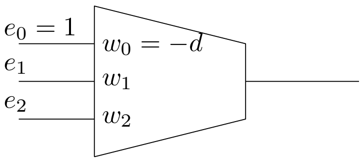
Neuronale Netze, Backpropagation

Das Lernverfahren für mehrschichtige Netze heißt Backpropagation (Rückwärtsausbreitung), da die Gewichte der Neuronen mit der Ausgabeschicht beginnend entgegen der Verarbeitungsrichtung geändert werden.

Um die Lernregeln übersichtlich formulieren zu können, wird das Neuronenmodell vereinfacht.
Statt eines Neurons mit den Eingängen e_1, e_2, den Gewichten w_1, w_2, dem Schwellenwert d und der Ausgabefunktion:

$$\begin{cases} 0 & \text{für} \quad e_1 w_1 + e_2 w_2 \ < \ d \\ 1 & \text{sonst} \end{cases}$$

verwenden wir ein Neuron mit drei Eingängen:



und der Ausgabefunktion:

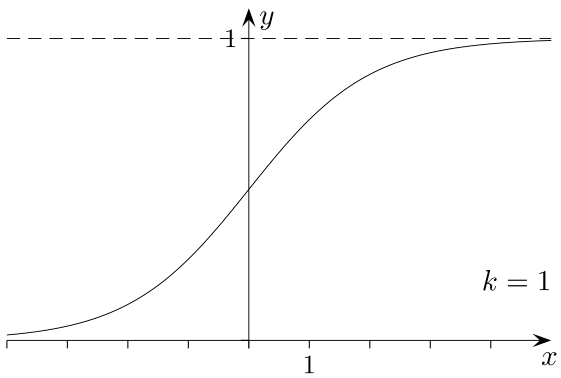
$$\begin{cases} 0 & \text{für} \quad e_1 w_1 + e_2 w_2 - d \ < \ 0 \\ 1 & \text{sonst} \end{cases}$$

Eine Unterscheidung zwischen Schwellenwert und Gewichten ist nun nicht mehr erforderlich.
Ein Eingang ist konstant 1.

Eine neuronale 0-1-Ausgabe ist für viele Anwendungen zu undifferenziert.
Daher wird eine stetige Ausgabefunktion wie z. B.

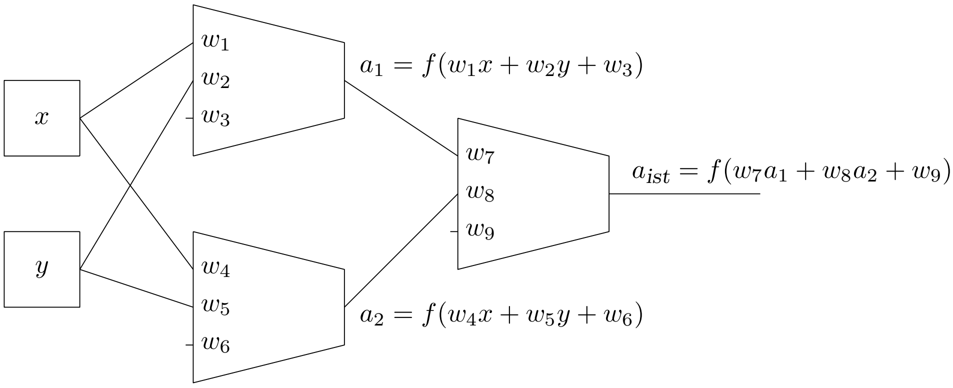
$$f(x) = \frac{1}{1 + e^{-k\,x}}$$ auf die gewichtete neuronale Eingabe $e_1 w_1 + e_2 w_2 + w_3$ angewendet.

Als besonders nützlich erweist sich, dass die Berechnung der Ableitung mit den Funktionswerten möglich ist, denn es gilt: $f'(x) = k\,f(x)\,(1 - f(x))$. Die Ableitung im Ursprung beträgt $\frac{k}{4}$.



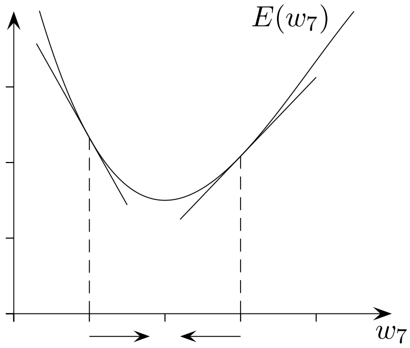
Der Lernalgorithmus

Betrachten wir ein Netz, das z. B. zur Berechnung der Exklusiv-oder-Funktion geeignet ist.



In der Lernphase werden die Gewichte so korrigiert, dass für jedes Muster (x, y) die Abweichung von a_{soll} zu a_{ist} möglichst gering wird. Als Maß für den Fehler (error) eignet sich $E = \frac{1}{2}\,(a_{soll} - a_{ist})^2$.
Der Faktor $\frac{1}{2}$ dient lediglich der Vereinfachung.
Um z. B. das Gewicht w_7 der Ausgabeschicht anzupassen, wird E als Funktion von w_7 dargestellt.

$$\begin{aligned} E(w_7) &= \frac{1}{2}\,(a_{soll} - a_{ist})^2 \\ &= \frac{1}{2}\,(a_{soll} - f(w_7a_1 + w_8a_2 + w_9))^2 \end{aligned}$$

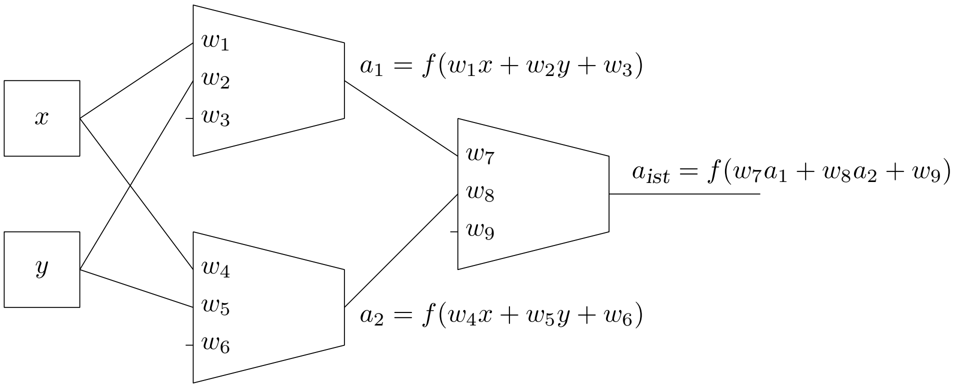


Die Korrektur von w_7 erfolgt nun in naheliegender Weise.
Um sich dem lokalen Minimum zu nähern, wird w_7 bei negativer Steigung ein wenig vergrößert, bei positiver Steigung ein wenig verkleinert. Zur Berechnung der Steigung wird die Kettenregel benötigt.

$$\begin{aligned} E(w_7) &= \frac{1}{2}\,(a_{soll} - f(\underbrace{w_7a_1 + w_8a_2 + w_9}_{\star}))^2 \\ E'(w_7) &= (a_{soll} - f(\,\star\,)) \cdot \underbrace{f'(\,\star\,)}_{f(\star)(1-f(\star))} \cdot a_1 \\ &= (a_{soll} - a_{ist}) \cdot a_{ist} \cdot (1 - a_{ist}) \cdot a_1 \end{aligned}$$

Die Korrektur eines Gewichts der inneren Schicht, z. B. w_2, ist mit etwas mehr Aufwand verbunden.

Der Lernalgorithmus,   Fortsetzung



E wird nun als Funktion von w_2 dargestellt.

$$
\begin{aligned}
E(w_2) &= \frac{1}{2}\,(a_{soll} - a_{ist})^2 \\
&= \frac{1}{2}\,(a_{soll} - f(w_7a_1 + w_8a_2 + w_9))^2 \\
&= \frac{1}{2}\,(a_{soll} - f(\underbrace{w_7 \cdot f(\underbrace{w_1x + w_2y + w_3}_{\star}) + w_8a_2 + w_9}_{\circ}))^2
\end{aligned}
$$

$$
\begin{aligned}
E'(w_2) &= (a_{soll} - f(\,\circ\,)) \cdot \underbrace{f'(\,\circ\,)}_{f(\circ)(1-f(\circ))} \cdot w_7 \cdot \underbrace{f'(\star)}_{f(\star)(1-f(\star))} \cdot y \\
&= (a_{soll} - a_{ist}) \cdot a_{ist} \cdot (1 - a_{ist}) \cdot w_7 \cdot a_1 \cdot (1 - a_1) \cdot y
\end{aligned}
$$

Ein Teil des Terms wurde bereits bei der Veränderung eines Gewichts der Ausgabeschicht bestimmt und kann hier eingefügt werden. Die Lernregeln sind nun erkennbar.
Im allgemeinen Fall wird entsprechend vorgegangen. Mit der Berücksichtigung des bereits Errechneten kann der Rechenaufwand in Grenzen gehalten werden.