Universität des Saarlandes
FR 6.2 Informatik

Prof. Dr. Kurt Mehlhorn
Dr. Antonios Antoniadis
André Nusser

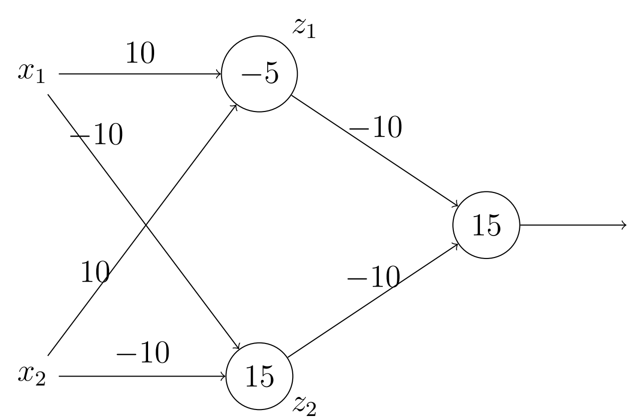
WiSe 2017/18

Übungen zu Ideen der Informatik

http://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter17/ideen/

Blatt 12 Abgabeschluss: Das Blatt muss nicht abgegeben werden

Aufgabe 1 (*0 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 (*0 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?

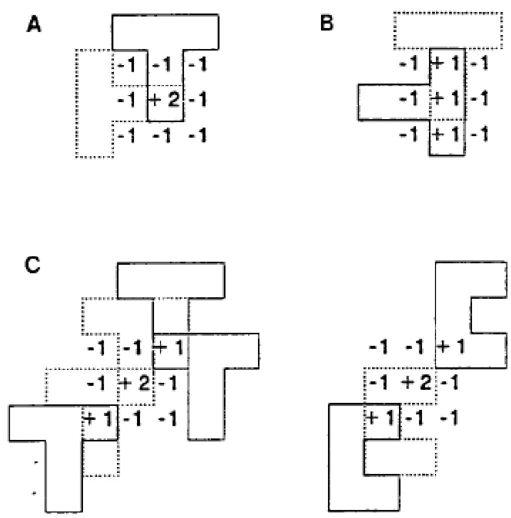
b) Wie sehen die Höhenlinien $z = c$ aus, wobei c ein fester Wert ist? Was ist der Zusammenhang zwischen Höhenlinien und Gradient?

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?

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

Aufgabe 3 (*0 Punkte*) In der Vorlesung haben wir das Netz gesehen, das C und T unterscheiden kann. Es wurde 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?

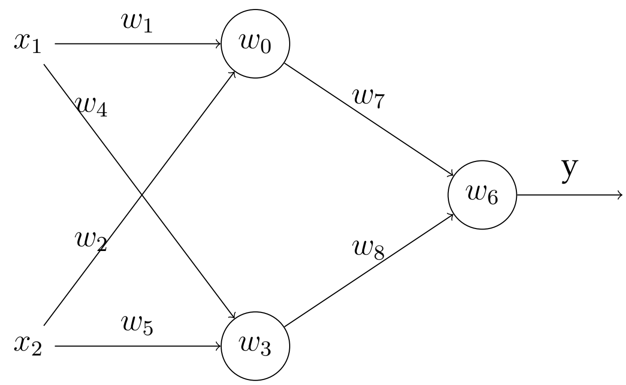
Aufgabe 4 (*0 Punkte*) [Schwierig]

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.

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

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

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?