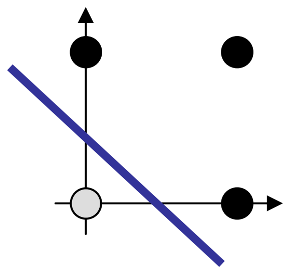
separiert $\mathbb{R}^2$

in 2 Klassen

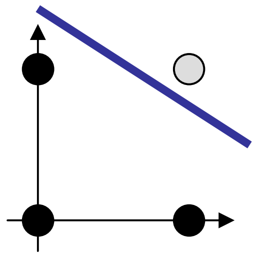
AND



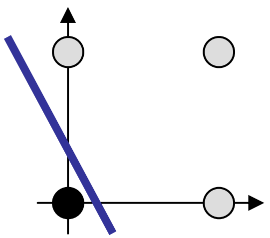
OR



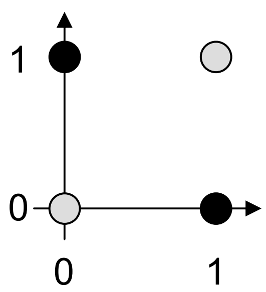
NAND



NOR



XOR



?

x_1	x_2	xor
0	0	0
0	1	1
1	0	1
1	1	0

$\Rightarrow 0 < \theta$

$\Rightarrow w_2 \geq \theta$

$\Rightarrow w_1 \geq \theta$

$\Rightarrow w_1 + w_2 < \theta$

$w_1, w_2 \geq \theta > 0$

$\Rightarrow w_1 + w_2 \geq 2\theta$

Widerspruch!

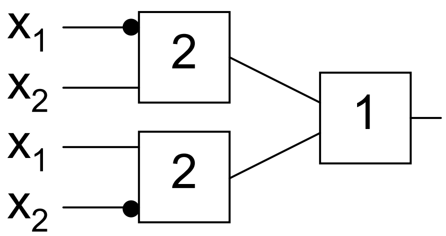
$$w_1 x_1 + w_2 x_2 \geq \theta$$

1969: Marvin Minsky / Seymor Papert

- Buch *Perceptrons* → Analyse math. Eigenschaften von Perzeptrons
- Ernüchterndes Ergebnis:
  Triviale Probleme können nicht mit Perzeptrons gelöst werden!
  - XOR-Problem
  - Parity-Problem
  - Connectivity-Problem
- „Folgerung“: Alle künstliche Neuronen haben diese Schwäche!
  ⇒ Forschung auf diesem Gebiet ist wissenschaftliche Sackgasse!
- Folge: Forschungsförderung bzgl. KNN praktisch eingestellt (~ 15 Jahre)

Wege aus der „Sackgasse“:

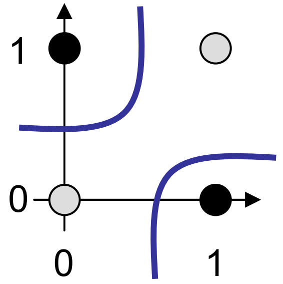
1. Mehrschichtige Perzeptrons:



$\Rightarrow$ realisiert XOR

2. Nichtlineare Trennfunktionen:

XOR



$g(x_1, x_2) = 2x_1 + 2x_2 - 4x_1x_2 - 1$ mit $\theta = 0$

$g(0,0) = -1$
$g(0,1) = +1$
$g(1,0) = +1$
$g(1,1) = -1$

Wie kommt man zu den Gewichten und θ ?

bisher: durch Konstruktion

Bsp: NAND-Gatter

x_1	x_2	**NAND**	
0	0	1	$\Rightarrow 0 \geq \theta$
0	1	1	$\Rightarrow w_2 \geq \theta$
1	0	1	$\Rightarrow w_1 \geq \theta$
1	1	0	$\Rightarrow w_1 + w_2 < \theta$

erfordert Lösung eines linearen Ungleichungssystems ($\in$ P)

(Bsp: $w_1 = w_2 = -2$, $\theta = -3$)

jetzt: durch „Lernen" bzw. Trainieren

Perceptron-Lernen

Annahme: Testbeispiele mit richtigem Ein-/Ausgabeverhalten bekannt

Prinzip:

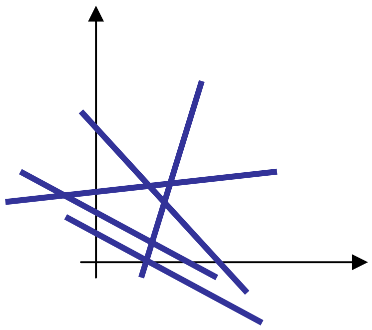
(1) wähle Gewichte irgendwie

(2) lege Testmuster an

(3) falls Perceptronausgabe falsch, dann verändere Gewichte

(4) gehe nach (2) bis richtige Perceptronausgabe für alle Testmuster

grafisch:



→ Verschieben und Drehen der Trenngeraden

Perceptron-Lernen

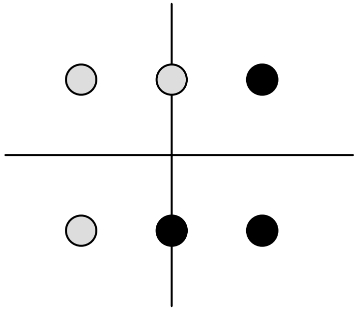
P: Menge positiver Beispiele
N: Menge negativer Beispiele

1. wähle w_0 zufällig, $t = 0$
2. wähle zufällig $x \in P \cup N$
3. falls $x \in P$ und $w_t'x > 0$ dann **goto** 2
   falls $x \in N$ und $w_t'x \leq 0$ dann **goto** 2
   } E/A richtig!
4. falls $x \in P$ und $w_t'x \leq 0$ dann
   $w_{t+1} = w_t + x$; t++; **goto** 2
   } sei $w'x \leq 0$, sollte > 0 sein! $(w+x)'x = w'x + x'x > w'x$
5. falls $x \in N$ und $w_t'x > 0$ dann
   $w_{t+1} = w_t - x$; t++; **goto** 2
   } sei $w'x > 0$, sollte ≤ 0 sein! $(w-x)'x = w'x - x'x < w'x$
6. Abbruch? Wenn E/A für alle Beispiele richtig!

Anmerkung: Verfahren konvergiert, ist endlich, worst case: exponentiell

Kapitel 5: Künstliche neuronale Netze

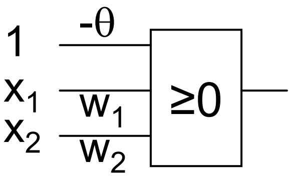
Beispiel



$$P = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\} \quad \bullet$$

$$N = \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \quad \circ$$

Schwellwert als Gewicht: $w = (-\theta, w_1, w_2)'$



$$\Downarrow$$

$$P = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

$$N = \left\{ \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

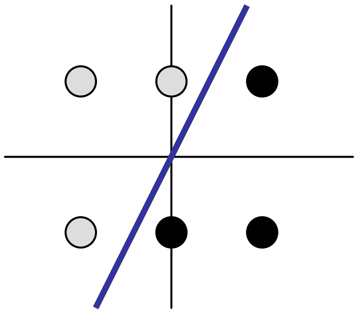
Sei initialer Gewichtvektor

$w^{(0)} = (0, 0, 0)'$

Trainingsmuster	Gewichte	Soll	Ist	Korrektur
1 1 1	0 0 0	> 0	0	add
1 1 1	1 1 1	> 0	3	☑
1 1 -1	1 1 1	> 0	1	☑
1 0 -1	1 1 1	> 0	0	add
1 1 1	2 1 0	> 0	3	☑
1 1 -1	2 1 0	> 0	3	☑
1 0 -1	2 1 0	> 0	2	☑
1 -1 -1	2 1 0	< 0	1	sub
1 -1 -1	1 2 1	< 0	-2	☑
1 -1 1	1 2 1	< 0	0	sub
1 -1 -1	0 3 0	< 0	-3	☑
1 -1 1	0 3 0	< 0	-3	☑
1 0 1	0 3 0	< 0	0	sub

Trainingsmuster	Gewichte	Soll	Ist	Korrektur
1 -1 -1	-1 3 -1	< 0	-3	☑
1 -1 1	-1 3 -1	< 0	-5	☑
1 0 1	-1 3 -1	< 0	-2	☑
1 1 1	-1 3 -1	> 0	1	☑
1 1 -1	-1 3 -1	> 0	3	☑
1 0 -1	-1 3 -1	> 0	0	add
1 -1 -1	0 3 -2	< 0	-1	☑
1 -1 1	0 3 -2	< 0	-5	☑
1 0 1	0 3 -2	< 0	-2	☑
1 1 1	0 3 -2	> 0	1	☑
1 1 -1	0 3 -2	> 0	5	☑
1 0 -1	0 3 -2	> 0	2	☑

$\Rightarrow 3\,x_1 - 2\,x_2 = 0 \qquad \Rightarrow x_2 = 3/2\,x_1$



Satz:

Sind die Beispieldaten linear separierbar, dann wird die Lösung / Trennebene (bzw. die Gewichte) mit dem Perzeptron-Lernalgorithmus gefunden.

Beweis: Sei

P die Menge der positiven Beispiele (mit $w^*x > 0$) und
N die Menge der negative Beispiele (mit $w^*x < 0$), wobei
w^* ein optimaler Gewichtsvektor

Kunstgriff: negative Beispiele aus N zu positiven Beispielen machen
$\Rightarrow B = P \cup \{ -x : x \in N \} \quad \Rightarrow \quad \forall x \in B: w^*x > 0$

Vereinfacht Lernalgorithmus:

1. Wähle $x \in B$
2. falls $w'x \leq 0$ dann $w = w + x$
3. goto 1 bis $\forall x \in B: w'x > 0$

Annahme: $\exists\, w^*: \forall\, x \in B: x'w^* > 0$ und $||w^*|| = 1$ (Trennebene existiert)

Wir betrachten $\dfrac{w_t' w^*}{||w_t||} = \cos \alpha_t$, weil α_t der Winkel zwischen w_t und w^* ist.

$\Rightarrow \cos \alpha_t \rightarrow 1 \quad \Rightarrow \quad \alpha_t \rightarrow 0 \quad \Rightarrow \quad w_t \,/\, ||w_t|| \rightarrow w^*$ für $t \rightarrow \infty$

Fakt:
Gewichte w werden dann und nur dann verändert, wenn $w'x \le 0$ gilt (Schritt 2).

Zähler: $w_{t+1}' w^* = (w_t + x)' w^* = w_t' w^* + x' w^* \geq w_t' w^* + \delta$

wobei $\delta = \min\{x' w^* : x \in B\}$

nach t Veränderungsschritten: $w_t' w^* \geq w_0' w^* + t \cdot \delta$ (♣)

Nenner: $\|w_{t+1}\|^2 = w'_{t+1}w_{t+1} = (w_t+x)'(w_t+x) = \|w_t\|^2 + \underbrace{2x'w_t}_{\leq 0} + \|x\|^2$

$$\leq \|w_t\|^2 + \|x\|^2 \leq \|w_t\|^2 + \gamma$$

wobei $\gamma = \max\{\|x\|^2 : x \in B\}$

nach t Veränderungsschritten: $\|w_t\|^2 \leq \|w_0\|^2 + t \cdot \gamma$

bzw.: $\|w_t\| \leq \sqrt{\|w_0\|^2 + t \cdot \gamma}$ (♣ ♣)

Zähler & Nenner:

$$\cos\alpha_t = \frac{w'_t w^*}{\|w_t\|} = \min\left\{\frac{w'_t w^*}{\|w_t\|}, 1\right\} \overset{(♣)}{\geq} \min\left\{\frac{w'_0 w^* + t \cdot \delta}{\|w_t\|}, 1\right\}$$

$$\overset{(♣ ♣)}{\geq} \min\left\{\frac{w'_0 w^* + t \cdot \delta}{\sqrt{\|w_0\|^2 + t \cdot \gamma}}, 1\right\} \to 1 \quad \text{für} \quad t \to \infty$$

q.e.d.