

# Künstliche Intelligenz

## Übungsblatt #6

### Lernen

### Version 1.0

Prof. Dr. J. Fürnkranz, Dr. G. Grieser

## Aufgabe 6.1

Es sei das folgende Neuronale Netz gegeben:

- das Neuron 0 ist ein Biasneuron, das immer den Wert +1 liefert.
- die Neuronen 1,2,3,4 sind die Eingabeneuronen, das Neuron 8 ist das Ausgabeneuron
- das Neuron 8 ist durch die Aktivierungsfunktion  $g_8(in_8) = \frac{1}{1+e^{-in_8}}$  beschrieben.
- die Neuronen 5, 6 und 7 sind durch die Aktivierungsfunktion  $g_i(in_i) = 0, 1 \cdot in_i$  beschrieben

Die Gewichte sind durch die folgende Matrix  $[w_{i,j}]_{i,j}$  gegeben:

$$\begin{array}{c} i \downarrow \\ \left[ \begin{array}{ccccccccc} & & & & j \rightarrow & & & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & -0,5 & 1,2 & 3 & \\ 0 & 0 & 0 & 0 & 0 & 2 & 0,5 & 0,3 & -0,8 & \\ 0 & 0 & 0 & 0 & 0 & -1,2 & 1,5 & 3 & 7 & \\ 0 & 0 & 0 & 0 & 0 & 3 & 0,4 & -3 & -4 & \\ 0 & 0 & 0 & 0 & 0 & -2 & 4 & 0,1 & 2 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \end{array} \right] \end{array}$$

- In Russel/Norvig ist das Bias-Neuron so definiert, daß es immer den Wert  $-1$  liefert. Diskutieren Sie den Unterschied in der Leistungsfähigkeit/Repräsentationsfähigkeit von Netzen mit Biasneuron  $+1$  bzw.  $-1$ .
- Visualisieren Sie die Netzwerkstruktur einschl. der Gewichte.
- Was ist die Ausgabe des Netzwerkes für die Eingabe  $\langle 0, 1, 2, 3 \rangle$ ?

- d) Sei nun die Zielklassifikation für die Instanz  $\langle 0, 1, 2, 3 \rangle$  der Wert 0. Passen Sie die Gewichte aller Neuronen mittels des Backpropagation-Algorithmus an, Als Lernrate benutzen Sie  $\alpha = 0,1$ .
- e) Berechnen Sie den Ausgabewert des neuen Netzes und vergleichen Sie ihn mit der ursprünglichen Berechnung.

## Aufgabe 6.2

Ein Agent bewegt sich in einer einfachen 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 Punkt. 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.

Als Discountfaktor setzen wir  $\gamma = 0,9$ .

- a) Formulieren Sie die Reward-Funktion.
- b) Berechnen Sie die Bewertungsfunktion  $V^\pi(s)$  für die Strategie  $\pi$ :
- 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) Berechnen Sie die optimale Bewertungsfunktion  $V^*(s)$ .
- d) Bestimmen Sie die  $Q$ -Funktion.
- e) Geben Sie eine optimale Policy an.
- f) Versuchen Sie, mittels  $Q$ -Learning die  $Q$ -Funktion 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), bis der Agent am Ziel angekommen ist und das ganze bis zur Konvergenz wiederholen. Als Lernrate können Sie 1 annehmen.