Machine Learning
Written Assignment 3

F. Raaijmakers 10886869

4 November 2016

Question 3

Part 1

The weights vectors are accordingly:

$$\theta^{(1)} = \begin{pmatrix} 0.5 \\ 0.1 \\ 0.5 \\ 0.1 \end{pmatrix} = \begin{pmatrix} \theta^{(1)}_{11} \\ \theta^{(1)}_{21} \\ \theta^{(1)}_{12} \\ \theta^{(1)}_{22} \end{pmatrix}$$

$$\theta^{(2)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \theta^{(2)}_{11} \\ \theta^{(2)}_{12} \end{pmatrix}$$

$$\theta^{(1)}_{10} = \theta^{(1)}_{20} = \theta^{(2)}_{10} = 0.2$$

The input vector is the following:

$$\mathrm{x} = \begin{pmatrix} x_0 = a_0 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.5 \\ 0.9 \end{pmatrix}$$

Thus we use the data above to calculate the activations:

$$\begin{aligned} a^{(2)}_1 &= g(\theta^{(1)}_{10} * x_0 + \theta^{(1)}_{11} * x_1 + \theta^{(1)}_{12} * x_2) \\ &= g(0.2 * 1 + 0.5 * 0.5 + 0.5 * 0.9) \\ &= \frac{1}{1 + e^{-0.9}} \approx 0.711 \end{aligned}$$

$$
\begin{aligned}
a_2^{(2)} &= g(\theta_{20}^{(1)} * x_0 + \theta_{21}^{(1)} * x_1 + \theta_{22}^{(1)} * x_2) \\
&= g(0.2 * 1 + 0.1 * 0.5 + 0.1 * 0.9) \\
&= \frac{1}{1 + e^{-0.9}} \approx 0.584
\end{aligned}
$$

$$
\begin{aligned}
a_1^{(3)} &= g(\theta_{10}^{(2)} * a_0^{(2)} + \theta_{11}^{(2)} * a_1^{(2)} + \theta_{12}^{(2)} * a_2^{(2)}) \\
&= g(0.2 * 1 + 1 * 0.711 + 2 * 0.584) \\
&= \frac{1}{1 + e^{-2.079}} \approx 0.888
\end{aligned}
$$

Part 2

$$
\begin{aligned}
\delta_1^{(3)} &= a_1^{(3)} - y_1 \\
&= 0.888 - 1 \\
&\approx 0.112
\end{aligned}
$$

$$
\begin{aligned}
\delta_1^{(2)} &= \theta_{11}^{(2)} * \delta_1^{(3)} * g'(z_1^{(2)}) \\
&= 1 * -0.112 * a_1^{(2)} * (1 - a_1^{(2)}) \\
&= 0.112 * 0.711 * 0.289 \\
&\approx 0.0230
\end{aligned}
$$

$$
\begin{aligned}
\delta_2^{(2)} &= \theta_{21}^{(2)} * \delta_1^{(3)} * g'(z_2^{(2)}) \\
&= 1 * -0.112 * a_2^{(2)} * (1 - a_2^{(2)}) \\
&= 0.112 * 0.711 * 0.243 \\
&\approx 0.0194
\end{aligned}
$$

Hence the updated weights are:

$$
\begin{aligned}
\theta_{11}^{(2)} &= \theta_{11}^{(2)} + \delta_1^{(2)} \\
&= 1 + 0.0230 \\
&= 1.0230
\end{aligned}
$$

$$
\begin{aligned}
\theta_{12}^{(2)} &= \theta_{12}^{(2)} + \delta_2^{(2)} \\
&= 2 + 0.0194 \\
&= 2.0194
\end{aligned}
$$

Question 4

Part 1

According to the slides the formula for the perceptron is:

$$o(x_1, \ldots, x_n = \begin{cases} 1, & if w_0 + w_1x_1 + \ldots + w_nx_n > 0 \\ -1, & \text{otherwise.} \end{cases}$$

Since $n = 2$, the equation of the decision line is:

$$w_0 + w_1x_1 + w_2x_2 = 0$$

Given the following coordinates:
a $= (\mathrm{a}_{x_1}, a_{x_2}) = (-1, 0)$
$b = (b_{x_1}, b_{x_2}) = (0, 2)$

We equate the slopes of both coordinates to find the equation of the line:

$$\begin{aligned} \frac{x_1 - a_{x_1}}{b_{x_1} - a_{x_1}} &= \frac{x_2 - a_{x_2}}{b_{x_2} - a_{x_2}} \\ \frac{x_1 - (-1)}{0 - (-1)} &= \frac{x_2 - 0}{2 - 0} \\ x_1 + 1 &= \frac{x_2}{2} \\ -2 - 2x_1 + x_2 &= 0 \end{aligned}$$

Thus comparing the above equation with the equation of the decision line, we can make out the values for w_0,w_1,w_2.
$\mathrm{w}_0 = -2$
$w_1 = -2$
$w_2 = 1$

Part 2

A
According to the truth table for 'and not' we find out the value for O given our two input system (A,B). As for the results of the output function, we can hypothesize a correct equation for the decision line for the A-B input. Computing the slopes as done in part 1, we yield an equation for the decision line and conclude the values for w_0,w_1,w_2. This output can be visualized that the first

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

$(A, B > 0)$, second $(A < 0, B > 0)$ and third $(A, B < 0)$ quadrant are negative (-1) and the fourth quadrant $(A > 0, B < 0)$ is positive (1). Hence the decision line will separate these regions. One of the possibilities is that the decision line will intersect with the A axis at 1 and the B axis at -1. Thus if we plug in these coordinates into the slope equation we yield:

$\mathrm{x} = (\mathrm{x}_A, x_B) = (1, 0)$
$y = (y_A, y_B) = (0, -1)$

$$\frac{A - x_A}{y_A - x_A} = \frac{A - x_B}{y_B - x_B}$$
$$\frac{A - 0}{1 - 0} = \frac{B - (-1)}{0 - (-1)}$$
$$A = B + 1$$
$$0 = -1 + A - B$$

Thus the weights are:
$\mathrm{w}_0 = -1$
$w_1 = 1$
$w_2 = -1$

B
In order to design a two-layer network of perceptrons that implements A XOR B, we should dissect A XOR B in two parts; namely (A AND NOT B) OR (NOT A and B). We observe that this is a combination of the result for the previous question and a variation of the previous result. Therefore we can interpret this as a two-layer network where the middle layer nodes are perception 1 (P1) and 2 (P2) and the inputs are A and B. The truth-table for perceptron P1 is as illustrated in the previous question, and the truth-table for P2 is the following (see-below):

The last layer is the concluding perceptron (P3), the truth table is the following (see below):

From the result of the concluding perceptron (P3) we see that only quad-

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

A	B	O(P1 or P2) "(A AND NOT B) OR (NOT A and B)"
-1	-1	-1
-1	1	1
1	-1	-1
1	1	-1

rant 3 ($A, B < 0$) is negative. Thus the decision line should separate quadrant 3 from the other quadrants. A possible line is intersecting the P1 axis (since the input of P3 is P1 and P2, the coordinate axis is P1 and P2) at -1 and the P2 axis at -1 as well.
x $= (\mathrm{x}_{P1}, x_{P2}) = (-1, 0)$
$y = (y_{P1}, y_{P2}) = (0, -1)$

Now we are performing the same procedure of equating the slopes to find the equationg for this two-layer perceptron:

$$\begin{aligned}
\frac{P1 - x_{P1}}{y_{P1} - x_{P1}} &= \frac{P2 - x_{P2}}{y_{P2} - x_{P2}} \\
\frac{P1 - (-1)}{0 - (-1)} &= \frac{x_2 - 0}{(-1) - 0} \\
P1 + 1 &= \frac{P2}{-1} \\
0 &= -P1 - P2 - 1
\end{aligned}$$

Thus the weights are:
$\mathrm{w}_0 = -1$
$w_1 = -1$
$w_2 = -1$