Home | Startseite | Impressum | Kontakt | Gästebuch

Aufgabe: Bilden Sie vermöge eines dreistufigen neuronalen Netzes mit verdeckter Schicht sämtliche neuronalen Bitmuster, die sich mittels dreier Eingangsneuronen bilden lassen, auf ihre komplementären Bitmuster ab.

Lösung: Die n_1 Inputneuronen $y_{i_1}^{(1)}$ der Eingabeschicht (1) mit den Gewichten $w_{i_1 i_2}^{(1)}$ können übersichtlich in folgender Matrixschreibweise zusammengefaßt werden:

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} & w_{33}^{(1)} \end{pmatrix} \begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \\ y_3^{(1)} \end{pmatrix}.$$

In Komponentenschreibweise erhalten wir mit der linearen Standard-Propagierungsfunktion $x_{i_2}^{(1)}$ das folgende lineare Gleichungssystem:

$$x_{i_2}^{(1)} = \sum_{i_1=1}^{n_1} w_{i_1 i_2}^{(1)} y_{i_1}^{(1)}.$$

Mit der Aktivierungsfunktion f ergeben sich die n_2 verborgenen Neuronen $y_{i_2}^{(2)}$ der verdeckten Schicht (2) mit den Gewichten $w_{i_2 i_3}^{(2)}$ zu

$$y_{i_2}^{(2)} = f\left(x_{i_2}^{(1)}\right) = f\left(\sum_{i_1=1}^{n_1} w_{i_1 i_2}^{(1)} y_{i_1}^{(1)}\right).$$

Die lineare Standard-Propagierungsfunktion $x_{i_3}^{(2)}$ kann wieder durch eine orthogonale Matrix[1] mit den Gewichten $w_{i_2 i_3}^{(2)}$ ausgedrückt werden:

$$\begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} & w_{31}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & w_{32}^{(2)} \\ w_{13}^{(2)} & w_{23}^{(2)} & w_{33}^{(2)} \end{pmatrix} \begin{pmatrix} y_1^{(2)} \\ y_2^{(2)} \\ y_3^{(2)} \end{pmatrix},$$

die wir in Komponentenschreibweise wieder wie folgt schreiben können:

$$x_{i_3}^{(2)} = \sum_{i_2=1}^{n_2} w_{i_2 i_3}^{(2)} y_{i_2}^{(2)}.$$

Die Optimierung startet an der Ausgangsebene der n_3 Outputneuronen der Ausgabeschicht (3) mit dem Output $y_{i_3}^{(3)}$ des Ausgangsneurons i_3:

[1] Eine Matrix, deren Inverse gleich ihrer Transponierten ist

$$y_{i_3}^{(3)} = f\left(x_{i_3}^{(2)}\right) = f\left(\sum_{i_2=1}^{n_2} w_{i_2 i_3}^{(2)} y_{i_2}^{(2)}\right),$$

mit $1 \le i_3 \le n_3$. Das Optimierungsproblem kann damit wie folgt beschrieben werden:

$$E = \frac{1}{2}\sum_{i_3=1}^{n_3}\left(\eta_{i_3} - y_{i_3}^{(3)}\right)^2 = \frac{1}{2}\sum_{i_3=1}^{n_3}\left(\eta_{i_3} - f\left(x_{i_3}^{(2)}\right)\right)^2 \to \min .$$

Der Gradient der Fehlerfunktion muß folglich null sein:

$$\nabla E = \left(\frac{\partial E}{\partial w_{11}^{(1)}}, \frac{\partial E}{\partial w_{12}^{(1)}}, \ldots, \frac{\partial E}{\partial w_{n_1 n_2}^{(1)}}, \frac{\partial E}{\partial w_{11}^{(2)}}, \frac{\partial E}{\partial w_{12}^{(2)}}, \ldots, \frac{\partial E}{\partial w_{n_2 n_3}^{(2)}}\right) = 0.$$

Zur Gewichtsadaption sind die partiellen Ableitungen der Fehlerfunktion

$$\Delta w_{ij}^{(1)} = -\lambda \frac{\partial E}{\partial w_{ij}^{(1)}} \quad \text{und} \quad \Delta w_{ij}^{(2)} = -\lambda \frac{\partial E}{\partial w_{ij}^{(2)}}$$

zu bilden.

Zunächst gilt an der Schnittstelle der Hidden-Output-Schicht für $w_{i_2 i_3}^{(2)}$ nach der Kettenregel

$$\frac{\partial E}{\partial w_{ij}^{(2)}} = \frac{\partial E}{\partial y_{i_3}^{(3)}} \frac{\partial f\left(x_{i_3}^{(2)}\right)}{\partial x_{i_3}^{(2)}} \frac{\partial x_{i_3}^{(2)}}{\partial w_{ij}^{(2)}}.$$

Da

$$\frac{\partial w_{i_2 i_3}^{(2)}}{\partial w_{ij}^{(2)}} = \begin{cases} 1 & \text{für} \quad (i,\, j) = (i_2,\, i_3), \\ 0 & \text{für} \quad (i,\, j) \ne (i_2,\, i_3), \end{cases}$$

ist die Summe

$$\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} \frac{\partial w_{i_2 i_3}^{(2)}}{\partial w_{ij}^{(2)}} = 1$$

und damit die Summe

$$\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} \frac{\partial x_{i_3}^{(2)}}{\partial w_{i_2 i_3}^{(2)}} \frac{\partial w_{i_2 i_3}^{(2)}}{\partial w_{ij}^{(2)}} = \frac{\partial x_{i_3}^{(2)}}{\partial w_{ij}^{(2)}}.$$

Folglich ist

$$\frac{\partial E}{\partial w_{ij}^{(2)}} = \sum_{i_3=1}^{n_3} \frac{\partial E}{\partial y_{i_3}^{(3)}} \frac{\partial f\left(x_{i_3}^{(2)}\right)}{\partial x_{i_3}^{(2)}} \sum_{i_2=1}^{n_2} \frac{\partial x_{i_3}^{(2)}}{\partial w_{i_2 i_3}^{(2)}} \frac{\partial w_{i_2 i_3}^{(2)}}{\partial w_{ij}^{(2)}}.$$

Die Ableitung von E nach $y_{i_3}^{(3)}$ liefert

$$\frac{\partial E}{\partial y_{i_3}^{(3)}} = -\left(\eta_{i_3} - y_{i_3}^{(3)}\right).$$

Für die Sigmoidfunktion

$$f(x) = \frac{1}{1+e^{-x}}$$

ist

$$f'(x) = \frac{1}{\left(1+e^{-x}\right)^2} e^{-x} = f(x) \frac{1+e^{-x}-1}{1+e^{-x}} = f(x)\left(1 - f(x)\right)$$

und damit

$$\frac{\partial y_{i_3}^{(3)}}{\partial x_{i_3}^{(2)}} = \frac{\partial f\left(x_{i_3}^{(2)}\right)}{\partial x_{i_3}^{(2)}} = f\left(x_{i_3}^{(2)}\right)\left(1 - f\left(x_{i_3}^{(2)}\right)\right).$$

Wegen

$$x_{i_3}^{(2)} = \sum_{i_2=1}^{n_2} w_{i_2 i_3}^{(2)} y_{i_2}^{(2)} \quad \text{ist} \quad \frac{\partial x_{i_3}^{(2)}}{\partial w_{i_2 i_3}^{(2)}} = y_{i_2}^{(2)}.$$

Setzen wir alle drei Beiträge ein, erhalten wir, da $\partial w_{i_2 i_3}^{(2)} / \partial w_{ij}^{(2)}$ für $(i,\ j) \neq (i_3, i_2)$ verschwindet, den Ausdruck

$$\begin{aligned} \frac{\partial E}{\partial w_{ij}^{(2)}} &= -\sum_{i_3=1}^{n_3} \left(\eta_{i_3} - f\left(x_{i_3}^{(2)}\right)\right) f\left(x_{i_3}^{(2)}\right)\left(1 - f\left(x_{i_3}^{(2)}\right)\right) \sum_{i_2=1}^{n_2} y_{i_2}^{(2)} \frac{\partial w_{i_3 i_2}^{(2)}}{\partial w_{ij}^{(2)}} \\ &= -\left(\eta_j - y_j^{(3)}\right) y_j^{(3)} \left(1 - y_j^{(3)}\right) y_i^{(2)}. \end{aligned}$$

Damit lautet die erste Gewichtung

$$\Delta w_{i_2 i_3}^{(2)} = -\lambda \frac{\partial E}{\partial w_{i_2 i_3}^{(2)}} = \lambda \left(\eta_{i_3} - y_{i_3}^{(3)}\right) y_{i_3}^{(3)} \left(1 - y_{i_3}^{(3)}\right) y_{i_2}^{(2)}.$$

An der Schnittstelle Input-Hidden-Schicht für $w_{i_1 i_2}^{(1)}$ gilt nach der Kettenregel analog

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = \sum_{i_3=1}^{n_3} \frac{\partial E}{\partial y_{i_3}^{(3)}} \frac{\partial f\left(x_{i_3}^{(2)}\right)}{\partial x_{i_3}^{(2)}} \sum_{i_2=1}^{n_2} \frac{\partial x_{i_3}^{(2)}}{\partial y_{i_2}^{(2)}} \frac{\partial f\left(x_{i_2}^{(1)}\right)}{\partial x_{i_2}^{(1)}} \sum_{i_1=1}^{n_1} \frac{\partial x_{i_2}^{(1)}}{\partial w_{i_1 i_2}^{(1)}} \frac{\partial w_{i_1 i_2}^{(1)}}{\partial w_{ij}^{(1)}}.$$

Ebenfalls in Analogie zu oben kommt hinzu

$$\frac{\partial y_{i_2}^{(2)}}{\partial x_{i_2}^{(1)}} = \frac{\partial f\left(x_{i_2}^{(1)}\right)}{\partial x_{i_2}^{(1)}} = f\left(x_{i_2}^{(1)}\right)\left(1 - f\left(x_{i_2}^{(1)}\right)\right).$$

Wegen

$$x_{i_3}^{(2)} = \sum_{i_2=1}^{n_2} w_{i_2 i_3}^{(2)} y_{i_2}^{(2)} \quad \text{ist} \quad \frac{\partial x_{i_3}^{(2)}}{\partial y_{i_2}^{(2)}} = w_{i_2 i_3}^{(2)},$$

und aus

$$x_{i_2}^{(1)} = \sum_{i_1=1}^{n_1} w_{i_1 i_2}^{(1)} y_{i_1}^{(1)} \quad \text{folgt} \quad \frac{\partial x_{i_2}^{(1)}}{\partial w_{i_1 i_2}^{(1)}} = y_{i_1}^{(1)}.$$

Setzen wir alle Größen ein, erhalten wir

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = -\sum_{i_3=1}^{n_3}\left(\eta_{i_3} - y_{i_3}^{(3)}\right) f\left(x_{i_3}^{(2)}\right)\left(1 - f\left(x_{i_3}^{(2)}\right)\right)\sum_{i_2=1}^{n_2} w_{i_2 i_3}^{(2)} f\left(x_{i_2}^{(1)}\right)\left(1 - f\left(x_{i_2}^{(1)}\right)\right)\sum_{i_1=1}^{n_1} y_{i_1}^{(1)} \frac{\partial w_{i_1 i_2}^{(1)}}{\partial w_{ij}^{(1)}}.$$

Mit

$$\frac{\partial w_{i_1 i_2}^{(2)}}{\partial w_{ij}^{(2)}} = \begin{cases} 1 & \text{für} \quad (i, j) = (i_1, i_2), \\ 0 & \text{für} \quad (i, j) \neq (i_1, i_2) \end{cases}$$

folgt weiter

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = -\sum_{i_3=1}^{n_3}\left(\eta_{i_3} - y_{i_3}^{(3)}\right) y_{i_3}^{(3)}\left(1 - y_{i_3}^{(3)}\right) w_{j i_3}^{(2)} y_j^{(1)}\left(1 - y_j^{(1)}\right) y_i^{(1)}.$$

Mit der Wahl $i = i_1$ und $j = i_2$ ist

$$\frac{\partial E}{\partial w_{i_1 i_2}^{(1)}} = -\sum_{i_3=1}^{n_3}\left(\eta_{i_3} - y_{i_3}^{(3)}\right) y_{i_3}^{(3)}\left(1 - y_{i_3}^{(3)}\right) w_{i_2 i_3}^{(2)} y_{i_2}^{(2)}\left(1 - y_{i_2}^{(2)}\right) y_{i_1}^{(1)}.$$

Damit erhalten wir die zweite Gewichtung zu

$$\Delta w_{i_1 i_2}^{(1)} = -\lambda \frac{\partial E}{\partial w_{i_1 i_2}^{(1)}} = \lambda \sum_{i_3=1}^{n_3}\left(\eta_{i_3} - y_{i_3}^{(3)}\right) y_{i_3}^{(3)}\left(1 - y_{i_3}^{(3)}\right) w_{i_2 i_3}^{(2)} y_{i_2}^{(2)}\left(1 - y_{i_2}^{(2)}\right) y_{i_1}^{(1)}.$$

Für die schrittweise Gewichtsadaption mit zufällig vorinitialisierten Verbindungsgewichten $w_{i_2 i_3}^{(2)}(0)$ und $w_{i_1 i_2}^{(1)}(0)$ ergibt sich insgesamt

$$\Delta w_{i_2 i_3}^{(2)}(0) = -\lambda \left.\frac{\partial E}{\partial w_{i_2 i_3}^{(2)}}\right|_{w_{i_2 i_3}^{(2)} = w_{i_2 i_3}^{(2)}(0)} = \lambda\left(\eta_{i_3} - y_{i_3}^{(3)}(0)\right) y_{i_3}^{(3)}(0)\left(1 - y_{i_3}^{(3)}(0)\right) y_{i_2}^{(2)}$$

und

$$\Delta w_{i_1 i_2}^{(1)}(0) = -\lambda \left.\frac{\partial E}{\partial w_{i_1 i_2}^{(1)}}\right|_{w_{i_1 i_2}^{(1)} = w_{i_1 i_2}^{(1)}(0)}$$
$$= \lambda \sum_{i_3=1}^{n_3} \left(\eta_{i_3} - y_{i_3}^{(3)}(0)\right) y_{i_3}^{(3)}(0)\left(1 - y_{i_3}^{(3)}(0)\right) w_{i_2 i_3}^{(2)}(0) y_{i_2}^{(2)}(0)\left(1 - y_{i_2}^{(2)}(0)\right) y_{i_1}^{(1)}.$$

Mit den Abkürzungen δ für die Fehler können wir die Ausdrücke noch einmal vereinfachen:

$$\delta_{i_3}^{(3)}(0) = \left(\eta_{i_3} - y_{i_3}^{(3)}(0)\right) y_{i_3}^{(3)}(0)\left(1 - y_{i_3}^{(3)}(0)\right),$$
$$\delta_{i_2}^{(2)}(0) = \sum_{i_3=1}^{n_3} \delta_{i_3}^{(3)}(0) w_{i_2 i_3}^{(2)}(0) y_{i_2}^{(2)}(0)\left(1 - y_{i_2}^{(2)}(0)\right),$$

so daß

$$\Delta w_{i_2 i_3}^{(2)}(0) = \lambda \delta_{i_3}^{(3)}(0) y_{i_2}^{(2)} \quad \text{und} \quad \Delta w_{i_1 i_2}^{(1)}(0) = \lambda \delta_{i_2}^{(2)}(0) y_{i_1}^{(1)}.$$

Nach Adaption der Gewichte im ersten Schritt können die im nächsten berechnet werden:

$$w_{i_2 i_3}^{(2)}(1) = w_{i_2 i_3}^{(2)}(0) + \Delta w_{i_2 i_3}^{(2)}(0) \quad \text{bzw.} \quad w_{i_1 i_2}^{(1)}(1) = w_{i_1 i_2}^{(1)}(0) + \Delta w_{i_1 i_2}^{(1)}(0).$$

Rekursiv gilt

$$w_{i_2 i_3}^{(2)}(k+1) = w_{i_2 i_3}^{(2)}(k) + \Delta w_{i_2 i_3}^{(2)}(k) \quad \text{bzw.} \quad w_{i_1 i_2}^{(1)}(k+1) = w_{i_1 i_2}^{(1)}(k) + \Delta w_{i_1 i_2}^{(1)}(k),$$

mit

$$\Delta w_{i_2 i_3}^{(2)}(k) = \lambda \delta_{i_3}^{(3)}(k) y_{i_2}^{(2)} \quad \text{bzw.} \quad \Delta w_{i_1 i_2}^{(1)}(k) = \lambda \delta_{i_2}^{(2)}(k) y_{i_1}^{(1)}.$$

Dieses rückwärtsgewandte Durchreichen der Soll-Ist-Abweichung zwischen Zielwert und Netzoutput zurück zur Inputschicht beschreibt das, was man gemeinhin unter einem Backpropagation-Algorithmus versteht.

$y_1^{(1)}$	$y_2^{(1)}$	$y_3^{(1)}$	$y_1^{(3)}$	$y_2^{(3)}$	$y_3^{(3)}$	η_1	η_2	η_3
1	1	1	0,0290	0,0290	0,0290	0	0	0
1	1	0	0,0315	0,0315	0,9722	0	0	1
1	0	1	0,0315	0,9722	0,0315	0	1	0
0	1	1	0,9722	0,0315	0,0315	1	0	0
1	0	0	0,0386	0,9651	0,9651	0	1	1
0	1	0	0,9651	0,0386	0,9651	1	0	1
0	0	1	0,9651	0,9651	0,0386	1	1	0
0	0	0	0,9462	0,9462	0,9462	1	1	1

Tabelle 1. Werte der Eingangs- und Ausgangsneuronen sowie die Soll-Ausgangswerte

In Tabelle 1 wurden als Soll-Ausgangswerte die Komplemente der Inputwerte der Eingabeschicht gewählt. Die Fehler zwischen Soll- und Ist-Ausgangswerten liegen im Bereich von 3-5 Prozent, werden aber je nach Lage der Schwelle ohnehin entweder auf- oder abgerundet. Der Anhang enthält ein MATLAB-Programm, das den Backpropagation-Algorithmus mit Angabe des Fehlers ausführt. Die Fehler liegen in unserem Fall in der Größenordnung von 10^{-4}.

Anhang

```
% Programm Hidden Layer

% Fehlerberechnung für ein dreistufiges neuronales Netz mit verdeckter
% Schicht mit beliebig vielen Ein- und Ausgabeneuronen und einer Sigmoidkurve
% als Aktivierungsfunktion
clear all

% Der Lernparameter sei lambda
lambda = 5;

kmax = 50;

n1 = 3;
n2 = 3;
n3 = 3;

for i1=1:n1
    for i2=1:n2
        w1(i1,i2,1) = 1;
    end
end

for i2=1:n2
    for i3=1:n3
        w2(i2,i3,1) = 1;
    end
end

% Input des Neurons i1 in Schicht 1, wobei 1 <= i1 <= n1
y1(1,1) = 0;
y1(2,1) = 0;
y1(3,1) = 0;

% eta(i3) Soll-Output des Neurons i3 in Schicht 3, wobei 1 <= i3 <= n3
eta(1) = 1;
eta(2) = 1;
eta(3) = 1;

% x1(i2,1)
x1(1,1) = w1(1,1,1)*y1(1,1) + w1(2,1,1)*y1(2,1) + w1(3,1,1)*y1(3,1);
x1(2,1) = w1(1,2,1)*y1(1,1) + w1(2,2,1)*y1(2,1) + w1(3,2,1)*y1(3,1);
x1(3,1) = w1(1,3,1)*y1(1,1) + w1(2,3,1)*y1(2,1) + w1(3,3,1)*y1(3,1);
% y2(i2,1)
y2(1,1) = 1/(1+exp(-x1(1,1)));
y2(2,1) = 1/(1+exp(-x1(2,1)));
y2(3,1) = 1/(1+exp(-x1(2,1)));
% x2(i3,1)
```



```
x2(1,1) = w2(1,1,1)*y2(1,1) + w2(2,1,1)*y2(2,1) + w2(3,1,1)*y2(3,1);
x2(2,1) = w2(1,2,1)*y2(1,1) + w2(2,2,1)*y2(2,1) + w2(3,2,1)*y2(3,1);
x2(3,1) = w2(1,3,1)*y2(1,1) + w2(2,3,1)*y2(2,1) + w2(3,3,1)*y2(3,1);
% y3(i3,1)
y3(1,1) = 1/(1+exp(-x2(1,1)));
y3(2,1) = 1/(1+exp(-x2(2,1)));
y3(3,1) = 1/(1+exp(-x2(3,1)));

summeE = 0;
for k=1:kmax
    summe3 = 0;
    for i3=1:n3
        for i2=1:n2
            x2(i3,k)    =    w2(1,i3,k)*y2(1,1)    +    w2(2,i3,k)*y2(2,1)    +
w2(3,i3,k)*y2(3,1);
            y3(i3,k) = 1/(1+exp(-x2(i3,k)));
            delta3(i3,k) = (eta(i3) - y3(i3,k))*y3(i3,k)*(1-y3(i3,k));
            x1(i2,k)    =    w1(1,i2,k)*y1(1,1)    +    w1(2,i2,k)*y1(2,1)    +
w1(3,i2,k)*y1(3,1);
            y2(i2,k) = 1/(1+exp(-x1(i2,k)));
            Deltaw2(i2,i3,k) = lambda*delta3(i3,k)*y2(i2,1);
            summe3    =    summe3    +    delta3(i3,k)*w2(i2,i3,k)*y2(i2,k)*(1-
y2(i2,k));
            delta2(i2,k) = summe3;
            w2(i2,i3,k+1) = w2(i2,i3,k) + Deltaw2(i2,i3,k);
            for i1=1:n1
                Deltaw1(i1,i2,k) = lambda*delta2(i2,k)*y1(i1,1);
                w1(i1,i2,k+1) = w1(i1,i2,k) + Deltaw1(i1,i2,k);
            end
        end
        E(k) = summeE + 1/2*(eta(i3) - y3(i3,k))^2;
    end
end

Error = E(kmax)
Input1 = y1(1,1)
Input2 = y1(2,1)
Input3 = y1(3,1)
Output1 = y3(1,kmax)
Output2 = y3(2,kmax)
Output3 = y3(3,kmax)
Solloutput1 = eta(1)
Solloutput2 = eta(2)
Solloutput3 = eta(3)
```

