

## Mathematikaufgabe 73

---

[Home](#) | [Startseite](#) | [Impressum](#) | [Kontakt](#) | [Gästebuch](#)

**Aufgabe:** Berechnen Sie die allgemeine Fehlerfunktion für ein zweistufiges neuronales Netz ohne verdeckte Schicht mit drei Eingabe- und drei Ausgabeneuronen, die anhand von drei Mustern trainiert werden sollen.

**Lösung:** Gegeben seien die Vektoren für überwachtes Lernen:

$$\begin{aligned}\vec{x} &= (x_1, x_2, x_3) && \text{Eingangsvektor,} \\ \vec{y} &= (y_1, y_2, y_3) && \text{Ist - Ausgangsvektor,} \\ \vec{\eta} &= (\eta_1, \eta_2, \eta_3) && \text{Soll - Ausgangsvektor.}\end{aligned}$$

Dabei wird dem Ausgangsvektor ein Zielwert gegenübergestellt. Die Gewichte zwischen den sendenden Neuronen  $j$  und den empfangenden Neuronen  $i$  sind durch folgende Matrixabbildung gegeben:

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

Da während der Lernphase Soll- und Istausgabe in der Regel nicht übereinstimmen, wird zunächst eine Fehlerfunktion  $E_p$  über alle drei Ausgabeneuronen  $i = 1, 2, 3$  gebildet:

$$E_p = \frac{1}{2} \sum_{i=1}^3 (\eta_{pi} - y_{pi})^2,$$

wobei  $\eta_{pi}$  der Sollwert und  $y_{pi}$  der Istwert der Ausgabeneinheit  $i$  für das Trainingsmuster  $p$  ist. Das Fehlersignal entspricht somit der Fehlerquadratsumme zwischen tatsächlicher und erwarteter Ausgabe für dieses Muster. Der Gesamtfehler  $E$  des Netzes ergibt sich damit als die Summe aller  $E_p$  über die Trainingsmuster  $p = 1, 2, 3$  gemäß der Summation

$$E = \sum_{p=1}^3 E_p = \frac{1}{2} \sum_{p=1}^3 \sum_{i=1}^3 (\eta_{pi} - y_{pi})^2.$$

Durch schrittweises Verändern der Gewichte kann der Gesamtfehler des neuronalen Netzes sukzessive verringert werden. Das Verfahren ist also analog zur linearen Regression mittels Least-squares-Fit. Die Ausgangsgrößen entsprechen dabei einer Koordinatentransformation bzw. einem linearen Gleichungssystem mit drei Unbekannten:

$$\begin{aligned}y_{p1} &= w_{11}x_{p1} + w_{12}x_{p2} + w_{13}x_{p3}, \\ y_{p2} &= w_{21}x_{p1} + w_{22}x_{p2} + w_{23}x_{p3}, \\ y_{p3} &= w_{31}x_{p1} + w_{32}x_{p2} + w_{33}x_{p3}.\end{aligned}$$

Unter Berücksichtigung einer zwischengeschalteten Aktivierungsfunktion

## Mathematikaufgabe 73

$$f(y_{pi}) = f(w_{i1}x_{p1} + w_{i2}x_{p2} + w_{i3}x_{p3}) = f\left(\sum_{j=1}^3 w_{ij}x_{pj}\right)$$

lautet die Fehlerfunktion

$$E = \frac{1}{2} \sum_{p=1}^3 \sum_{i=1}^3 \left( \eta_{pi} - f\left(\sum_{j=1}^3 w_{ij}x_{pj}\right) \right)^2.$$

Eine notwendige Bedingung für die Lösung dieser Optimierungsaufgabe ist, daß der Gradient der Fehlerfunktion verschwindet, d.h.

$$\nabla E = \left( \frac{\partial E}{\partial w_{11}}, \frac{\partial E}{\partial w_{12}}, \dots, \frac{\partial E}{\partial w_{33}} \right) = 0.$$

Im folgenden verzichten wir aus Gründen der Übersichtlichkeit auf die Mitführung des Musterparameters  $p$  und schreiben

$$y_i \equiv w_{i1}x_1 + w_{i2}x_2 + w_{i3}x_3 = \sum_{j=1}^3 w_{ij}x_j$$

bzw.

$$E = \frac{1}{2} \sum_{i=1}^3 \left( \eta_i - f\left(\sum_{j=1}^3 w_{ij}x_j\right) \right)^2 = \frac{1}{2} \sum_{i=1}^3 (\eta_i - f(y_i))^2.$$

Damit ergibt sich für die partiellen Ableitungen nach mehrmaliger Anwendung der Kettenregel

$$\begin{aligned} \frac{\partial E}{\partial w_{ij}} &= \frac{1}{2} \frac{\partial}{\partial w_{ij}} \sum_{k=1}^3 (\eta_k - f(y_k))^2 \\ &= \frac{1}{2} \sum_{k=1}^3 \frac{\partial}{\partial (\eta_k - f(y_k))} (\eta_k - f(y_k))^2 \frac{\partial}{\partial y_k} (\eta_k - f(y_k)) \frac{\partial}{\partial w_{ij}} \sum_{l=1}^3 w_{kl}x_l. \end{aligned}$$

Nach Ausführen der ersten Differentiation verbleibt

$$\frac{\partial E}{\partial w_{ij}} = - \sum_{k=1}^3 (\eta_k - f(y_k)) \frac{\partial f(y_k)}{\partial y_k} \frac{\partial}{\partial w_{ij}} \sum_{l=1}^3 w_{kl}x_l$$

und nach Ausführung der letzten

$$\frac{\partial E}{\partial w_{ij}} = -(\eta_i - f(y_i)) \frac{\partial f(y_i)}{\partial y_i} x_j.$$

Als Aktivierungsfunktion im offenen Intervall  $]0, 1[$  wählen wir die sogenannte Sigmoidfunktion. Diese lautet für Neuron  $i$ :

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

wobei deren Ableitung gegeben ist durch

$$\begin{aligned} f'(y_i) &= \frac{\partial f(y_i)}{\partial y_i} = \frac{e^{-y_i}}{(1 + e^{-y_i})^2} = \frac{1}{1 + e^{-y_i}} \frac{e^{-y_i}}{1 + e^{-y_i}} \\ &= f(y_i) \frac{1 + e^{-y_i} - 1}{1 + e^{-y_i}} = f(y_i)(1 - f(y_i)). \end{aligned}$$

Damit ergibt sich für die partiellen Ableitungen als endgültiges Resultat

$$\frac{\partial E}{\partial w_{ij}} = -(\eta_i - f(y_i))f(y_i)(1 - f(y_i))x_j.$$

Damit der Lernfehler abnehmen kann, muß gelten:

$$\frac{\partial E}{\partial w_{ij}} < 0.$$

Die Änderung des Gewichtsvektors  $\vec{w} = (w_{11}, w_{12}, \dots, w_{33})$  in allen seinen Komponenten ist dann gegeben durch

$$\Delta w_{ij} = -\lambda \frac{\partial E}{\partial w_{ij}}$$

oder in Vektorschreibweise

$$\Delta \vec{w} = -\lambda \nabla E(\vec{w}).$$

In unserem konkreten Fall lautet also der anzuwendende Algorithmus

$$\Delta w_{ij} = \lambda(\eta_i - f(y_i))f(y_i)(1 - f(y_i))x_j \equiv \lambda \delta_i x_j,$$

wobei  $\delta_i$  eine Hilfsfunktion für die Ausgangsgrößen ist. Mit einem entsprechenden Startwert  $w_{ij}(0)$  und der an dieser Stelle genommenen Ableitung

$$\Delta w_{ij}(0) = -\lambda \left. \frac{\partial E}{\partial w_{ij}} \right|_{w_{ij}=w_{ij}(0)} = \lambda \delta_i(0) x_j$$

läßt sich dann das neue Verbindungsgewicht

$$w_{ij}(1) = w_{ij}(0) + \Delta w_{ij}(0)$$

berechnen. Dessen Ableitung an der Stelle  $w_{ij}(1)$ ,

## Mathematikaufgabe 73

---

$$\Delta w_{ij}(1) = -\lambda \left. \frac{\partial E}{\partial w_{ij}} \right|_{w_{ij}=w_{ij}(1)} = \lambda \delta_i(1) x_j,$$

gestattet wiederum die Berechnung eines noch näher am Minimum befindlichen Gewichts, i.e.

$$w_{ij}(2) = w_{ij}(1) + \Delta w_{ij}(1),$$

usw. Setzt man dieses Verfahren iterativ fort, landet man im Idealfall im globalen Minimum des Fehlergebirges. Da es nur ein globales Minimum gibt, sollte der ideale Startwert schon in relativ guter Nähe zu diesem Minimum sein, sonst kann es passieren, daß das Gradientenabstiegsverfahren in ein lokales Nebenminimum führt oder in eine Ebene ausläuft oder auf einen Sattelpunkt trifft. Auch Oszillationen sind möglich. Es gibt offenbar auch im Gehirn des Menschen Entscheidungen, die einem nicht leicht fallen und die daher durchaus falsch getroffen werden können. Hier aber bereits von einem freien Willen zu reden ist ebenso gewagt wie zweifelhaft. Natürlich weiß der Mensch, daß er mehrere Möglichkeiten hat, zwischen denen er entscheiden kann, allerdings – die Entscheidung trifft nicht er, sondern sein neuronales Netz nach ausgiebiger Abwägung. Führt die Abwägung dazu, daß die Neuronen feuern, wird die Entscheidung ausgeführt, andernfalls unterbleibt sie. Nicht der Mensch bestimmt über die Natur, sondern die Natur über den Menschen. In dieser Hinsicht kann man wohl sagen, daß sich die Natur die Freiheit der Entscheidung offenbar vorbehält, wenn etwa die Anfangsbedingungen der Neuronen zu vage sind.