Written Assignment 3

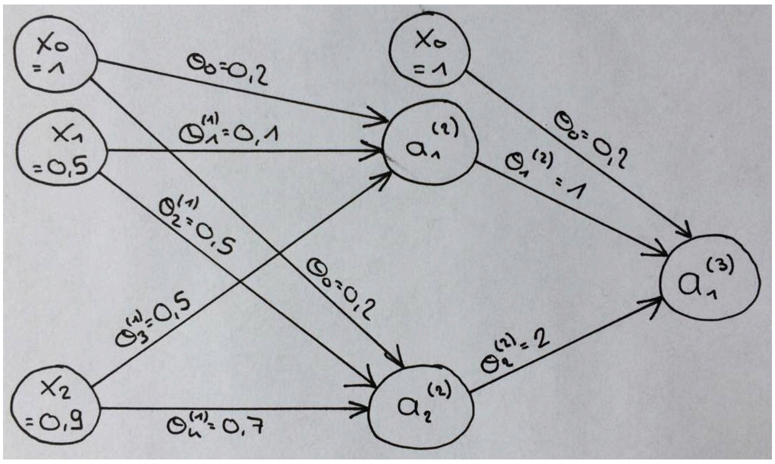
Machine Learning

Leah Dickhoff

S11155515

due November 4th, 2016

3. Neural Network figure:



1. Let $a_j^{(i)}$ denote the activation of node *j* in layer *i*.

   $a_1^{(2)} = g(\theta_0^{(1)}x_0 + \theta_1^{(1)}x_1 + \theta_3^{(1)}x_2)$ || where: Sigmoid function $g(z) = \frac{1}{1+e^{-z}}$

   $= g(0,2 \bullet 1 + 0,1 \bullet 0,5 + 0,5 \bullet 0,9)$

   $= g(0,7)$

   $= \frac{1}{1+e^{-0,7}}$

   $\approx 0,6682$

   $a_2^{(2)} = g(\theta_0^{(1)}x_0 + \theta_2^{(1)}x_1 + \theta_4^{(1)}x_2)$

   $= g(0,2 \bullet 1 + 0,5 \bullet 0,5 + 0,7 \bullet 0,9)$

   $= g(1,08)$

   $= \frac{1}{1+e^{-1,08}}$

   $\approx 0,7466$

   $a_1^{(3)} = g(\theta_0^{(1)}x_0 + \theta_1^{(2)}a_1^{(2)} + \theta_2^{(2)}a_2^{(2)})$

   $= g(0,2 \bullet 1 + 1 \bullet 0,6682 + 2 \bullet 0,7466)$

   $= g(2,3814)$

   $= \frac{1}{1+e^{-2,3814}}$

   $\approx 0,9154$

   ≈ 1

   Choice of theta values:

   Question 3.2. states that the correct output should be 1, so I chose the theta values such that $\frac{1}{1+e^{-z}}$ is closest to 1, i.e. such that z is as large as possible. For example, multiplying the largest x value, x_2, with the largest theta value, $\theta_4^{(1)}$=0,7, gives a larger 'z' and thus an output closer to 1.

2. Let $\delta_j^{(i)}$ denote the error of node *j* in layer *i.*
   Then for the final node: $\delta_1^{(3)} = a_1^{(3)} - y_1$ || from $\delta_j^{(i)} = a_j^{(i)} - y_j$
   $= 0{,}9154 - 1$
   $= 0{,}0846$
   For the previous nodes, we use: $\delta^{(i)} = \theta^{(i)}\, \delta^{(i+1)} \bullet g'(z^{(i)})$
   So for the 2nd layer:

$$\begin{aligned}\delta_1^{(2)} &= \theta_1^{(2)}\, \delta_1^{(3)} \bullet g'(z_1^{(2)})\\ &= \theta_1^{(2)}\, \delta_1^{(3)} \bullet [a_1^{(2)} \bullet (1-a_1^{(2)})]\\ &= 1 \bullet \delta_1^{(3)} \bullet 0{,}6682 \bullet (1-0{,}6682)\\ &= 0{,}0846 \bullet 0{,}6682 \bullet 0{,}3318\\ &\approx 0{,}01876\end{aligned}$$

$$\begin{aligned}\delta_2^{(2)} &= \theta_2^{(2)}\, \delta_2^{(3)} \bullet g'(z_2^{(2)})\\ &= \theta_2^{(2)}\, \delta_2^{(3)} \bullet [a_2^{(2)} \bullet (1-a_2^{(2)})]\\ &= 2 \bullet \delta_1^{(3)} \bullet 0{,}7466 \bullet (1-0{,}7466)\\ &= 2 \bullet 0{,}0846 \bullet 0{,}7466 \bullet 0{,}2534\\ &\approx 0{,}03201\end{aligned}$$

   With updating the weights $\theta_1^{(2)} \to \theta_1^{(2)} + \delta_1^{(2)}$, we have:

$$\begin{aligned}a_1^{(3)} &= g[\theta_0^{(1)}x_0 + (\theta_1^{(2)}+\delta_1^{(2)})\, a_1^{(2)} + (\theta_2^{(2)}+\delta_2^{(2)})\, a_2^{(2)}]\\ &= g(0{,}2\bullet 1 + (1+0{,}01876)\bullet 0{,}6682 + (2+0{,}03201)\bullet 0{,}7466)\\ &= g(2{,}3978)\\ &= \frac{1}{1+e^{-2{,}3978}}\\ &\approx 0{,}9167 \quad \text{(which is a little closer to one than before)}\end{aligned}$$

4.

1. For the perceptron, we have:

$$p(x_1, \ldots, x_n) = \begin{cases} 1 \text{ if } w_0 + w_1x_1 + \cdots + w_nx_n > 0 \\ -1 \text{ if } w_0 + w_1x_1 + \cdots + w_nx_n \leq 0 \end{cases}$$

   Since the line in figure a) is 2-dimensional, we can use $w_0+w_1x_1+w_2x_2$:

$$w_0+w_1x_1+w_2x_2 = 0$$
$$\Leftrightarrow x_2 = \frac{-w_1}{w_2}x_1 - \frac{w_0}{w_2} \qquad (1)$$

   Two points that lie on this line are A(-1,0) and B(0,2), so the slope of the line is given by:

$$\frac{\Delta y}{\Delta x} = \frac{0-2}{-1-0} = \frac{-2}{-1} = 2.$$

   Thus

$$\frac{-w_1}{w_2} = 2 \Leftrightarrow w_1 = -2w_2 \qquad (2)$$

   Now, the coordinates of A and (2) are plugged into (1):

$$0 = \frac{2w_2}{w_2}(-1) - \frac{w_0}{w_2}$$
$$\Leftrightarrow \frac{w_0}{w_2} = -2$$
$$\Leftrightarrow w_0 = -2w_2$$

Hence we have the system of equations: $\begin{cases} w_0 = -2w_2 \\ w_1 = -2w_2 \\ w_2 = w_2 \end{cases}$

So one possibility (setting w_2=1) for the values of the weights is:

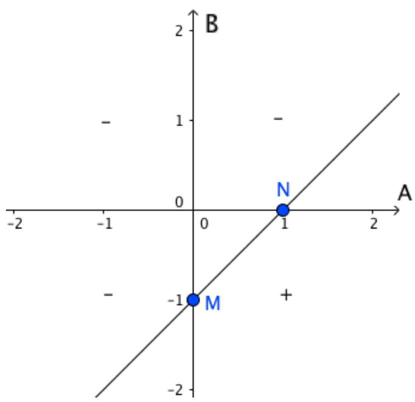
$$\begin{cases} w_0 = -2 \\ w_1 = -2 \\ w_2 = 1 \end{cases}$$

2.

a) All possible perceptron p(x) outputs:

A	B	p(A AND (NOT B))
1	1	-1
1	-1	1
-1	1	-1
-1	-1	-1

The graph of the perceptron is:



where the line of eq. B = $\frac{\Delta y}{\Delta x}$ A + k goes through N(1,0) and M(0,-1).

The slope of the line is $\frac{\Delta y}{\Delta x} = \frac{0-(-1)}{1-0}$ = 1.

So (filling in x_N, y_N) the line intersects the B-axis at 0 = 1•1+k ⇔ k = -1

Hence, the equation of the line is:

B = A – 1

⇔ –1 + A – B = 0

Comparing this to 3.2., a possibility for the values of the weights would be

$$\begin{cases} w_0 = -1 \\ w_1 = 1 \\ w_2 = -1 \end{cases}$$

b) All possible perceptron p(x) outputs:

A	B	p(A XOR B)
1	1	-1
1	-1	1
-1	1	-1
-1	-1	-1

A XOR B is the same thing as [A∩(NOT B)] ∪ [(NOT A)∩B].

Let's define one perceptron for each: P_1 gives A∩(NOT B)

P_2 gives (NOT A)∩B

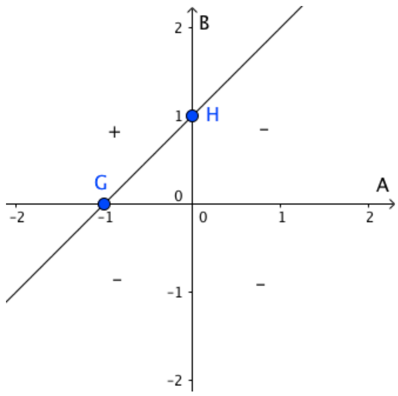
P_3 gives A XOR B

P_1 is described in exercise 4.2.a).

P_2 has the following possible outputs:

A	B	p((NOT A) AND B)
1	1	-1
1	-1	-1
-1	1	1
-1	-1	-1

with graph:

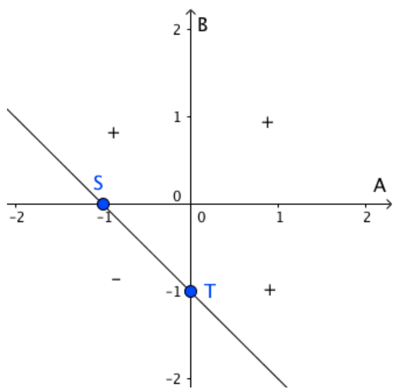


and with weights $\begin{cases} w_0 = 1 \\ w_1 = 1 \\ w_2 = -1 \end{cases}$ (found in the same way as in 4.2.a))

P_3 has the following possible outputs:

A	B	p($P_1 \cup P_2$)
1	1	1
1	-1	1
-1	1	1
-1	-1	-1

with graph:



and with weights $\begin{cases} w_0 = 1 \\ w_1 = 1 \\ w_2 = 1 \end{cases}$ (found in the same way as in 4.2.a))

Figure of the network in the case A = B = 1: