Universität
des
Saarlandes
FR 6.2 Informatik

Prof. Dr. Kurt Mehlhorn
Michael Dirnberger

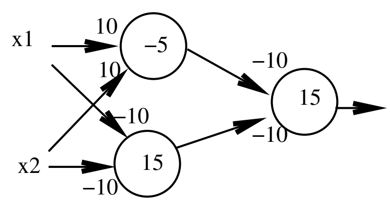
WiSe 2015/16

Übungen zu Ideen der Informatik

http://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter15/ideen/

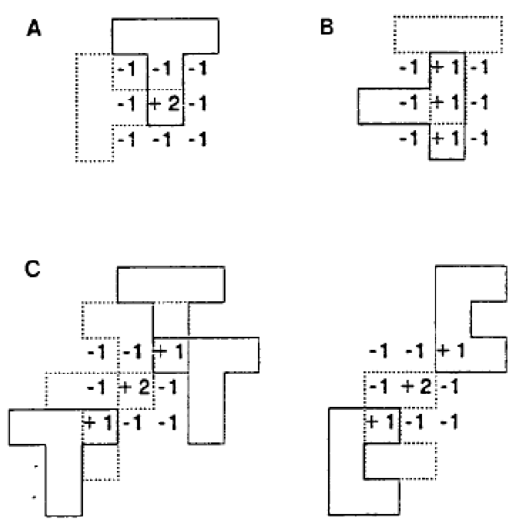
Blatt 9 Abgabeschluss: 11.1.2016

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						

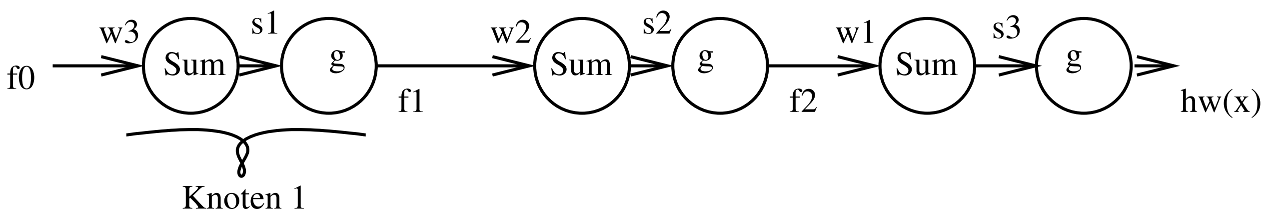
Aufgabe 2 (*10 Punkte*) In der Vorlesung haben wir das Netz gesehen, das C und T unterscheiden kann. Ich habe in der Vorlesung erklärt, wie die Filter A und D funktionieren. Erklären Sie, wie die Filter B und C funktionieren.



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

b) Was muss das Ausgabeneuron leisten?

Aufgabe 3 (*15 Punkte*) Die folgende Abbildung zeigt einen Pfad durch ein Neuronales Netz. Ich habe dabei jeden Knoten als zwei Teilknoten gezeichnet. Der erste Teilknoten bildet eine gewichtete Summe seiner Eingabeleitungen. Die Ausgabeleitung des Summenknoten geht dann in einen Knoten welcher die Sigmoidfunktion g berechnet, wobei $g(z) = \frac{1}{1+e^{-z}}$. Die Leitung f_0 trägt ein Signal, das von der Eingabe x und Parametern abhängt, die uns für diese Aufgabe nicht interessieren.



Der Knoten ganz links bildet die Summe $s_1 = w_3 f_0 + t_0$, wobei t_0 alle anderen Eingaben des Summenknoten zusammenfasst; t_0 hängt nicht von w_3 ab. Dann ist $f_1 = g(s_1)$, $s_2 = w_2 f_1 + t_1$, wobei t_1 alle anderen Eingaben des Summenknoten zusammenfasst; t_1 hängt nicht von w_3 ab. Dann $s_3 = w_1 f_2 + t_2$, wobei t_2 alle anderen Eingaben des Summenknoten zusammenfasst; t_2 hängt nicht von w_3 ab. Schließlich erhalten wir die Ausgabe $h_w(x)$ als $g(s_3)$. Die Ausgabe ist also folgende Funktion von w_3:

$$h_w(x) = g(t_2 + w_1 \cdot g(t_1 + w_2 \cdot g(t_0 + w_3 f_0))).$$

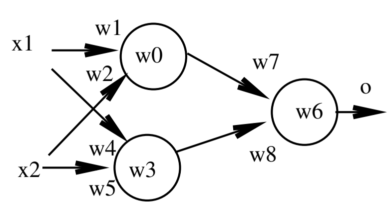
a) Verifizieren Sie $g(z) + g(-z) = 1$ und $g'(z) = g(z)(1 - g(z))$ für alle z.

b) Bestimmen Sie die partielle Ableitung von $h_w(x)$ nach w_3. Hinweis: Das ist eine Übung im Anwenden der Kettenregel. Schlagen Sie bitte notfalls nach. Es gilt:

$$\frac{\partial h_w(x)}{\partial w_3} = \frac{\partial h_w(x)}{\partial s_3} \cdot \frac{\partial s_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_3}.$$

Was ist nun $\frac{\partial h_w(x)}{\partial w_3}$? Hinweis: Formulieren Sie die Lösung als reines Produkt mit Hilfe von $g'(z)$.

c) Betrachten Sie nun konkret 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 wiederum die Funktionen $g(z)$ und $g'(z)$ um die Lösungen kompakt zu schreiben.

d) Der Gesamtfehler ist definiert als

$$E(w) = \sum_{i=1}^{N} (y_i - h_w(x^{(i)}))^2.$$

Dabei ist (x^i, y_i) das i-te Trainingsbeispiel. Bestimmen Sie die partiellen Ableitungen von $E(w)$ nach w_k. Hinweis: Wenden Sie wieder die Kettenregel an. Damit Sie dabei nicht durcheinanderkommen, definieren Sie die Funktionen $h_i(w) = h_w(x^i)$ und leiten Sie dann gemäß der Kettenregel ab.

$$\frac{\partial E(w)}{\partial w_k} = \sum_{i=1}^{N} \frac{\partial E_w}{\partial h_i} \cdot \frac{\partial h_i}{\partial w_k}.$$

e) Algorithmus für den Newtonschritt.

Für alle Trainingsbeispiele (x^i, y_i) tue.

(a) Setze $x = x^{(i)}$ und $y = y_i$.

(b) Berechne s_1, s_2, f_1, f_2, s und $h_i = h_w(x^{(i)})$. Beachte, dass man dazu im wesentlichen von den Eingängen zum Ausgang über das Netz läuft.

(c) Berechne $\delta = 2(y - h)$.

(d) Berechne $\partial h_i / \partial w_k$ für $k = 0, \ldots, 8$ gemäß der Formeln unter d) und e) und multipliziere jeweils mit δ. Beachte, dass man dazu im wesentlichen von den Ausgang zum den Eingängen über das Netz läuft (die Größe $g'(s)$ braucht man für alle partiellen Ableitungen, die Größe $g'(s)w_7g'(s_1)$ braucht man für die partiellen Ableitungen nach w_0 bis w_2). Daher heißt der Algorithmus Backpropagation.

Summiere für jedes w_k alle diese Änderungen auf und multipliziere noch mit der Schrittweite der Newtoniteration. Ändere w_k um diesen Betrag.

In der obigen Beschreibung köennte ein Fehler sein. Finden Sie ihn.