

# Grundlagen der künstlichen Intelligenz: Hausaufgabe 8

Tom Nick - 340528  
Niklas Gebauer - 340942  
Leonard Witte - 341457  
Johannes Herrmann - 341091

## Aufgabe 1

a.) Die erste drei Beispiele werden korrekt klassifiziert (wir gehen in der Tabelle von links nach rechts vor):

$$\text{sgn}(w_{11} \cdot x_1 + w_{12} \cdot x_2 - w_{10}) = \text{sgn}((-1) + (-1) + 1) = \text{sgn}(-1) = -1 = x_1 \oplus x_2$$

$$\text{sgn}(w_{11} \cdot x_1 + w_{12} \cdot x_2 - w_{10}) = \text{sgn}(1 + (-1) + 1) = \text{sgn}(1) = 1 = x_1 \oplus x_2$$

$$\text{sgn}(w_{11} \cdot x_1 + w_{12} \cdot x_2 - w_{10}) = \text{sgn}((-1) + 1 + 1) = \text{sgn}(1) = 1 = x_1 \oplus x_2$$

Nur Beispiel 4 ändert also die Gewichte:

$$w_{20} = w_{10} + \lambda y x_0 = -1 + 0.7 \cdot (-1) \cdot (-1) = -0.3$$

$$w_{21} = w_{11} + \lambda y x_1 = 1 + 0.7 \cdot (-1) \cdot 1 = 0.3$$

$$w_{22} = w_{12} + \lambda y x_2 = 1 + 0.7 \cdot (-1) \cdot 1 = 0.3$$

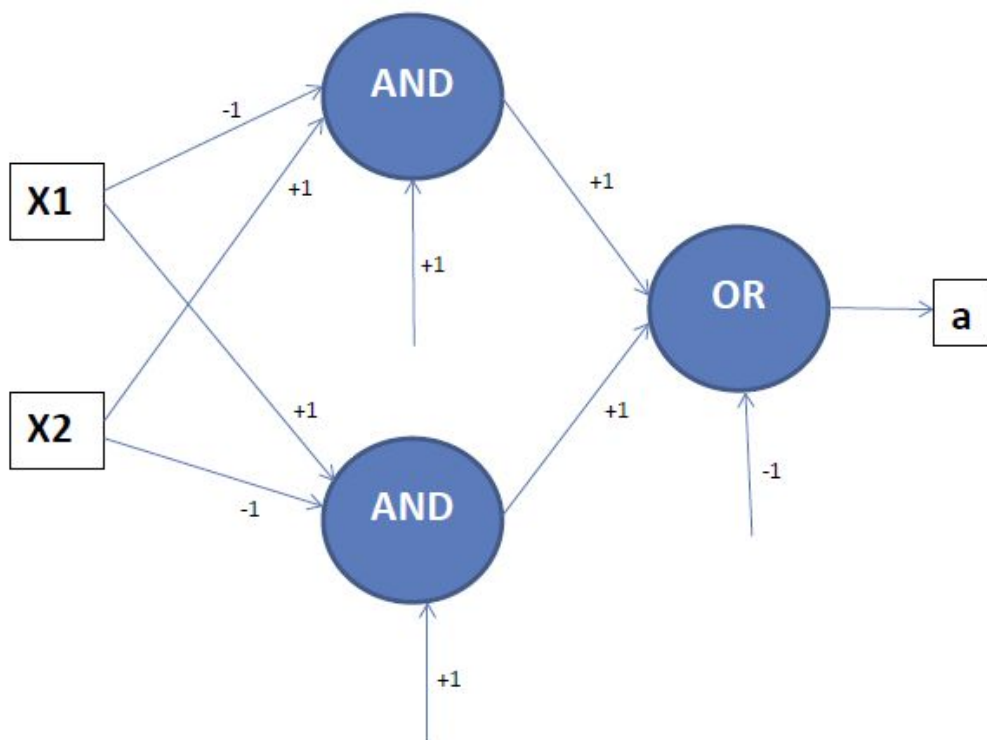
Nun werden wieder die ersten drei Beispiele korrekt klassifiziert, das 4. allerdings immer noch nicht. Da wir jedes Beispiel nur einmal betrachten sollten, können wir nun aufhören.

b.) Nein, ein Perzeptron alleine kann die XOR-Funktion nicht erlernen, da die Datenpunkte (Beispiele) nicht linear separabel sind. Ein einzelnes Perzeptron kann allerdings nur linear separable Datensätze komplett korrekt klassifizieren.

c.) Man kann die Funktion  $x_1 \oplus x_2$  umformen:

$$x_1 \oplus x_2 \equiv (\neg x_1 \wedge x_2) \vee (x_1 \wedge \neg x_2)$$

Das folgende neuronale Netz beschreibt genau diese äquivalente Formel:



Die erste Schicht bilden zwei Neuronen, die die UND-Funktion realisieren und jeweils einen der beiden Inputs invertieren. In der zweiten Schicht hingegen haben wir nur ein Neuron, das die beiden Ausgänge der Neuronen der ersten Schicht durch eine ODER-Funktion zum Ausgabewert 'a' verarbeitet:

## Aufgabe 2

Da die  $K$  versteckten Neuronen lineare Funktionen haben, kann deren Funktionalität durch eine Funktion am Ausgabeneuron dargestellt werden:

$$a = \beta_0 + \sum_{i=1}^K \beta_i \cdot (-\alpha_{i0} + \sum_{j=1}^N \alpha_{ij} e_j)$$

Da dies eine Komposition von linearen Funktionen ist, ist  $a$  eine lineare Funktion.

## Aufgabe 3

a.) Die Ausgabe ist abhängig von  $p$ :

$$a(x) = \begin{cases} 1 & , p > 0,5 \\ -1 & , p < 0,5 \\ x \in [-1; 1] & , p = 0,5 \end{cases}$$

Für  $p = 0,5$  ist der Wert von  $E_1$  immer gleich, daher kann die Funktion dann auf einen beliebigen Wert aus dem Intervall abbilden. Dieser muss natürlich fest gewählt werden.

b.) Die gesuchte Funktion ist  $E_2$ . Denn wenn wir ihr Minimum berechnen, indem wir die erste Ableitung gleich Null setzen, sehen wir, dass wir genau die beschriebene Formel für  $a(x)$  erhalten:

$$\begin{aligned} E'_2 &= \left( \sum_{i=1}^N |y_i - a(x_i)|^2 \right)' = (p(1 - a(x))^2 + (1 - p)(1 + a(x))^2)' = -4p + 2a(x) + 2 \\ -4p + 2a(x) + 2 &= 0 \Leftrightarrow a(x) = 2p - 1 \end{aligned}$$

c.)  $p = 0,9$  und  $N = 100$  bedeutet wir haben 90 positive und 100 negative Datenpunkte. Für  $E_2$  wissen wir, dass das Minimum bei  $a(x) = 2p - 1$  liegt.  
Also haben wir für unsere Werte ein Minimum bei:  
 $2 \cdot 0,9 - 1 = 0,8$ !

Für  $E_3$  bilden wir die Ableitung und setzen gleich Null:

$$\begin{aligned} E'_3 &= \left( \sum_{i=1}^N |y_i - a(x_i)|^3 \right)' = (90(1 - a(x))^3 + 10(1 + a(x))^3)' = -240a(x)^2 + 600a(x) - 240 \\ -240a(x)^2 + 600a(x) - 240 &= 0 \Leftrightarrow a(x)_1 = 0,5 \text{ oder } a(x)_2 = 2 \end{aligned}$$

Da für  $a(x)$  nur Werte aus dem Intervall  $[-1; 1]$  in Frage kommen, liegt unser Minimum der kubischen Fehlerfunktion mit den gegebenen Parametern bei 0,5.