Universität Kassel
Fachbereich Mathematik/Informatik
Fachgebiet Wissensverarbeitung
Hertie-Stiftungsprofessur
Wilhelmshöher Allee 73
34121 Kassel

Email: hotho@cs.uni-kassel.de
Tel.: ++49 561 804-6252

Dr. Andreas Hotho
Prof. Gerd Stumme

12.07.04

Knowledge Discovery

Lösungsblatt 5

Sommersemester 2004

Aufgabe 1: Künstliche Neuronale Netze

a) Gegeben sei folgende Funktion:

X_1	X_2	O
0	0	1
0	1	1
1	0	1
1	1	0

Lernen Sie die obige Funktion mit einem einfachen Perzeptron. Verwenden sie dabei folgende Output-Funktion:

$$o = \begin{Bmatrix} 1 \text{ if } \sum_{i=0}^{n} w_i x_i > 0 \\ 0 \text{ otherwise} \end{Bmatrix}$$

Dabei sei $w_0 = -s,\ x_0 = 1$.

Trainieren sie das Perzeptron indem Sie nacheinander die Zeilen der obigen Tabelle anlegen und die Gewichte gemäß der folgenden Formel anpassen:

$$w_{i,neu} = w_{i,alt} + \eta x_i (o_{soll} - o_{ist})$$

Verwenden sie dabei die folgenden Anfangswerte:

$$s = 0.5, w_0 = -0.5, w_1 = 0.5, w_2 = 0.5, \eta = 0.5$$

Anlegen von (0,0):

0.5 * 0 + 0.5*0 – 0.5*1=-0.5 < 0 => o(0,0)=0

$$w_{0,neu} = -0.5 + 0.5 * 1 * (1 - 0) = 0$$
$$w_{1,neu} = 0.5 + 0.5 * 0 * (1 - 0) = 0.5$$
$$w_{2,neu} = 0.5 + 0.5 * 0 * (1 - 0) = 0.5$$

Anlegen von (0,1):

0.5*0+0.5*1+0*1=0.5 => o(0,1) =1

Ergebnis korrekt, also keine Gewichtsanpassung!

Anlegen von (1,0):

0.5*1+0.5*+0*1=0.5 => o(1,0)=1

Ergebnis korrekt, also keine Gewichtsanpassung!

Anlegen von (1,1):

0.5*1+0.5*1+0*1=1 => o(1,1)=0

$$w_{0,neu} = 0 + 0.5 * 1 * (0 - 1) = -0.5$$
$$w_{1,neu} = 0.5 + 0.5 * 1 * (0 - 1) = 0$$
$$w_{2,neu} = 0.5 + 0.5 * 1 * (0 - 1) = 0$$

b) Um welche logischen Operator handelt es sich bei der obigen Funktion? Wählen Sie die Gewichte für das obige Perzeptron derart, dass es tatsächlich die gewünschte Funktion berechnet.

Es handelt sich dabei um den NAND-Operator. Eine mögliche Wahl der Gewichte ist z.B.

$$s = -0.8, w_1 = -0.5, w_2 = -0.5$$

c) Gegeben sei nun folgende Funktion

X_1	X_2	O
0	0	0
0	1	1
1	0	1
1	1	0

Wählen Sie für das einfache Perzeptron die Gewichte von Hand derart, dass es diese Funktion berechnet.

Geht nicht... ;-)
Das ist nämlich genau der XOR-Operator!!!
Aus der Vorlesung wissen wir, dass das Problem nicht linear separierbar ist.

e) Was ist das Prinzip der Backpropagation?

Das Prinzip des Backpropagation-Algorithmus besteht darin, den aktuellen Fehler (tatsächlicher Output Minus gewollter Input) vom Output bis zum Input über jedes Gewicht zurückzuverfolgen und diese Information entsprechend vom Output zum Input zurückzupropagieren. (back propagate). Für jedes Gewicht wird sein Anteil zum Fehler berechnet und danach mit der Gradientenmethode das Gewicht vergrößert oder verkleinert.

f) Nennen Sie Vorteile/Nachteile von KNN gegenüber anderen Klassifikatoren, die Sie im Rahmen der Vorlesung kennengelernt haben, wie z.B. Entscheidungsbäumen.

Vorteile:

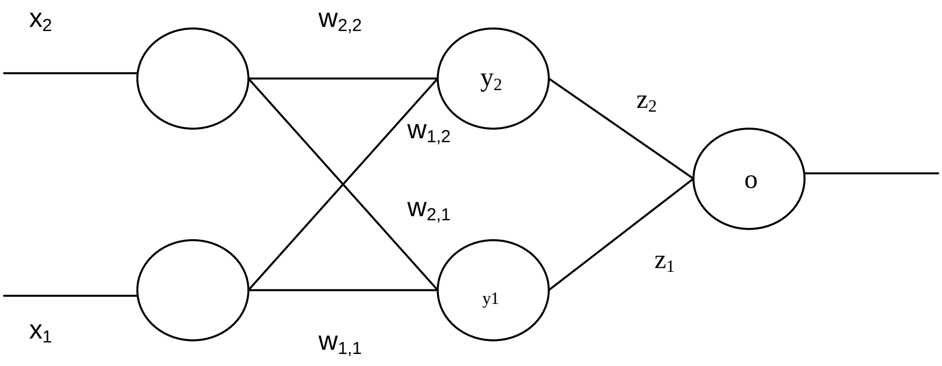
- gute Mustererkenner
- verarbeiten fehlerhafte Daten
- funktionieren für reellwertige Funktionen

Nachteile:

- lange Trainingszeiten
- Konvergenz auf eine Lösung kann nicht garantiert werden
- Ergebnisse können nicht interpretiert werden (blackbox)
- unklare Entwurfsmethodik

Aufgabe 2: Künstliche Neuronale Netze

a) Gegeben sei folgendes Neuronale Netz, bei dem die Aktivierungsfunktion das Skalarprodukt und die Ausgabefunktion die Sigmoidfunktion ist (Steigung ist 1).



Berechnen Sie nun für die XOR-Funktion (gegeben durch folgende Tabelle) durch Backpropagation die Gewichtsänderungen nach einer Iteration indem sie die letzte Zeile der Tabelle an das Netz anlegen.

X_1	X_2	O
0	0	0
0	1	1
1	0	1
1	1	0

Die Gewichte seien zufällig wie folgt initialisiert worden:

$$w_{1,1} = 0{,}5$$
$$w_{1,2} = 0{,}75$$
$$w_{2,1} = 0{,}25$$
$$w_{2,2} = 0{,}25$$
$$z_1 = 0{,}5$$
$$z_2 = 0{,}5$$

Die Sigmoid-Funktion sieht wie folgt aus: $\theta(x) = \frac{1}{1+e^{-nx}}$

Ihre Ableitung ist gegeben durch: $\theta'(x) = \theta(x)(1-\theta(x))$

Passen Sie die Gewichte durch folgende Formeln (aus dem Vorlesungsskript) an:

$$\Delta z_i = \frac{\delta F}{\delta z_i} = 2 \cdot (t-o) \cdot \theta'(Z \cdot Y) \cdot y_i$$

$$\Delta w_{i,j} = \frac{\delta F}{\delta w_{i,j}} = 2 \cdot (t-o) \cdot \theta'(Z \cdot Y) \cdot z_i \cdot \theta'(W_i \cdot X) \cdot x_j$$

Wir berechnen erst die Δz_i's . Dazu berechnen wir vorerst $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, sowie den Output o:

$$Y = \theta(W \cdot X) = \theta\begin{pmatrix} 0.5 & 0.75 \\ 0.25 & 0.25 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \theta\begin{pmatrix} 1.25 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.62 \end{pmatrix}$$

$$o = \theta(Z \cdot Y) = \theta\left(\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0.78 \\ 1.62 \end{pmatrix}\right) = \theta(0.7) = 0.67$$

Weiter berechnen wir:

$$2 \cdot (t-o) \cdot \theta'(Z \cdot Y) = 2 \cdot (t-o) \cdot \theta(Z \cdot Y) \cdot (1-\theta(Z \cdot Y)) = 2 \cdot (0-0.67) \cdot 0.67 \cdot (1-0.67) = -0{,}3$$

Und damit:

$$\Delta z_1 = -0.3 \cdot 0.78 = -0.23$$
$$\Delta z_2 = -0.3 \cdot 1.62 = -0.19$$

$z_1 = 0.27$ und $z_2 = 0.31$

Nun berechnen wir die $\Delta w_i{}'s$:

Wir erhalten $\Delta w_{i,j} = -0.3 \cdot z_i \cdot \theta'(W_i \cdot X) \cdot x_j$

Wir berechnen also zunächst die $W_i \cdot X$ Skalarprodukte:

$$W_1 \cdot X = \begin{pmatrix} 0.5 \\ 0.75 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1.25$$

$$W_2 \cdot X = \begin{pmatrix} 0.25 \\ 0.25 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0.5$$

Damit ergibt sich:

$$\Delta w_{1,1} = -0.3 \cdot z_1 \cdot \theta(1.25) \cdot (1 - \theta(1.25)) \cdot x_1 = -0.3 \cdot 0.5 \cdot 0.78 \cdot 0.22 \cdot 1 = -0.09$$
$$\Delta w_{1,2} = -0.3 \cdot z_1 \cdot \theta(1.25) \cdot (1 - \theta(1.25)) \cdot x_2 = -0.3 \cdot 0.5 \cdot 0.78 \cdot 0.22 \cdot 1 = -0.09$$
$$\Delta w_{2,1} = -0.3 \cdot z_2 \cdot \theta(0.5) \cdot (1 - \theta(1.25)) \cdot x_1 = -0.3 \cdot 0.5 \cdot 0.62 \cdot 0.38 \cdot 1 = -0.12$$
$$\Delta w_{2,2} = -0.3 \cdot z_2 \cdot \theta(0.5) \cdot (1 - \theta(1.25)) \cdot x_2 = -0.3 \cdot 0.5 \cdot 0.62 \cdot 0.38 \cdot 1 = -0.12$$

und

$$w_{1,1} = w_{1,1} + \Delta w_{1,1} = 0.41$$
$$w_{1,2} = w_{1,2} + \Delta w_{1,2} = 0.66$$
$$w_{2,1} = w_{2,1} + \Delta w_{2,1} = 0.13$$
$$w_{2,2} = w_{2,2} + \Delta w_{2,2} = 0.13$$

Neuer Output bei (1,1) ist 0.57 und nicht mehr 0.67 durch die Anpassung der Gewichte.

b) Schreiben Sie für gegebene Gewichte $w_{1,1},\dots,w_{2,2},z_1,z_2$ explizit die Auswertungsfunktion $f : \mathrm{R}^2 \to \mathrm{R}$ des neuronalen Netzes hin.

$$f(x_1, x_2) = \theta(z_1 \cdot \theta(w_{1,1} \cdot x_1 + w_{2,1} \cdot x_1) + z_2 \cdot \theta(w_{2,1} \cdot x_2 + w_{2,2} \cdot x_2)) = \frac{1}{1 + e^{-(z_1 \cdot \theta(\dots))}} = \dots$$

c) Sei $(x_1^i, x_2^i, o^i)_{i=1,\dots,n}$ ein Trainigsdatensatz mit $x_1^i, x_2^i, o^i \in \mathrm{R}$ für i=1,...,n. Beschreiben Sie die Lernphase als Optimierungsproblem.

Das Ziel ist nun eben, die Gewichte so zu wählen, dass $\sum_i \left(f(x_1, x_2) - o_i \right)^2$ minimal wird,

d.h. der quadratische Fehler minimal wird.