Backpropagation – Beispiel

R.Brause, υ8ρς8ςπρτ

Zur Veranschaulichung des Ablaufs beim Backpropagation-Algorithmus sei hier ein konkretes Beispiel gerechnet. Angenommen, wir haben eine Netzwerkarchitektur wie sie in Abbildung 1 gezeigt ist.

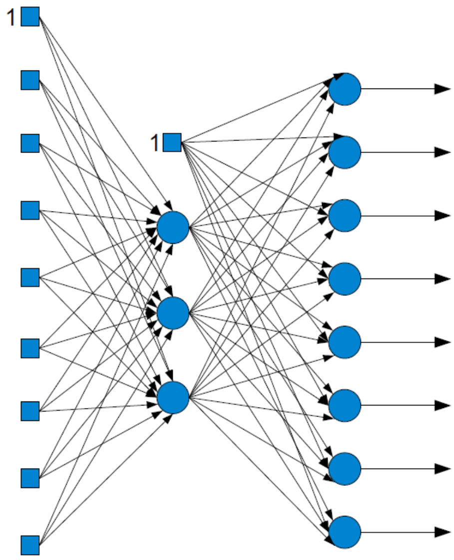

Abb. 1 Eine 8x3x8-Architektur

Errechnen der Ausgabe

Die 8-dimensionale Eingabe **x** = $(x_1,\ldots,x_8)$ wird ergänzt durch die feste Bias-Eingabe 1, so dass **x** = $(1,x_1,\ldots,x_8)$ wird. Seien anfangs $x_1 = 0.1$, $x_2 = 0.2$,…, $x_8 = 0.8$ und die Gewichte der ersten Schicht für jedes Neuron mit $w_0 = 0$, $w_1 = 1$, …, $w_8 = 1$ vorinitialisiert. Dann ergibt sich die Ausgabe für das erste Neuron der ersten Schicht mit S(z) als Fermi-Funktion zu

$$z_1^{(1)} = \sum_{j=0}^{8} w_{1j}^{(1)} x_j^{(1)} = w_0 + w_1x_1 + \ldots + w_8x_8 = 0 + 0.1 + \ldots +0.8 = 3.6,\quad S(3.6) = 0.97 \tag{1}$$

für jedes der 3 Neuronen, so dass $\mathbf{y}^{(1)} = (1, 0.97, 0.97, 0.97)$ die um die Bias-Aktivität 1 ergänzte Eingabe für die nächste Schicht darstellt. Auch hier seien die Gewichte vorinitialisiert, und zwar alle gleichmäßig mit $w_{ij}^{(2)} = 1$. Somit ist die Aktivität für das erste Neuron der zweiten Schicht

$$y_1^{(2)} = \sum_{j=0}^{3} w_{1j}^{(2)} x_j^{(2)} = w_0^{(2)} + w_1^{(2)}x_1^{(2)} + w_2^{(2)}x_2^{(2)} + w_3^{(2)}x_3^{(2)} = 1 + 0.97 + 0.97 + 0.97 = 3.91 \tag{2}$$

Gleiches gilt für alle anderen Neuronen, so dass für $\mathbf{y}^{(2)} = (3.91, 3.91, \ldots, 3.91)$ ein 8-dim Ausgabevektor errechnet ist.

Fehlerrückführung

Angenommen, die 8-dim. Lehrervorgabe sei **L** = (1,0,…,0). Da die Ausgabe in dieser Hinsicht zu groß ist, müssen also überall im Netz die Gewichte verkleinert werden, um der Lehrervorgabe zu genügen. Der Fehler der zweiten Schicht ist dann

$$\delta^{(2)} = -\left((y_1^{(2)}-L_1)S'(z_1^{(2)}), \ldots, (y_8^{(2)}-L_8)S'(z_8^{(2)}) \right) = -\left((y_1^{(2)}-L_1), \ldots, (y_8^{(2)}-L_8) \right) \tag{3}$$

$$= -(2.91, 3.91, \ldots, 3.91) \quad \text{da hier } S(z) = z \text{ und damit } S'(z) = 1$$

Für die erste Schicht errechnet sich der Fehler für das *i*-te Neuron (i = 1, 2, 3) zu

$$\delta_i^{(1)} = \left(\sum_{j=1}^{8} \delta_j^{(2)} w_{ji}^{(2)} \right) S'\left(z_i^{(1)}\right) \tag{4}$$

Hier wird von dem Fehler am j-ten Neuron, der an der 2.Schicht festgestellt wurde, über die 8 Gewichte, über die Neuron *i* mit den 8 Ausgabeneuronen verbunden ist, auf eine falsche Ausgabe von Neuron *i* zurückgerechnet.
Konkret errechnet sich der Fehler mit

$$\left(\sum_{j=1}^{8} \delta_j^{(2)} w_{ji}^{(2)} \right) = -\,(2.91 \cdot 1 + 3.91 \cdot 1 + \ldots + 3.91 \cdot 1) = -\,30.28 \tag{5}$$

zu

$$\delta_1^{(1)} = \left(\sum_{j=1}^{8} \delta_j^{(2)} w_{j1}^{(2)} \right) S'\left(z_1^{(1)}\right) = -\,30.28 \cdot (1-S(z_1))S(z_1) = -\,30.28 \cdot (1-S(3.6))S(3.6)$$

$$= -\,30.28 \cdot (1-0.97) \cdot 0.97 = -\,30.28 \cdot (0.03) \cdot 0.97 = -\,0.88$$

$$\delta_2^{(1)} = \left(\sum_{j=1}^{8} \delta_j^{(2)} w_{j2}^{(2)} \right) S'\left(z_2^{(1)}\right) = -\,30.28 \cdot (1-S(z_2))S(z_2) = -\,0.88 \tag{6}$$

$$\delta_3^{(1)} = \left(\sum_{j=1}^{8} \delta_j^{(2)} w_{j3}^{(2)} \right) S'\left(z_3^{(1)}\right) = -\,30.28 \cdot (1-S(z_3))S(z_3) = -\,0.88$$

Verbessern der Gewichte

Sodann wird für die Änderung der Gewichte

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \tag{7}$$

die Verbesserungen errechnet mit

$$\Delta w_{ij}(t) = \gamma \cdot \delta_i(t) \cdot x_j(t) \tag{8}$$

Für die Gewichte der zweiten Schicht mit $S(z) = z$ gilt

$$\Delta w_{ij}^{(2)}(t) = \gamma\, \delta_i^{(2)} x_j^{(2)} = -\gamma\, (y_i^{(2)} - L_i)\, x_j^{(2)} \tag{9}$$

und somit für das 1. Neuron

$$\Delta w_{10}^{(2)}(t) = -\gamma\, (y_1^{(2)} - L_1)\, x_0^{(2)} = -\gamma \cdot 2.91 \cdot 1 \qquad \textit{Bias-Gewicht} \tag{10}$$

$$\Delta w_{11}^{(2)}(t) = -\gamma\, (y_1^{(2)} - L_1)\, x_1^{(2)} = -\gamma \cdot 2.91 \cdot 0.97$$

$$\Delta w_{12}^{(2)}(t) = -\gamma\, (y_1^{(2)} - L_1)\, x_2^{(2)} = -\gamma \cdot 2.91 \cdot 0.97$$

$$\Delta w_{13}^{(2)}(t) = -\gamma\, (y_1^{(2)} - L_1)\, x_3^{(2)} = -\gamma \cdot 2.91 \cdot 0.97$$

und für das 2. Neuron

$$\Delta w_{20}^{(2)}(t) = -\gamma\, (y_2^{(2)} - L_2)\, x_0^{(2)} = -\gamma \cdot 3.91 \cdot 1 \qquad \textit{Bias-Gewicht} \tag{11}$$

$$\Delta w_{21}^{(2)}(t) = -\gamma\, (y_2^{(2)} - L_2)\, x_1^{(2)} = -\gamma \cdot 3.91 \cdot 0.97$$

$$\Delta w_{22}^{(2)}(t) = -\gamma\, (y_2^{(2)} - L_2)\, x_2^{(2)} = -\gamma \cdot 3.91 \cdot 0.97$$

$$\Delta w_{23}^{(2)}(t) = -\gamma\, (y_2^{(2)} - L_2)\, x_3^{(2)} = -\gamma \cdot 3.91 \cdot 0.97$$

bis zum 8. Neuron

$$\Delta w_{80}^{(2)}(t) = -\gamma\, (y_8^{(2)} - L_8)\, x_0^{(2)} = -\gamma \cdot 3.91 \cdot 1 \qquad \textit{Bias-Gewicht} \tag{12}$$

$$\Delta w_{81}^{(2)}(t) = -\gamma\, (y_8^{(2)} - L_8)\, x_1^{(2)} = -\gamma \cdot 3.91 \cdot 0.97$$

$$\Delta w_{82}^{(2)}(t) = -\gamma\, (y_8^{(2)} - L_8)\, x_2^{(2)} = -\gamma \cdot 3.91 \cdot 0.97$$

$\Delta w_{83}^{(2)}(t) = -\gamma\,(y_8^{(2)}-L_8)\,x_3^{(2)} = -\gamma\cdot 3.91\cdot 0.97$

Für die Gewichte der ersten Schicht ist mit Gl.(4) für das Gewicht der *k*-ten Eingabe zum *i*-ten Neuron

$$\Delta w_{ik}^{(1)}(t) = \gamma\,\delta_i^{(1)} x_k^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{ji}^{(2)}\right) S'\left(z_i^{(1)}\right) x_k^{(1)} \qquad (13)$$

und damit für das 1. Neuron

$$\Delta w_{10}^{(1)}(t) = \gamma\,\delta_1^{(1)} x_0^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j1}^{(2)}\right) S'\left(z_1^{(1)}\right) x_0^{(1)} \qquad \textit{Bias-Gewicht} \qquad (14)$$

$$= -\gamma\cdot 30.28\cdot(1-0.97)\cdot 0.97\cdot 1 = -\gamma\cdot 0.88$$

$$\Delta w_{11}^{(1)}(t) = \gamma\,\delta_1^{(1)} x_1^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j1}^{(2)}\right) S'\left(z_1^{(1)}\right) x_1^{(1)}$$

$$= -\gamma\cdot 30.28\cdot(1-0.97).0.97\cdot 0.1 = -\gamma\cdot 0.088$$

…

$$\Delta w_{18}^{(1)}(t) = \gamma\,\delta_1^{(1)} x_8^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j1}^{(2)}\right) S'\left(z_1^{(1)}\right) x_8^{(1)}$$

$$= -\gamma\cdot 30.28\cdot(1-0.97).0.97\cdot 0.8 = -\gamma\cdot 0.7$$

Für das 3. Neuron ist dies dann auch

$$\Delta w_{30}^{(1)}(t) = \gamma\,\delta_3^{(1)} x_0^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j3}^{(2)}\right) S'\left(z_3^{(1)}\right) x_0^{(1)} \qquad \textit{Bias-Gewicht} \qquad (15)$$

$$= -\gamma\cdot 30.28\cdot(1-0.97).0.97\cdot 1 = -\gamma\cdot 0.88$$

$$\Delta w_{31}^{(1)}(t) = \gamma\,\delta_3^{(1)} x_1^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j3}^{(2)}\right) S'\left(z_3^{(1)}\right) x_1^{(1)}$$

$$= -\gamma\cdot 30.28\cdot(1-0.97).0.97\cdot 0.1 = -\gamma\cdot 0.088$$

…

$$\Delta w_{38}^{(1)}(t) = \gamma\,\delta_3^{(1)} x_8^{(1)} = \gamma\left(\sum_{j=1}^{8}\delta_j^{(2)} w_{j3}^{(2)}\right) S'\left(z_3^{(1)}\right) x_8^{(1)}$$

$$= -\gamma\cdot 30.28\cdot(1-0.97).0.97\cdot 0.8 = -\gamma\cdot 0.7$$

Für die eigentliche Korrektur der Gewichte müssen wir unterscheiden: Beim *offline*-Learning werden die Verbesserungen $\Delta w_{ij}(t)$ über alle Muster **x**(t) (alle *T* Zeitpunkte *t*) gesammelt (addiert)

$$\Delta w_{ij} = \gamma\,\frac{1}{T}\sum_{t=1}^{T}\Delta w_{ij}(t) \qquad (16)$$

und dann erst tatsächlich umgesetzt; beim *online*-Learning dagegen sofort nach jedem Muster.