

Einführung in die Künstliche Intelligenz

SS09 - Prof. Dr. J. Fürnkranz



TECHNISCHE
UNIVERSITÄT
DARMSTADT

6. Übungsblatt (14.07.2009)

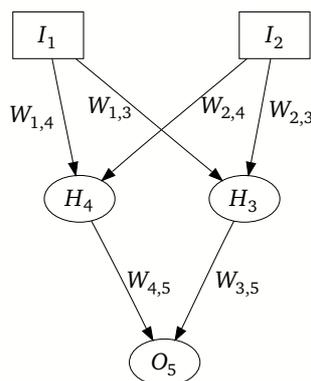
Aufgabe 1 Perceptrons, Neuronale Netze

a) Die Funktion f sei folgendermaßen definiert:

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

Was ist an dieser Funktion ungewöhnlich ?

- b) Konstruieren Sie ein neuronales Netz, das die obige Funktion f berechnet. Spezifizieren Sie dabei genau die verwendeten Knoten (Gewichte, Aktivierungsfunktion, ...), die Sie verwenden.
- c) Betrachten Sie das folgende neuronale Netz.



Die benutzte Aktivierungsfunktion ist die Sigmoidfunktion ($g(x) = \frac{1}{1+e^{-x}}$), die Gewichte sind im Ausgangszustand wie folgt:

$w_{1,3}$	$w_{1,4}$	$w_{2,3}$	$w_{2,4}$	$w_{3,5}$	$w_{4,5}$
0.4	0.5	0.2	-0.7	-0.2	0.7

Als Element einer Trainingsmenge wird $(I_1 = 1, I_2 = 0, T = 0)$ angelegt, wobei T das korrekte Ergebnis für die Eingabe $I_1 = 1, I_2 = 0$ beschreibt. Berechnen Sie zunächst die Ausgabe des Netzes und den Fehler. Geben Sie dann die neuen Gewichte $w_{1,3}$ und $w_{3,5}$ nach dem BACKPROPAGATION-Algorithmus unter Verwendung einer Lernrate von $\alpha = 0.6$ an.

(Beachten Sie, daß alle Bias Weight-Werte in dieser Aufgabe auf 0 gesetzt sind und nicht eingezeichnet wurden.)

Aufgabe 2 Reinforcement Learning

Ein Agent bewegt sich in einer einfachen deterministischen Welt, die wie folgt angeordnet ist:

a	b	c
d	e	f
g	h	i

Der Agent kann sich jeweils ein Feld nach unten, oben, links, oder rechts bewegen, falls dort ein Feld ist. Jeder Schritt kostet 0.1 Punkte. Wenn der Agent im Feld f landet, erhält er einen Reward von 1 Punkt und kann sich von dort nicht mehr wegbewegen. Auf allen anderen Feldern erhält er einen Reward von 0 Punkten.

Benutzen sie im Folgenden als Discountfaktor $\gamma = 0.9$.

- a) Formulieren Sie die Reward-Funktion, d.h. geben Sie für alle möglichen Zustands-Aktions Paare (s, a) die unmittelbare Belohnung $r(s, a)$ an.
- b) Berechnen Sie die Bewertungsfunktion $V^\pi(s)$ für alle Zustände s , wobei Policy π wie folgt definiert ist:

- wenn dies möglich ist, gehe nach oben; ansonsten:
- wenn dies möglich ist, gehe nach rechts; ansonsten:
- wenn dies möglich ist, gehe nach unten; ansonsten:
- gehe nach links

→	→	↓
↑	↑	
↑	↑	↑

- c) Welche Änderung würde POLICYIMPROVEMENT an der Strategie π für das Feld e vornehmen ? (Benutzen Sie Teilaufgabe a und b)
- d) Überlegen Sie sich für jedes Feld s , welches ein optimaler Weg zum Ziel wäre. Berechnen Sie damit die optimale Bewertung $V^*(s)$ für dieses Feld. Bestimmen Sie zusätzlich die optimale $Q(s, a)$ -Funktion für alle möglichen Zustands-Aktion Paare (s, a) .
- e) Bilden Sie aus der optimalen Q -Funktion eine optimale Policy.
- f) Versuchen Sie, mittels Q-LEARNING die Q -Funktion direkt zu lernen, indem Sie den Agenten auf ein zufällig gewähltes Anfangsfeld stellen und die jeweils beste Aktion nach der momentanen Q -Funktion ausführen (bei Gleichheit zufällige Auswahl), die Update Regel anwenden, bis der Agent am Ziel angekommen ist und das ganze bis zur Konvergenz wiederholen. Benutzen Sie als Lernrate $\alpha = 1$.