Universität des Saarlandes
FR Informatik

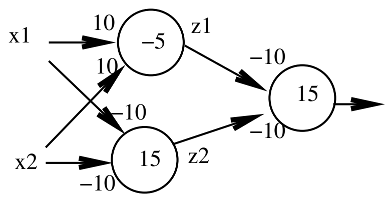
Kurt Mehlhorn **WiSe 2018/19**

Übungen zu Ideen der Informatik

https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter18/ideen/

Blatt 12 Abgabeschluss: 28.1.2019

Aufgabe 1 (*5 Punkte*) Vervollständigen Sie die Tabelle und geben Sie an welche logische Funktion das abgebildete Netzwerk berechnet wird?



x_1	x_2	$z_1 =$	$z_1 \approx$	$z_2 =$	$z_2 \approx$	$o =$	$o \approx$
0	0	$g(-5)$	0				
0	1						
1	0						
1	1						

Lösung:

x_1	x_2	$z_1 =$	$z_1 \approx$	$z_2 =$	$z_2 \approx$	$o =$	$o \approx$
0	0	$g(-5)$	0	$g(15)$	1	$g(5)$	1
0	1	$g(5)$	1	$g(5)$	1	$g(-5)$	0
1	0	$g(5)$	1	$g(5)$	1	$g(-5)$	0
1	1	$g(15)$	1	$g(-5)$	0	$g(5)$	1

Das Netz berechnet die Funktion $x_1 \equiv x_2$. (Die Antworten $x_1 = x_2$ oder x_1 gleich x_2 sind auch OK.)

Aufgabe 2 (*5 Punkte*) Betrachte die Funktion $z = z(x, y) = x^2 + 2y^2$.

a) Was sind die Ableitungen von z nach x und y? Der Gradient ∇z von z ist der Vektor bestehend aus den beiden Ableitungen. Was ist der Gradient ∇z?

   Lösung: $\partial z/\partial x = 2x$ and $\partial z/\partial y = 4y$. Daher $\nabla z = \begin{pmatrix} 2x \\ 4y \end{pmatrix}$.

b) Wie sehen die Höhenlinien $z = c$ aus?. Dabei ist c ein fester Wert? Was ist der Zusammenhang zwischen Höhenlinien und Gradient?

   Lösung: Die Höhenlinien sind Ellipsen mit Halbachsen der Länge 1 und $1/\sqrt{2}$. Der Gradient steht senkrecht auf der Höhenlinie.

c) Gradientenabstieg: Wir beginnen mit einem Punkt (x_0, y_0) und definieren dann eine Folge (x_i, y_i), $i \geq 1$, durch $(x_{i+1}, y_{i+1}) = (x_i, y_i) - h\nabla z(x_i, y_i) = (x_i - 2hx_i, y_i - 4hy_i)$. Dabei ist h die Schrittweite.

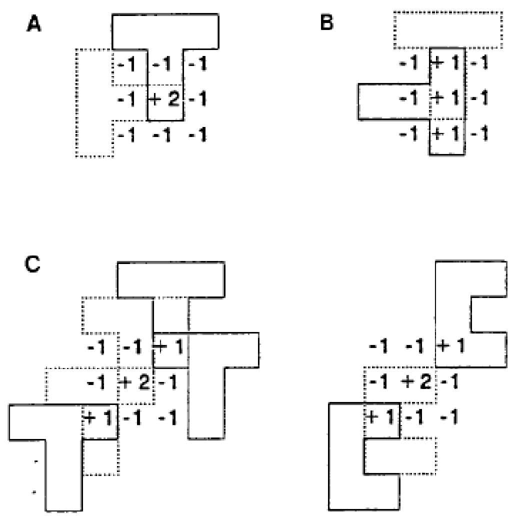
Starten sie mit $(x_0, y_0) = (2, 3)$ und bestimmen sie die ersten vier Schritte bei Verwendung der Schrittweite $h = 1/4$. Das Minimum ist der Punkt $(0, 0)$. Wie nahe kommen sie ihm in 10 Schritten?

Lösung: We have $x_{i+1} = (1 - 2h)x_i = 1/2x_i$. Also $x_0 = 2$, $x_1 = 1$, $x_2 = 1/2$, $x_3 = 1/4$, $x_4 = 1/8$ und $x_{10} = 2^{-10} \cdot 2 = 2/1014$. Für y erhalten wir: $y_1 = (1 - 4h)y_0 = 0$ und dann $y_3 = y_2 = y_1 = 0$.

d) Was passiert, wenn sie die Schrittweite $h = 1$ wählen?

Lösung: $x_{i+1} = (1 - 2h)x_i = -x_i$ und $y_{i+1} = (1 - 4h)y_i = -3y_i$. Also alterniert der x-Wert zwischen $+2$ und -2. Der y-Wert explodiert.

Aufgabe 3 (*15 Punkte*) In der Vorlesung haben wir das Netz gesehen, das C und T unterscheiden kann. Ich habe in der Vorlesung erklärt, wie die Filter A und D funktionieren. Erklären Sie, wie die Filter B und C funktionieren.



a) Welche Werte können die Filter B und C liefern bei Eingabe C bzw. T.

b) Was muss das Ausgabeneuron leisten?

Lösung:

Filter B: Beim T liefert mindestens ein Neuron der Eingabeschicht einen Wert ≥ 2. Bei Eingabe C ist der Wert immer ≤ 1.

Wenn das T normal oder auf dem Kopf steht und die mittlere Spalte mit 2 Kästchen überlappt, bekommt man den Wert 2. Wenn das T liegt und der Balken des T mit der mittleren Spalte übereinstimmt, bekommt man den Wert 2.

Wenn das C die mittlere Spalte nicht oder nur in einem Quadrat überlappt, dann ist der Gesamtwert sicher ≤ 1. Wenn das C die mittlere Spalte in genau 2 Quadraten überlappt, dann muss es auch eines der Felder mit Wert -1 überlappen. Also ist der Gesammtwert ≤ 1. Wenn das C die mittlere Spalte in 3 Quadraten überlappt, dann steht es aufrecht und überlappt auch zwei Felder mit Wert -1. Also ist der Gesamtwert ≤ 1.

Das Ausgabeneuron sagt T, wenn mindestens ein Neuron der ersten Schicht den Wert 2 liefert.

Filter C: Beim C liefert mindestens ein Filter den Wert -3. Beim T sind die Werte immer ≥ -2.

Wenn das C normal steht und mit dem linken Rand des Filters aligniert ist, dann ist der Wert -3. Analog für die drei anderen Lagen des C.

Nehmen wir an, das T steht normal. Wenn es drei Kästchen -1 überlappt, dann muss es muss sowohl der Stamm als auch der Balken des T den Filter überlappen. Also liegt der Stamm entweder in der linken Spalte des Filters (dann Gesamtwert -2) oder in der mittleren Spalte (dann Gesamtwert 0) oder in der rechten Spalte (Gesamtwert -2). Analog argumentiert bei den anderen Lagen des T.

Das Ausgabeneuron muss also nur entscheiden, ob es ein Eingabeneuron gibt mit Wert -3 gibt.

Aufgabe 4 (*10 Zusatzpunkte Punkte*)

a) Neuronale Netze benutzen die Sigmoidfunktion $g(z) = \frac{1}{1+e^{-z}}$ als Approximation für den Sprung von 0 nach 1 an der Stelle 0. Verifizieren Sie $g(z) + g(-z) = 1$ und $g'(z) = g(z)(1 - g(z))$ für alle z.

Lösung:

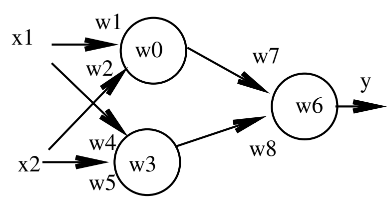
$$g(z) + g(-z) = \frac{1}{1+e^{-z}} + \frac{1}{1+e^{z}} = \frac{1+e^{-z}+1+e^{z}}{(1+e^{-z})(1+e^{z})} = \frac{2+e^{z}+e^{-z}}{1+e^{z}+e^{-z}+1} = 1$$

$$g'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \frac{1}{1+e^{-z}} \cdot (1 - \frac{1}{1+e^{-z}}) = g(z)(1-g(z)).$$

b) Erinnern Sie sich an die Kettenregel. Wenn f und g Funktionen sind, dann

$$\frac{d}{dx} f(g(x)) = f'(g(x)) \cdot g'(x).$$

Betrachten Sie das folgende Netz mit den 9 Parametern w_0 bis w_8.



Es berechnet die Funktion

$$h_w(x) := g(w_6 + w_7 \cdot g(w_0 + w_1 x_1 + w_2 x_2) + w_8 \cdot g(w_3 + w_4 x_1 + w_5 x_2)).$$

Was sind die partiellen Ableitungen von h_w nach w_6, w_7, w_0 und w_1?

Hinweis: Definieren Sie $s_1 = w_0 + w_1 x_1 + w_2 x_2$, $s_2 = w_3 + w_4 x_1 + w_5 x_2$, $f_1 = g(s_1)$, $f_2 = g(s_2)$, $s = w_6 + w_7 f_1 + w_8 f_2$. Nutzen Sie die Funktionen g und g', um die Lösungen kompakt zu schreiben. Es ist zum Beispiel

$$\frac{\partial h_w}{\partial w_5} = g'(s) w_8 g'(s_2) x_2.$$

Lösung:

$$\frac{\partial h_w}{\partial w_6} = g'(s) \cdot 1 = g(s)(1-g(s))$$
$$\frac{\partial h_w}{\partial w_7} = g'(s) \cdot f_1 = g(s)(1-g(s)) \cdot f_1$$
$$\frac{\partial h_w}{\partial w_0} = g'(s) w_7 g'(s_1)$$
$$\frac{\partial h_w}{\partial w_1} = g'(s) w_7 g'(s_1) x_1$$

c) Sei (x, y) ein Trainingsbeispiel. Wenn w den aktuellen Parametersatz bezeichnet, dann ist der quadratische Fehler an diesem Trainingsbeispiel definiert als

$$E(w) = (y - h_w(x))^2.$$

Beachten sie, dass $h_w(x)$ die Ausgabe des Netzes an der Eingabe w ist und y die gewünschte Ausgabe ist. Verifizieren sie die folgende Formel für die Ableitung von $E(w)$ nach dem Parameter w_k.

$$\frac{\partial E(w)}{\partial w_k} = -2(y - h_w(x)) \cdot \frac{\partial h_w}{\partial w_k}(x).$$

Hinweis: Benutzen sie wieder die Kettenregel. Beachten sie dabei, dass wir $h_w(x)$ als Funktion der Parameter betrachten und NICHT als Funktion von x.

d) Was ist für unser Beispiel die Ableitung von $E(w)$ nach w_0?

Lösung:

$$\frac{\partial E(w)}{w_0} = -2(y - h_w(x)) \cdot g'(s) w_7 g'(s_1).$$

Maschinelles Lernen war spannend ☐ okay ☐ langweilig ☐
schwierig ☐ okay ☐ einfach ☐