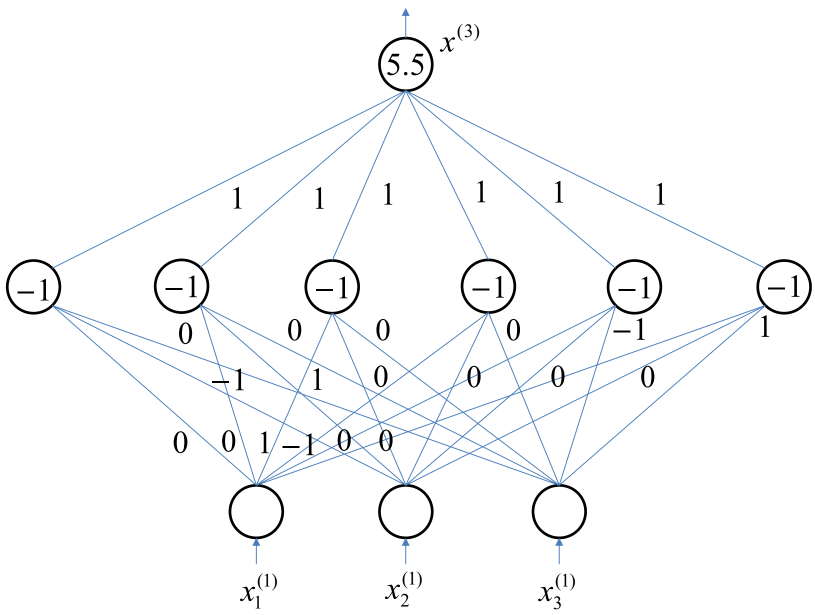
Aufgabe: Separieren Sie die Musterklasse eines Würfels der Kantenlänge 2 anhand einer ausreichenden Zahl verdeckter Knoten.

Lösung: Zur Musterklasse sollen alle Punkte $x_i^{(1)}$ innerhalb des Kubus der Kantenlänge 2 gehören, die der Bedingung $\left[-1 < x_1^{(1)} < 1;\ -1 < x_2^{(1)} < 1;\ -1 < x_3^{(1)} < 1\right]$ genügen. Um diese Klasse zu separieren, sind 6 verdeckte Knoten notwendig, die jeweils eine Begrenzungsebene festlegen.



Mit den Eingangsgewichten zu den verdeckten Knoten $\mathbf{w}_i^{(1)} = (w_{i0}^{(1)}, w_{i1}^{(1)}, w_{i2}^{(1)}, w_{i3}^{(1)})$ für $i = 1, \ldots, 6$ und den Ausgangsgewichten $\mathbf{w}^{(2)} = (w_0^{(2)}, w_1^{(2)}, w_2^{(2)}, w_3^{(2)}, w_4^{(2)}, w_5^{(2)}, w_6^{(2)})$ wird durch die Vorschrift

$$
\begin{aligned}
x_1^{(2)} &= g\left(w_{10}^{(1)}x_0^{(1)} + w_{11}^{(1)}x_1^{(1)} + w_{12}^{(1)}x_2^{(1)} + w_{13}^{(1)}x_3^{(1)}\right)\\
x_2^{(2)} &= g\left(w_{20}^{(1)}x_0^{(1)} + w_{21}^{(1)}x_1^{(1)} + w_{22}^{(1)}x_2^{(1)} + w_{23}^{(1)}x_3^{(1)}\right)\\
x_3^{(2)} &= g\left(w_{30}^{(1)}x_0^{(1)} + w_{31}^{(1)}x_1^{(1)} + w_{32}^{(1)}x_2^{(1)} + w_{33}^{(1)}x_3^{(1)}\right)\\
x_4^{(2)} &= g\left(w_{40}^{(1)}x_0^{(1)} + w_{41}^{(1)}x_1^{(1)} + w_{42}^{(1)}x_2^{(1)} + w_{43}^{(1)}x_3^{(1)}\right)\\
x_5^{(2)} &= g\left(w_{50}^{(1)}x_0^{(1)} + w_{51}^{(1)}x_1^{(1)} + w_{52}^{(1)}x_2^{(1)} + w_{53}^{(1)}x_3^{(1)}\right)\\
x_6^{(2)} &= g\left(w_{60}^{(1)}x_0^{(1)} + w_{61}^{(1)}x_1^{(1)} + w_{62}^{(1)}x_2^{(1)} + w_{63}^{(1)}x_3^{(1)}\right)\\
x^{(3)} &= g\left(w_0^{(2)}x_0^{(2)} + w_1^{(2)}x_1^{(2)} + w_2^{(2)}x_2^{(2)} + w_3^{(2)}x_3^{(2)} + w_4^{(2)}x_4^{(2)} + w_5^{(2)}x_5^{(2)} + w_6^{(2)}x_6^{(2)}\right)
\end{aligned}
$$

jedem Vektor $\mathbf{x}^{(1)} = \left(x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, x_3^{(1)}\right)$ ein Mustervektor $\mathbf{X}^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}\right)$ mit den Gewichten $\mathbf{W}_i^{(1)} = (w_{i1}^{(1)}, w_{i2}^{(1)}, w_{i3}^{(1)})$ so zugeordnet, daß gilt:

$$\mathbf{W}_i^{(1)}\mathbf{X}^{(1)} = -w_{i0}^{(1)}x_0^{(1)}$$

bzw.

$$\mathbf{w}_i^{(1)} \cdot \mathbf{x}^{(1)} = \sum_{j=0}^{3} w_{ij}^{(1)} x_j^{(1)} = 0.$$

Die Schwellen werden darin durch die konstante Eingabe $x_0^{(1)} = 1$ und die Gewichte $w_{i0}^{(1)}$ berücksichtigt.

Analog werden jedem Vektor $\mathbf{x}^{(2)} = \left(x_0^{(2)}, x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}, x_5^{(2)}, x_6^{(2)}\right)$ ein Mustervektor $\mathbf{X}^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}, x_5^{(2)}, x_6^{(2)}\right)$ und Gewichte $\mathbf{W}^{(2)} = (w_1^{(2)}, w_2^{(2)}, w_3^{(2)}, w_4^{(2)}, w_5^{(2)}, w_6^{(2)})$ zugeordnet, derart daß gilt:

$$\mathbf{W}^{(2)}\mathbf{X}^{(2)} = -w_0^{(2)} x_0^{(2)}$$

bzw.

$$\mathbf{w}^{(2)} \cdot \mathbf{x}^{(2)} = \sum_{j=0}^{6} w_j^{(2)} x_j^{(2)} = 0.$$

Die Schwellen werden darin durch die konstante Eingabe $x_0^{(2)} = 1$ und das Gewicht $w_0^{(2)}$ bedient.

Das Netzwerk mit Gewichten und zugehörigen Schwellen zur Selektion der Punkte innerhalb des definierten Kubus stellt sich in folgender Tabelle dar:

i	Ebenengl.	$w_{i0}^{(1)}$	$w_{i1}^{(1)}$	$w_{i2}^{(1)}$	$w_{i3}^{(1)}$	$w_i^{(2)}$
1	$-x_2^{(1)} + 1 = 0$	1	0	-1	0	1
2	$x_2^{(1)} + 1 = 0$	1	0	1	0	1
3	$x_1^{(1)} + 1 = 0$	1	1	0	0	1
4	$-x_1^{(1)} + 1 = 0$	1	-1	0	0	1
5	$-x_3^{(1)} + 1 = 0$	1	0	0	-1	1
6	$x_3^{(1)} + 1 = 0$	1	0	0	1	1

Die Ausgabevektoren der verdeckten Schicht sowie der Ausgabeschicht sind damit gegeben durch

$$\begin{aligned}
x_1^{(2)} &= g\left(1 - x_2^{(1)}\right),\\
x_2^{(2)} &= g\left(1 + x_2^{(1)}\right),\\
x_3^{(2)} &= g\left(1 + x_1^{(1)}\right),\\
x_4^{(2)} &= g\left(1 - x_1^{(1)}\right),\\
x_5^{(2)} &= g\left(1 - x_3^{(1)}\right),\\
x_6^{(2)} &= g\left(1 + x_3^{(1)}\right),\\
x^{(3)} &= g\left(w_0^{(2)} + x_1^{(2)} + x_2^{(2)} + x_3^{(2)} + x_4^{(2)} + x_5^{(2)} + x_6^{(2)}\right).
\end{aligned}$$

Mit den Definitionen

$$g\left(1 - x_i^{(1)}\right) = \begin{cases} 1 & \text{für} \quad x_i^{(1)} \leq 1 \\ 0 & \text{für} \quad x_i^{(1)} \geq 1 \end{cases}$$

und

$$g\left(1 + x_i^{(1)}\right) = \begin{cases} 1 & \text{für} \quad x_i^{(1)} \geq -1 \\ 0 & \text{für} \quad x_i^{(1)} \leq -1 \end{cases}$$

ergibt sich für Werte $x_i^{(1)}$ innerhalb des Kubus mit einem Offset von $w_0^{(2)} = -5{,}5$ und einem Temperaturparameter $T = 0{,}1$ unter Zugrundelegung der Sigmoidfunktion ein Schwellwert von

$$\sigma(z) = \frac{1}{1 + e^{-z/T}} \approx 0{,}99.$$

Der Netzinput ist dabei gegeben durch $z = w_0^{(2)} + x_1^{(2)} + x_2^{(2)} + x_3^{(2)} + x_4^{(2)} + x_5^{(2)} + x_6^{(2)} = 0{,}5.$ Eigentlich bedarf es gar keiner Sigmoidfunktion, denn für $T \to 0$ nähert man sich ohnehin der Θ-Funktion. Im Gegensatz zur 0-1-Entscheidung der Θ-Funktion für Werte rechts oder links der Hyperebene erhält man mit der Sigmoidfunktion allerdings ein kontinuierliches Maß für den Abstand von der Hyperebene. Je größer wir den Temperaturparameter wählen, desto stärkere Abrundungen erhalten wir bei der Klassifikation, womit die Konturen des Klassenvolumens besser approximiert werden können.

