Assignment 2 Machine Learning COS4852

Adriaan Louw (53031377)

July 15, 2018

Contents

1 Question 1

1.1 Question 1(a)

Firstly we calculate the line

$$x_2 = mx_1 + c \tag{1}$$

for the intersect points (2,0) and (0,6).
Calculating slope m,

$$
\begin{aligned}
m &= \frac{6-0}{0-2} \\
&= -3
\end{aligned}
\tag{2}
$$

x_2 intercept c is 6.
This makes equation 1

$$x_2 = -3x_1 + 6 \tag{3}$$

Nils J Nilsson (1998) gives the equation for the hyperplane as

$$\sum_{i=1}^{n} x_i \omega_i \geq \theta \tag{4}$$

which in this case gives the equation for the hyperplane to be

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 = 0 \tag{5}$$

We need to get equation 5 in the form of equation 1

$$\omega_1 x_1 + \omega_2 x_2 + + \omega_3 = 0$$
$$\omega_2 x_2 = -\omega_1 x_1 - \omega_3 \tag{6}$$
$$x_2 = -\frac{\omega_1 x_1}{\omega_2} - \frac{\omega_3}{\omega_2}$$

Comparing coefficients m and c from equation 3 to 6 we get

$$-\frac{\omega_1}{\omega_2} = -3$$
$$\omega_1 = 3\omega_2 \tag{7}$$

and

$$-\frac{\omega_3}{\omega_2} = 6$$
$$\omega_3 = -6\omega_2 \tag{8}$$

If we choose $\omega_3 = -2$ then $\omega_1 = 1$ and $\omega_2 = \frac{1}{3}$. This makes the hyperplane equation from equation 5

$$x_1 + \frac{x_2}{3} - 2 = 0 \tag{9}$$

Now we need to test this hyperplane. For positive instance (2,6)

$$x_1 + \frac{x_2}{3} - 2 =$$
$$2 + \frac{6}{2} - 2 = \tag{10}$$
$$2$$

Which is as expected.
And the negative instance (-1,2)

$$x_1 + \frac{x_2}{3} - 2 =$$
$$-1 + \frac{2}{3} - 2 = \tag{11}$$
$$-\frac{7}{3}$$

This is also as expected. The perceptron now classifies the the data correctly

1.2 Question 1(b)
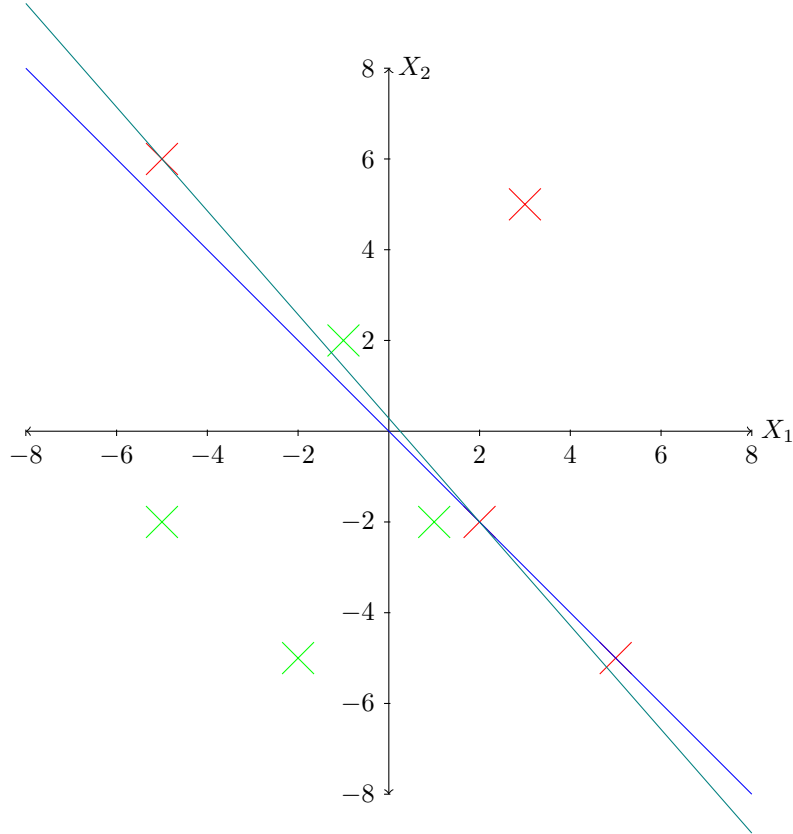
From the above image we can see that any that it is not possible to create a hyperplane that correctly classifies all negative instances and positive instances. The blue line is the line $x_2 = -x_1$ and the teal line is the line $x_2 = -\frac{7}{8}x_1 + \frac{2}{7}$. The any minimum plane that correctly classifies all the negative instances will classify the positive instance (-1,2) incorrectly as negative.

We can create a hyperplane from regression from all the points close to where the hyperplane should be. Using negative points (-5,6),(2,-2),(5,-5) and positive points (-1,2),(1,-2)

For the equation of the regressed line $x_2 = mx_1 + c$

$$m = r\frac{S(x_2)}{S(x_1)}$$

$$m = \frac{\sum((x_1 - \bar{x_1})(x_2 - \bar{x_2}))}{\sqrt{\sum(x_1 - \bar{x_1})^2 \sum(x_2 - \bar{x_2})^2}}\frac{\sqrt{\frac{\sum(x_2 - \bar{x_2})^2}{n-1}}}{\sqrt{\frac{\sum(x_1 - \bar{x_1})^2}{n-1}}} \tag{12}$$

where r is Pearsons Correlation Coefficient and S is standard deviation of axis x_2 or x_1.
Here follows the calculation

x_1	x_2	$x_1 - \bar{x_1}$	$x_2 - \bar{x_2}$	$(x_1 - \bar{x_1})(x_2 - \bar{x_2})$	$(x_1 - \bar{x_1})^2$	$(x_2 - \bar{x_2})^2$
-5	6	-5.4	6.2	-33.48	29.16	38.44
2	-2	1.6	-1.8	-2.88	2.56	3.24
5	-5	4.6	-4.8	-22.08	21.16	23.04
-1	2	-1.4	2.2	-3.08	1.96	4.84
1	-2	0.6	-1.8	-1.08	0.36	3.24

From the above table we have $\bar{x_1} = 0.4, \bar{x_2} = -0.2, \sum((x_1 - \bar{x_1})(x_2 - \bar{x_2})) = -62.6, (x_1 - \bar{x_1})^2 = 55.2, and (x_2 - \bar{x_2})^2 = 72.8$

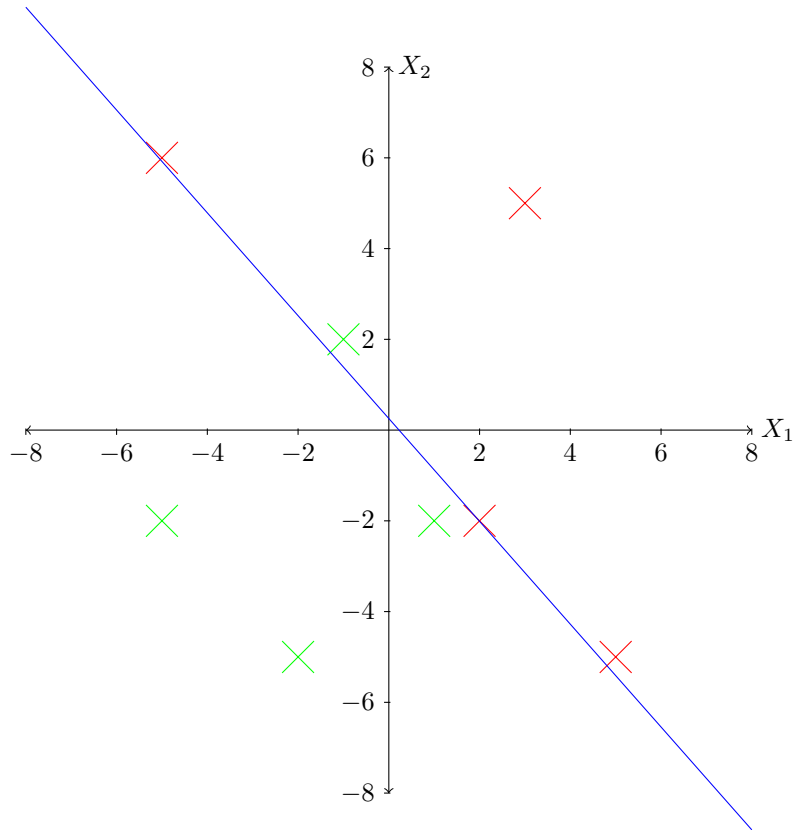Passing these into equation 12 we get m = -1.13

$$c = \bar{x_2} - m\bar{x_1}$$
$$= -0.2 - (-1.13)(0.4) \tag{13}$$
$$= 0.25$$

Which gives us equation

$$x_2 = -1.13x_1 + 0.25 \tag{14}$$

Visually it would be

This will minimise the error even though it would incorrectly classify (-1,2)

1.3   Question 1(c)

The equation for the hyperplane is:

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_0 = 0 \tag{15}$$

Passing in the point (14.0,0) gives us into equation 15

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_0 = 0$$
$$\omega_1(14) + \omega_2(0) + \omega_3(0) + \omega_0 = 0$$
$$14\omega_1 + \omega_0 = 0 \tag{16}$$
$$\omega_1 = -\frac{\omega_0}{14}$$

Passing the point (0,8,0) into equation 15

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_0 = 0$$
$$\omega_1(0) + \omega_2(8) + \omega_3(0) + \omega_0 = 0$$
$$8\omega_2 + \omega_0 = 0 \tag{17}$$
$$\omega_2 = -\frac{\omega_0}{8}$$

Passing the point (0,0,6) into equation 15

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_0 = 0$$
$$\omega_1(0) + \omega_2(0) + \omega_3(6) + \omega_0 = 0$$
$$6\omega_3 + \omega_0 = 0 \tag{18}$$
$$\omega_3 = -\frac{\omega_0}{6}$$

Now we need to solve the 3 simultaneous equations. Any value of ω_0 will solve the equation but we still need to check for polarity

If $\omega_0 = -1$

Then $\omega_1 = \frac{1}{14}, \omega_2 = \frac{1}{8}$ and $\omega_3 = \frac{1}{6}$.

Checking for polarity with positive instance (3,5,5)

$$\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + \omega_0 =$$
$$\frac{1}{14}(3) + \frac{1}{8}(5) + \frac{1}{6}(5) - 1 =$$
$$\frac{36 + 105 + 140 - 168}{168} = \tag{19}$$
$$\frac{113}{168}$$

Which is positive thus meaning the polarity is correct.

Thus $\omega_0 = -1; \omega_1 = \frac{1}{14}; \omega_2 = \frac{1}{8}; \omega_3 = \frac{1}{6}$.
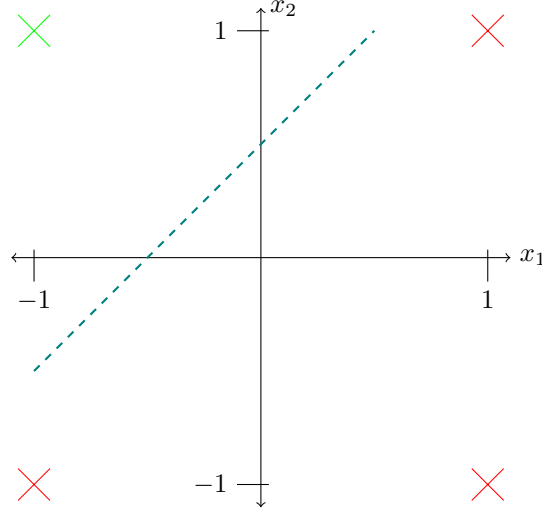
1.4 Question 1(d)

inputs

$$net = \sum_{i=0}^{3} \omega_1 x_i$$

5
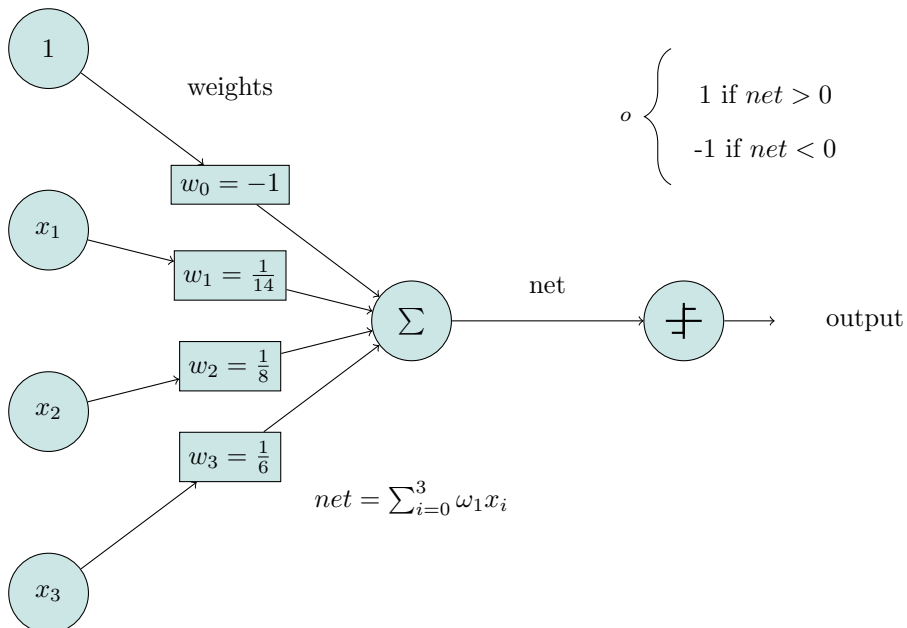
2 Question 2

2.1 Question 2(a)

The truth table for this function is

x_1	x_2	$\neg x_1$	$\neg x_1 \vee x_2$
-1	-1	1	-1
-1	1	1	1
1	-1	-1	-1
1	1	-1	-1

Visually this is:

The dashed line represents the function

$$x_2 = x_1 + 0.5 \tag{20}$$

This hyperplane will classify the boolean function correctly because it linearly separates all the positive instances from the negative ones.

The equation for the hyperplane with weights are given by equation 5 and we know from equation 6 what the relationship is from the weight equation to the line equation. Thus for case

$$m = -\frac{\omega_1}{\omega_2}$$
$$1 = -\frac{\omega_1}{\omega_2} \tag{21}$$
$$\omega_2 = -\omega_1$$

and

$$c = -\frac{\omega_0}{\omega_2}$$
$$0.5 = -\frac{\omega_0}{\omega_2} \tag{22}$$
$$\omega_0 = -\frac{\omega_2}{2}$$

Now we have the relationship between the weights. If we try $\omega_0 = 1$ we get $\omega_1 = 2$ and $\omega_2 = -2$ Which makes our weight equation

$$2x_1 - 2x_2 + 1 \tag{23}$$

6

Testing for polarity with point (-1,-1) we get

$$2(-1) - 2(-1) + 1 = 1 \tag{24}$$

which indicates a positive result but we expect a negative result. Thus our polarity is wrong
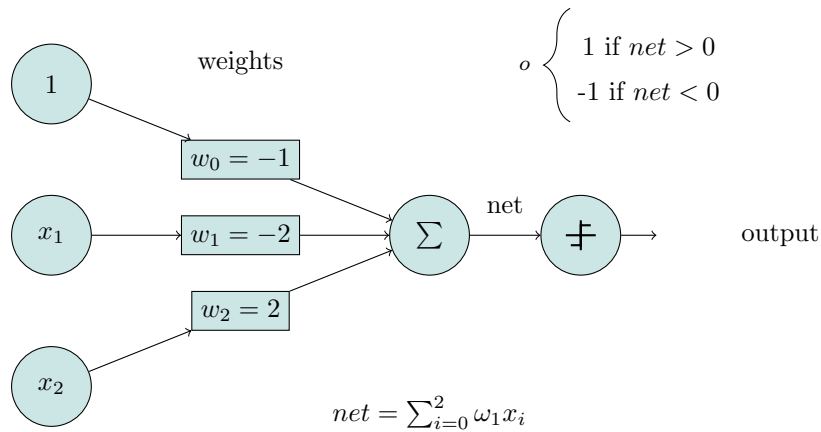We try $\omega_0 = -1$ which gives $\omega_1 = -2$ and $\omega_2 = 2$
Which makes the weight equation

$$-2x_1 + 2x_2 - 1 \tag{25}$$

Trying all our data we can sum it up in the following table

x_1	x_2	Result
-1	-1	-1
-1	1	4
1	-1	-5
1	1	-1

This shows that weights $\omega_0 = -1$; $\omega_1 = -2$ and $\omega_2 = 2$ correctly classify the data

$$net = \sum_{i=0}^{2} \omega_1 x_i$$

2.2   Question 2(b)

x_1	x_2	$x_1 \oplus x_2$	$f_2 = \neg(x_1 \oplus x_2)$
-1	-1	-1	1
-1	1	1	-1
1	-1	1	-1
1	1	-1	1

Visually this is:

7
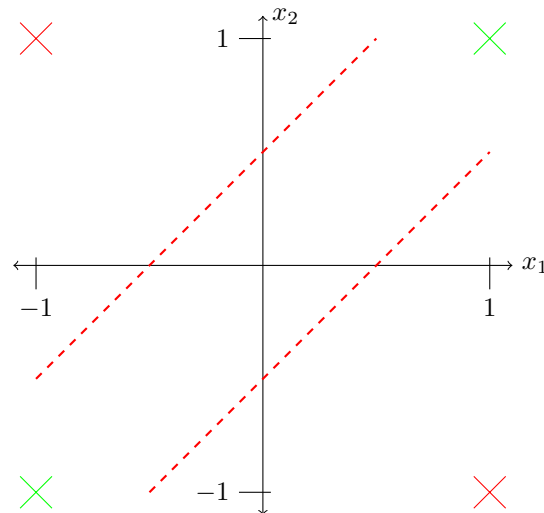
We can see from the previous diagram that f_2 cannot be linearly separated by 1 hyperplane. In other words a single perceptron cannot classify this function. We need to try to decompose f_2 into multiple linearly separable functions that can be modelled by multiple perceptrons.

$$
\begin{aligned}
f_2 &= \neg(x_1 \bigoplus x_2) \\
&= \neg((x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2)) \\
&= \neg(x_1 \wedge \neg x_2) \wedge \neg(\neg x_1 \wedge x_2) \\
&= (\neg x_1 \vee x_2) \wedge (x_1 \vee \neg x_2)
\end{aligned}
\tag{26}
$$

We can thus create new function which is equivalent to f_2

$$
g(h_1, h_2) = h_1 \wedge h_2 \tag{27}
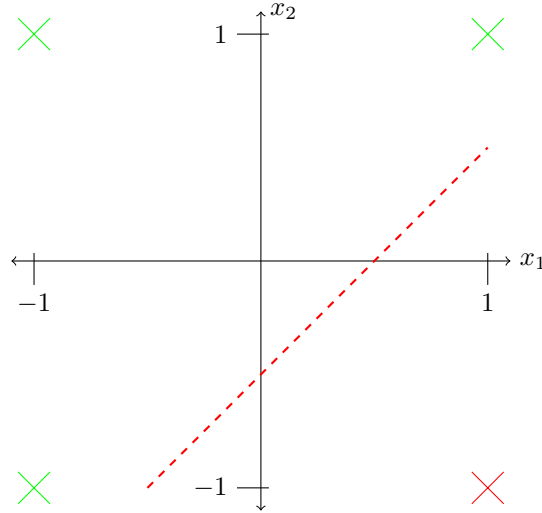$$

where

$$
h_1 = \neg x_1 \vee x_2 \tag{28}
$$

and

$$
h_2 = x_1 \vee \neg x_2 \tag{29}
$$

The truth table will thus become

x_1	x_2	$x_1 \bigoplus x_2$	$f_2 = \neg(x_1 \bigoplus x_2)$	h_1	h_2	$g(h_1, h_2)$
-1	-1	-1	1	1	1	1
-1	1	1	-1	1	-1	-1
1	-1	1	-1	-1	1	-1
1	1	-1	1	1	1	1

Drawing h_1 gives

Using equation 5 and knowing this hyperplane goes through points (-0.5,-1) and (1.0,0.5). We have

$$
\begin{aligned}
\omega_1(-0.5) + \omega_2(-1) + \omega_0 &= 0 \\
\frac{-\omega_1}{2} - \omega_2 + \omega_0 &= 0
\end{aligned}
\tag{30}
$$

for the point (-0.5,-1) and

$$
\begin{aligned}
\omega_1(1) + \omega_2(0.5) + \omega_0 &= 0 \\
\omega_1 + \frac{\omega_2}{2} + \omega_0 &= 0
\end{aligned}
\tag{31}
$$

for point (1.0,0.5)
Equating equations 30 and 31 we get

$$\frac{-\omega_1}{2} - \omega_2 + \omega_0 = \omega_1 + \frac{\omega_2}{2} + \omega_0$$
$$\frac{3\omega_1}{2} = \frac{3\omega_2}{2} \tag{32}$$
$$\omega_2 = -\omega_1$$

Passing equation 32 into 31 we get

$$\omega_1 + \frac{-\omega_1}{2} + \omega_0 = 0$$
$$\omega_0 = -\frac{\omega_1}{2} \tag{33}$$

If we take $\omega_0 = 1$ then $\omega_1 = -2$ and $\omega_2 = 2$.
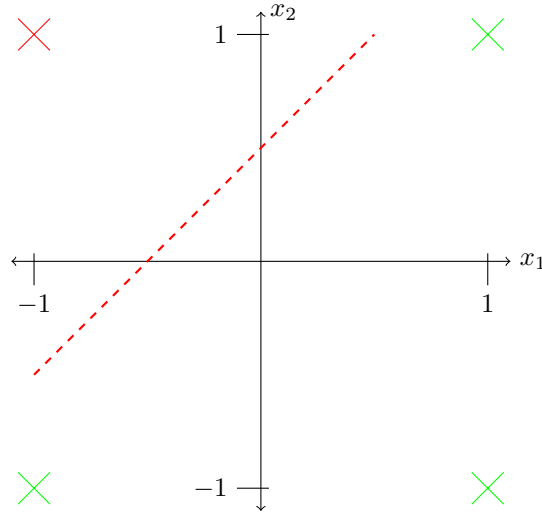Testing for polarity we use the point (-1,-1) where we expect a positive answer
For point (-1,-1)

$$-2x_1 + 2x_2 + 1 =$$
$$-1(-1) + 2(-1) + 1 = 1 \tag{34}$$

Which is greater than 0 as expected. Thus the weights for h_1 is $\omega_0 = 1; \omega_1 = -2; \omega_2 = 2$
Now drawing h_2

This is a similar perceptron from as in Question 2a. It uses the same hyperplane but the polarity is reversed. Thus from question 2a $\omega_2 = -\omega_1; \omega_0 = -\frac{\omega_2}{2}$.
Testing polarity with $\omega_0 = 1$ then $\omega_1 = 2; \omega_2 = -2$. This gives with positive instance (-1,-1)

$$2x_1 - 2x_2 + 1 =$$
$$2(-1) - 2(-1) + 1 = 1 \tag{35}$$

Which indicates the preceptron calssifies the point (-1,-1) correctly as a positive instance
This means $\omega_0 = 1; \omega_1 = 2; \omega_2 = -2$ for h_2
Now we need to create a perceptron for function g. The domain of g as a function of h_1 and h_2 can be shown as follows
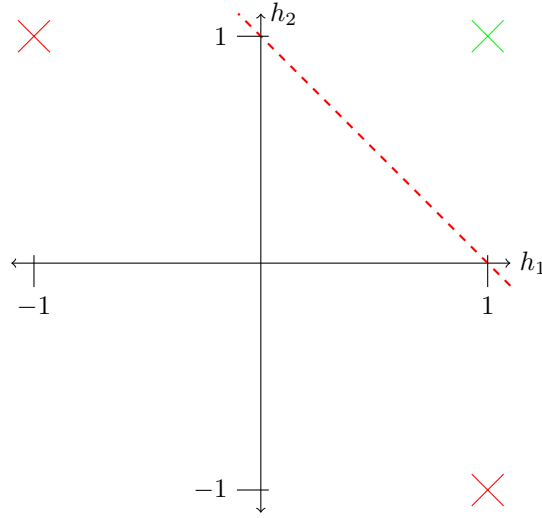
From the above diagram we can see that a hyperplane through the points (0,1) and (1,0) will classify this perceptron

From applying equation 5 to point $(h_1, h_2) = (1, 0)$

$$
\begin{aligned}
\omega_1 x_1 + \omega_2 x_2 + \omega_0 &= 0 \\
\omega_1(1) + \omega_2(0) + \omega_0 &= 0 \\
\omega_1 &= -\omega_0
\end{aligned}
\tag{36}
$$

and for point $(h_1, h_2) = (0, 1)$

$$
\begin{aligned}
\omega_1 x_1 + \omega_2 x_2 + \omega_0 &= 0 \\
\omega_1(0) + \omega_2(1) + \omega_0 &= 0 \\
\omega_2 &= -\omega_0
\end{aligned}
\tag{37}
$$

Now equation equations 36 and 37 we get $\omega_1 = \omega_2$

If we let $\omega_0 = -1$ then from substitution into equations 36 and 37 we get $\omega_1 = \omega_2 = 1$.

Testing for polarity we have positive instance (1,1)

$$
\begin{aligned}
\omega_1 x_1 + \omega_2 x_2 + \omega_0 &= \\
x_1 + x_2 - 1 &= \\
(1) + (1) - 1 &= 1
\end{aligned}
\tag{38}
$$

Which is positive as expected

Figure 1 shows the whole network where ω_{1i} denotes the weights for the perceptron modeling h_1, ω_{2i} denotes the weights for the perceptron of h_2 and ω_{3i} denotes the weights of the final perceptron.

2.3 Question 2(c)

Given function

$$
f_3(x_1, x_2) = (x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_2)
\tag{39}
$$

The truth table for the function is the following:

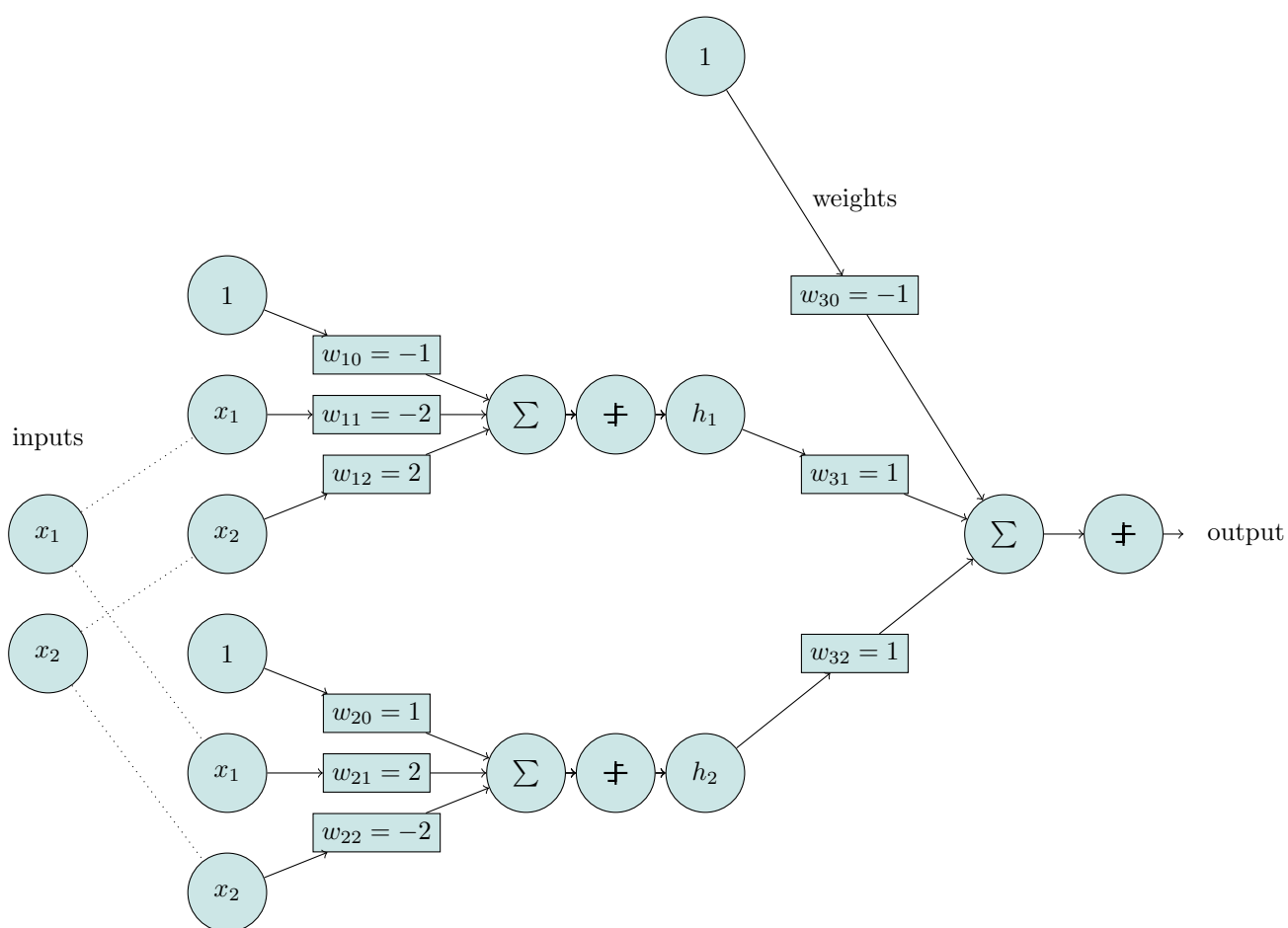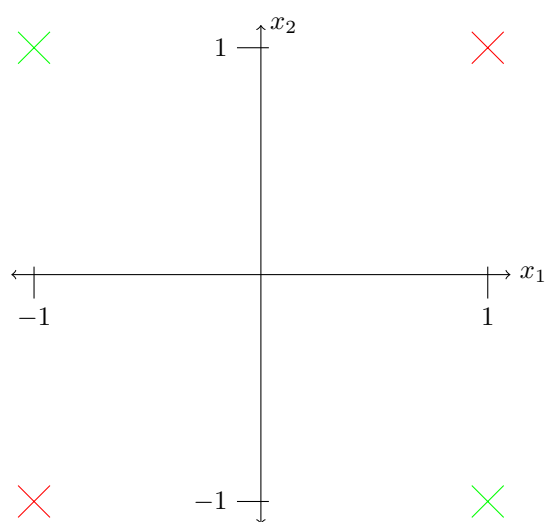| x_1	x_2	$\neg x_1$	$\neg x_2$	$x_1 \vee x_2$	$\neg x_1 \vee \neg x_2$	$(x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_2)$
-1	-1	1	1	-1	1	-1
-1	1	1	-1	1	1	1
1	-1	-1	1	1	1	1
1	1	-1	-1	1	-1	-1

Figure 1: Network for question 2b

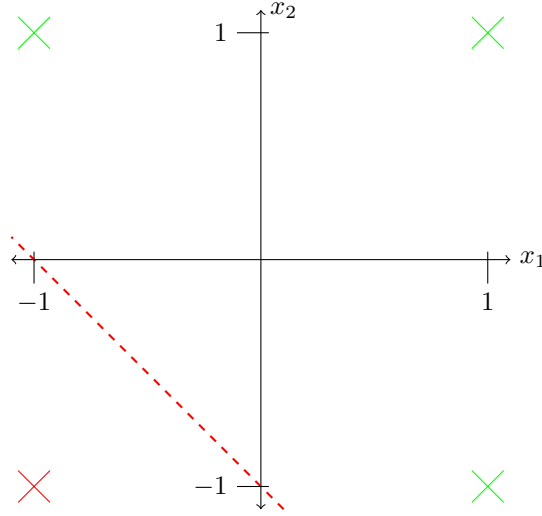Figure 2: Sketch of function $f_3(x_1, x_2)$

Figure 3: Sketch of function for $h_1(x_1, x_2)$ for Question 2c

Visually this can be seen in figure 2

No single hyperplane can classify all the data points. Multiple perceptrons will be needed. We will break the function up similar to Question 2b.

$$f_3(x_1, x_2) = g(h_1(x_1, x_2), h_2(x_1, x_2))$$
$$= h_1(x_1, x_2) \wedge h_2(x_1, x_2) \tag{40}$$

where

$$h_1(x_1, x_2) = x_1 \vee x_2 \tag{41}$$

and

$$h_2 x_1, x_2 = \neg x_1 \vee \neg x_2 \tag{42}$$

The truth table for $h_1(x_1, x_2)$ is:

x_1	x_2	$x_1 \vee x_2$
-1	-1	-1
-1	1	1
1	-1	1
1	1	1

Drawing the function $h_1(x_1, x_2)$ in Figure 3

We can see that h_1 will be classified by the hyperplane

$$x_2 = -x - 1 \tag{43}$$

Now we solve the weights equation (Eqn 5) with the points (0,-1) and (-1,0) on the hyperplane. Firstly with (0,-1)

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 = 0$$
$$\omega_1(0) + \omega_2(-1) + \omega_0 = 0 \tag{44}$$
$$\omega_2 = \omega_0$$

and for point (-1,0)

Figure 4: Sketch of function for $h_2(x_1, x_2)$ for Question 2c

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 = 0$$
$$\omega_1(-1) + \omega_2(0) + \omega_0 = 0 \tag{45}$$
$$\omega_1 = \omega_0$$

Thus $\omega_0 = \omega_1 = \omega_2$.

Since any value of ω_0 will solve the equation, we try $\omega_0 = 1$. Now we test for polarity with negative instance (-1,-1).

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 =$$
$$1(-1) + 1(-1) + 1 = \tag{46}$$
$$-1$$

The answer is negative thus correctly classifying the function $h_1(x_1, x_2)$

Now for the function $h_2(x, x_2)$.

The truth table for $h_2(x, x_2)$ is:

x_1	x_2	$\neg x_1 \vee \neg x_2$
-1	-1	1
-1	1	1
1	-1	1
1	1	-1

Drawing the function $h_1(x_1, x_2)$ in Figure 4

From Figure 4 we can see that function $h_2(x_1, x_2)$ will be classified by the hyperplane

$$x_2 = -x_1 + 1 \tag{47}$$

Now to solve the weights equation (Eqn 5) with the 2 points (0,1) and (1,0). For the point (0,1)

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 = 0$$
$$\omega_1(0) + \omega_2(1) + \omega_0 = 0 \tag{48}$$
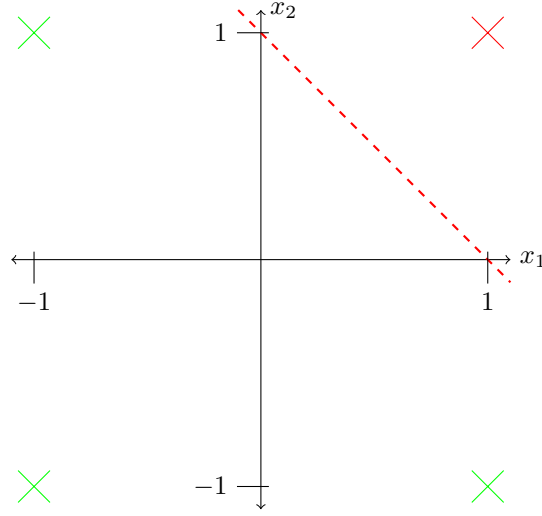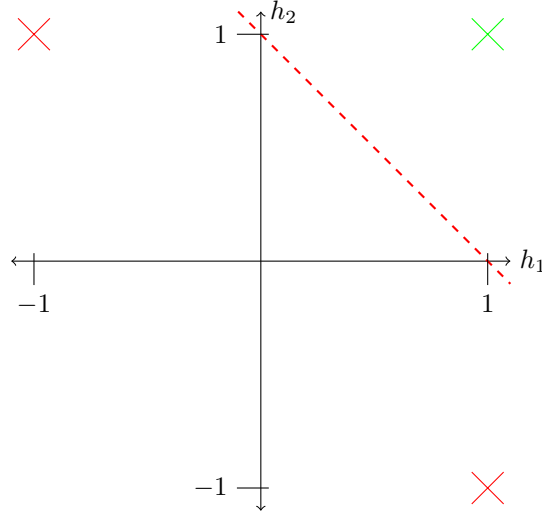$$\omega_0 = -\omega_2$$

For the point (1,0)

13

Figure 5: Sketch of function for $g(h_1, h_2)$ for Question 2c

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 = 0$$
$$\omega_1(1) + \omega_2(0) + \omega_0 = 0 \tag{49}$$
$$\omega_0 = -\omega_1$$

Now any value of ω_0 will solve this equation, but we need to check for polarity.
If we assume $\omega_0 = -1$ then $\omega_1 = \omega_2 = 1$. We check with positive instance (-1,-1).

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 =$$
$$1(-1) + 1(-1) + (-1) = \tag{50}$$
$$-3$$

Which gives a negative answer. Thus the polarity is swapped. Now trying $\omega_0 = 1$ and $\omega_1 = \omega_2 = -1$ with positive instance (-1,-1)

$$\omega_1 x_1 + \omega_2 x_2 \omega_0 =$$
$$-1(-1) + -1(-1) + (1) = \tag{51}$$
$$3$$

$\omega_0 = 1$ correctly classifies the function $h_2(x_1, x_2)$
Now we need to determine the values of g with respect to h_1 and h_2

x_1	x_2	$\neg x_1$	$\neg x_2$	$h_1 = x_1 \vee x_2$	$h_2 = \neg x_1 \vee \neg x_2$	$g = (x_1 \vee x_2) \wedge (\neg x_1 \vee \neg x_2)$
-1	-1	1	1	-1	1	-1
-1	1	1	-1	1	1	1
1	-1	-1	1	1	1	1
1	1	-1	-1	1	-1	-1

Now drawing the function $g_{(h_1, h_2)}$ in Figure 5.
Note that the function $g_{(h_1, h_2)}$ is not defined for the point $(h_1, h_2) = (-1, -1)$. This perceptron will be classified by the hyperplane

$$h_2 = -h_1 + 1 \tag{52}$$

Solving the weights equation (Eqn 5) for points $(h_1, h_2) = (0, 1)$ and $(h_1, h_2) = (1, 0)$.

14

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	b_1	b_2
0.15	0.2	0.25	0.3	0.4	0.45	0.5	0.55	0.35	0.6

Table 1: Start weights

pattern	x_1	x_2	d_1	d_2
p_1	0.1	0.1	0.1	0.9
p_2	0.1	0.9	0.9	0.1
p_3	0.9	0.1	0.9	0.1
p_4	0.9	0.9	0.1	0.9

Table 2: Input Patterns

For $(h_1, h_2) = (0, 1)$

$$
\begin{aligned}
\omega_1 h_1 + \omega_2 h_2 \omega_0 &= 0 \\
\omega_1(0) + \omega_2(1) + \omega_0 &= 0 \\
\omega_0 &= -\omega_2
\end{aligned}
\tag{53}
$$

For the point $(h_1, h_2) = (1, 0)$

$$
\begin{aligned}
\omega_1 h_1 + \omega_2 h_2 \omega_0 &= 0 \\
\omega_1(1) + \omega_2(0) + \omega_0 &= 0 \\
\omega_0 &= -\omega_1
\end{aligned}
\tag{54}
$$

Since any value of ω_0 will solve this equation we choose $\omega_0 = -1$. This gives $\omega_1 = \omega_2 = 1$. Testing polarity with with positive instance $(h_1, h_2) = (1, 1)$

$$
\begin{aligned}
\omega_1 h_1 + \omega_2 h_2 \omega_0 &= \\
1(1) + 1(1) + (-1) &= \\
1
\end{aligned}
\tag{55}
$$

Which is greater than 0 thus classifying the data correctly.

Thus we have weights $\omega_0 = -1$ and $\omega_1 = \omega_2 = 1$ for the perceptron classifying function $g(h_1, h_2)$.

A diagram for this perceptron can be found in Figure 6

3   Question 3

Rumelhart, Hinton, and Williams (1986) is the original article introducing the Backpropagation algorithm. The URL for this document is in the reference.

3.1   Question 3(a)

We start with the following weights (Table 1).

3.1.1   Pattern 1

Doing the forward pass. Starting with the values on the hidden units

$$
\begin{aligned}
net_{h1} &= \omega_1 x_1 + \omega_2 x_2 + b_1 \\
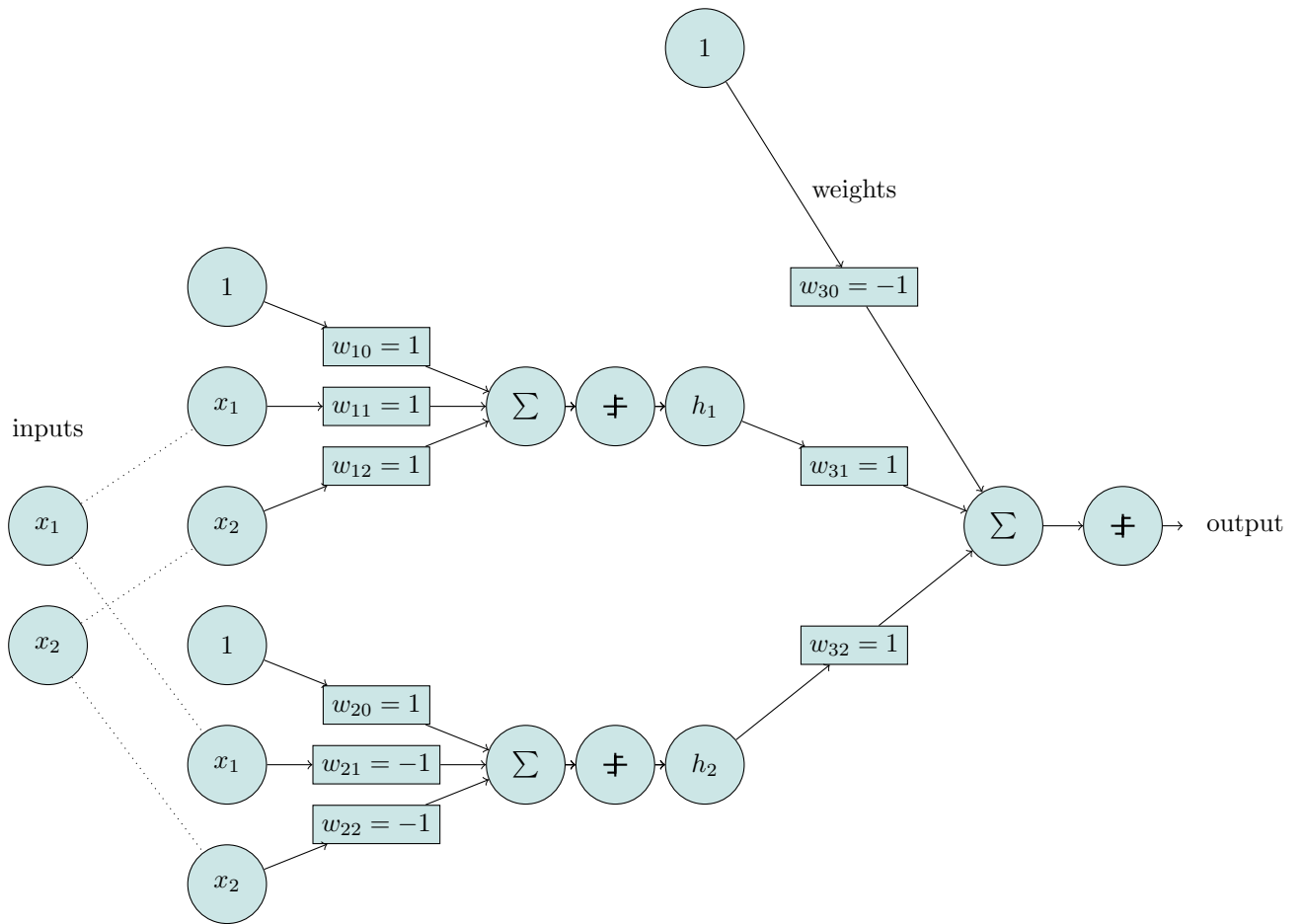&= 0.15(0.1) + 0.2(0.1) + 0.35 \\
&= 0.385
\end{aligned}
\tag{56}
$$

Figure 6: Network for question 2c

$$out_{h1} = \frac{1}{1 + e^{-net_{h1}}}$$
$$= \frac{1}{1 + e^{-0.385}} \quad (57)$$
$$= 0.595$$

$$net_{h2} = \omega_3 x_1 + \omega_4 x_2 + b_1$$
$$= 0.25(0.1) + 0.3(0.1) + 0.35 \quad (58)$$
$$= 0.405$$

$$out_{h2} = \frac{1}{1 + e^{-net_{h2}}}$$
$$= \frac{1}{1 + e^{-0.405}} \quad (59)$$
$$= 0.6$$

Continuing with the outputs of the network

$$net_{o1} = \omega_5 out_{h1} + \omega_6 out_{h2} + b_2$$
$$= 0.4(0.595) + 0.45(0.6) + 0.6 \quad (60)$$
$$= 1.108$$

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}}$$
$$= \frac{1}{1 + e^{-1.108}} \quad (61)$$
$$= 0.752$$

$$net_{02} = \omega_7 out_{h1} + \omega_8 out_{h2} + b_2$$
$$= 0.5(0.595) + 0.55(0.6) + 0.6 \quad (62)$$
$$= 1.227$$

$$out_{o1} = \frac{1}{1 + e^{-net_{h1}}}$$
$$= \frac{1}{1 + e^{-1.223}} \quad (63)$$
$$= 0.773$$

Now we need to calculate the error

$$E_{total} = E_{o1} + E_{02}$$
$$= \frac{1}{2}(0.1 - 0.752)^2 + \frac{1}{2}(0.9 - 0.773)^2 \quad (64)$$
$$= 0.220$$

$$\delta_{o1} = -(target_{o1} - out_{o1})(out_{o1})(1 - out_{o1})$$
$$= -(0.1 - 0.752)(0.752)(1 - 0.752) \quad (65)$$
$$= 0.122$$

$$\frac{\partial E_{total}}{\partial \omega_5} = \delta_{o1} out_{h1}$$
$$= 0.122(0.595)$$
$$= 0.073$$
$$(66)$$

We calculate the new value for ω_5

$$\omega_5^1 = \omega_5^0 - \eta \frac{\partial E_{total}}{\partial \omega_5}$$
$$= 0.4 - 0.5(0.073)$$
$$= 0.364$$
$$(67)$$

$$\frac{\partial E_{total}}{\partial \omega_6} = \delta_{o1} out_{h2}$$
$$= 0.122(0.6)$$
$$= 0.073$$
$$(68)$$

We calculate the new value for ω_6

$$\omega_6^1 = \omega_6^0 - \eta \frac{\partial E_{total}}{\partial \omega_6}$$
$$= 0.45 - 0.5(0.073)$$
$$= 0.414$$
$$(69)$$

$$\delta_{o2} = -(target_{o2} - out_{o2})(out_{o2})(1 - out_{o2})$$
$$= -(0.9 - 0.773)(0.773)(1 - 0.773)$$
$$= -0.022$$
$$(70)$$

$$\frac{\partial E_{total}}{\partial \omega_7} = \delta_{o2} out_{h1}$$
$$= -0.022(0.595)$$
$$= -0.013$$
$$(71)$$

We calculate the new value for ω_7

$$\omega_7^1 = \omega_7^0 - \eta \frac{\partial E_{total}}{\partial \omega_7}$$
$$= 0.5 - 0.5(-0.013)$$
$$= 0.507$$
$$(72)$$

$$\frac{\partial E_{total}}{\partial \omega_8} = \delta_{o2} out_{h2}$$
$$= -0.022(0.6)$$
$$= -0.013$$
$$(73)$$

We calculate the new value for ω_8

$$\omega_8^1 = \omega_8^0 - \eta \frac{\partial E_{total}}{\partial \omega_8}$$
$$= 0.55 - 0.5(-0.013)$$
$$= 0.557$$
$$(74)$$

Now we need to pass the error back to the hidden units

$$
\begin{aligned}
\delta_{h1} &= (\delta_{o1}\omega_5^0 + \delta_{o2}\omega_7^0)(out_{h1})(1 - out_{h1}) \\
&= (0.122(0.4) - 0.022(0.5))(0.595)(1 - 0.595) \\
&= 0.009
\end{aligned}
\tag{75}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_1} &= \delta_{h1}x_1 \\
&= 0.009(0.1) \\
&= 0.001
\end{aligned}
\tag{76}
$$

$$
\begin{aligned}
\omega_1^1 &= \omega_1^0 - \eta\frac{\partial E_{total}}{\partial \omega_1} \\
&= 0.15 - 0.5(0.001) \\
&= 0.149
\end{aligned}
\tag{77}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_2} &= \delta_{h1}x_2 \\
&= 0.009(0.1) \\
&= 0.001
\end{aligned}
\tag{78}
$$

$$
\begin{aligned}
\omega_2^1 &= \omega_2^0 - \eta\frac{\partial E_{total}}{\partial \omega_2} \\
&= 0.2 - 0.5(0.001) \\
&= 0.199
\end{aligned}
\tag{79}
$$

$$
\begin{aligned}
\delta_{h2} &= (\delta_{o1}\omega_6^0 + \delta_{o2}\omega_8^0)(out_{h2})(1 - out_{h2}) \\
&= (0.122(0.45) - 0.022(0.55))(0.6)(1 - 0.6) \\
&= 0.010
\end{aligned}
\tag{80}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_3} &= \delta_{h2}x_1 \\
&= 0.010(0.1) \\
&= 0.001
\end{aligned}
\tag{81}
$$

$$
\begin{aligned}
\omega_3^1 &= \omega_3^0 - \eta\frac{\partial E_{total}}{\partial \omega_3} \\
&= 0.25 - 0.5(0.001) \\
&= 0.249
\end{aligned}
\tag{82}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_4} &= \delta_{h2}x_2 \\
&= 0.010(0.1) \\
&= 0.001
\end{aligned}
\tag{83}
$$

$$
\begin{aligned}
\omega_4^1 &= \omega_4^0 - \eta\frac{\partial E_{total}}{\partial \omega_4} \\
&= 0.3 - 0.5(0.001) \\
&= 0.299
\end{aligned}
\tag{84}
$$

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	b_1	b_2
0.149	0.199	0.249	0.299	0.364	0.414	0.507	0.557	0.35	0.6

Table 3: Weights after applying pattern 1

3.1.2 Pattern 2

Continuing with the forward pass for pattern 2

$$
\begin{aligned}
net_{h1} &= \omega_1 x_1 + \omega_2 x_2 + b_1 \\
&= 0.149(0.1) + 0.199(0.9) + 0.35 \\
&= 0.544
\end{aligned}
\tag{85}
$$

$$
\begin{aligned}
out_{h1} &= \frac{1}{1 + e^{-net_{h1}}} \\
&= \frac{1}{1 + e^{-0.544}} \\
&= 0.633
\end{aligned}
\tag{86}
$$

$$
\begin{aligned}
net_{h2} &= \omega_3 x_1 + \omega_4 x_2 + b_1 \\
&= 0.249(0.1) + 0.299(0.9) + 0.35 \\
&= 0.644
\end{aligned}
\tag{87}
$$

$$
\begin{aligned}
out_{h2} &= \frac{1}{1 + e^{-net_{h2}}} \\
&= \frac{1}{1 + e^{-0.644}} \\
&= 0.656
\end{aligned}
\tag{88}
$$

Continuing with the outputs of the network

$$
\begin{aligned}
net_{o1} &= \omega_5 out_{h1} + \omega_6 out_{h2} + b_2 \\
&= 0.364(0.633) + 0.414(0.656) + 0.6 \\
&= 1.102
\end{aligned}
\tag{89}
$$

$$
\begin{aligned}
out_{o1} &= \frac{1}{1 + e^{-net_{o1}}} \\
&= \frac{1}{1 + e^{-1.102}} \\
&= 0.751
\end{aligned}
\tag{90}
$$

$$
\begin{aligned}
net_{02} &= \omega_7 out_{h1} + \omega_8 out_{h2} + b_2 \\
&= 0.507(0.633) + 0.557(0.644) + 0.6 \\
&= 1.280
\end{aligned}
\tag{91}
$$

$$
\begin{aligned}
out_{o1} &= \frac{1}{1 + e^{-net_{h1}}} \\
&= \frac{1}{1 + e^{-1.280}} \\
&= 0.782
\end{aligned}
\tag{92}
$$

Now we need to calculate the error

$$
\begin{aligned}
E_{total} &= E_{o1} + E_{02} \\
&= \frac{1}{2}(0.9 - 0.751)^2 + \frac{1}{2}(0.1 - 0.782)^2 \\
&= 0.244
\end{aligned}
\tag{93}
$$

$$
\begin{aligned}
\delta_{o1} &= -(target_{o1} - out_{o1})(out_{o1})(1 - out_{o1}) \\
&= -(0.9 - 0.751)(0.751)(1 - 0.751) \\
&= -0.028
\end{aligned}
\tag{94}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_5} &= \delta_{o1} out_{h1} \\
&= -0.028(0.633) \\
&= -0.018
\end{aligned}
\tag{95}
$$

We calculate the new value for ω_5

$$
\begin{aligned}
\omega_5^2 &= \omega_5^1 - \eta \frac{\partial E_{total}}{\partial \omega_5} \\
&= 0.364 - 0.5(-0.018) \\
&= 0.373
\end{aligned}
\tag{96}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_6} &= \delta_{o1} out_{h2} \\
&= -0.028(0.656) \\
&= -0.018
\end{aligned}
\tag{97}
$$

We calculate the new value for ω_6

$$
\begin{aligned}
\omega_6^2 &= \omega_6^1 - \eta \frac{\partial E_{total}}{\partial \omega_6} \\
&= 0.414 - 0.5(-0.018) \\
&= 0.423
\end{aligned}
\tag{98}
$$

$$
\begin{aligned}
\delta_{o2} &= -(target_{o2} - out_{o2})(out_{o2})(1 - out_{o2}) \\
&= -(0.1 - 0.782)(0.782)(1 - 0.782) \\
&= 0.166
\end{aligned}
\tag{99}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_7} &= \delta_{o2} out_{h1} \\
&= 0.166(0.633) \\
&= 0.073
\end{aligned}
\tag{100}
$$

We calculate the new value for ω_7

$$
\begin{aligned}
\omega_7^2 &= \omega_7^1 - \eta \frac{\partial E_{total}}{\partial \omega_7} \\
&= 0.507 - 0.5(0.073) \\
&= 0.471
\end{aligned}
\tag{101}
$$

$$\frac{\partial E_{total}}{\partial \omega_8} = \delta_{o2}out_{h2}$$
$$= 0.166(0.656)$$
$$= 0.076$$

(102)

We calculate the new value for ω_8

$$\omega_8^2 = \omega_8^1 - \eta\frac{\partial E_{total}}{\partial \omega_8}$$
$$= 0.557 - 0.5(0.076)$$
$$= 0.519$$

(103)

Now we need to pass the error back to the hidden units

$$\delta_{h1} = (\delta_{o1}\omega_5^1 + \delta_{o2}\omega_7^1)(out_{h1})(1 - out_{h1})$$
$$= (-0.028(0.364) + 0.116(0.507))(0.507)(1 - 0.507)$$
$$= 0.011$$

(104)

$$\frac{\partial E_{total}}{\partial \omega_1} = \delta_{h1}x_1$$
$$= 0.011(0.1)$$
$$= 0.001$$

(105)

$$\omega_1^2 = \omega_1^1 - \eta\frac{\partial E_{total}}{\partial \omega_1}$$
$$= 0.149 - 0.5(0.001)$$
$$= 0.149$$

(106)

$$\frac{\partial E_{total}}{\partial \omega_2} = \delta_{h1}x_2$$
$$= 0.011(0.9)$$
$$= 0.010$$

(107)

$$\omega_2^2 = \omega_2^1 - \eta\frac{\partial E_{total}}{\partial \omega_2}$$
$$= 0.199 - 0.5(0.010)$$
$$= 0.194$$

(108)

$$\delta_{h2} = (\delta_{o1}\omega_6^1 + \delta_{o2}\omega_8^1)(out_{h2})(1 - out_{h2})$$
$$= (-0.028(0.414) + 0.116(0.557))(0.656)(1 - 0.656)$$
$$= 0.012$$

(109)

$$\frac{\partial E_{total}}{\partial \omega_3} = \delta_{h2}x_1$$
$$= 0.012(0.1)$$
$$= 0.001$$

(110)

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	b_1	b_2
0.149	0.194	0.249	0.294	0.373	0.423	0.471	0.519	0.35	0.6

Table 4: Weights after applying pattern 2

$$
\begin{aligned}
\omega_3^2 &= \omega_3^1 - \eta \frac{\partial E_{total}}{\partial \omega_3} \\
&= 0.249 - 0.5(0.001) \\
&= 0.249
\end{aligned}
\tag{111}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_4} &= \delta_{h2} x_2 \\
&= 0.012(0.9) \\
&= 0.011
\end{aligned}
\tag{112}
$$

$$
\begin{aligned}
\omega_4^2 &= \omega_4^1 - \eta \frac{\partial E_{total}}{\partial \omega_4} \\
&= 0.299 - 0.5(0.011) \\
&= 0.294
\end{aligned}
\tag{113}
$$

3.1.3 Pattern 3

Continuing with the forward pass for pattern 3

$$
\begin{aligned}
net_{h1} &= \omega_1 x_1 + \omega_2 x_2 + b_1 \\
&= 0.149(0.9) + 0.194(0.1) + 0.35 \\
&= 0.504
\end{aligned}
\tag{114}
$$

$$
\begin{aligned}
out_{h1} &= \frac{1}{1 + e^{-net_{h1}}} \\
&= \frac{1}{1 + e^{-0.504}} \\
&= 0.623
\end{aligned}
\tag{115}
$$

$$
\begin{aligned}
net_{h2} &= \omega_3 x_1 + \omega_4 x_2 + b_1 \\
&= 0.249(0.9) + 0.294(0.1) + 0.35 \\
&= 0.604
\end{aligned}
\tag{116}
$$

$$
\begin{aligned}
out_{h2} &= \frac{1}{1 + e^{-net_{h2}}} \\
&= \frac{1}{1 + e^{-0.604}} \\
&= 0.646
\end{aligned}
\tag{117}
$$

Continuing with the outputs of the network

$$
\begin{aligned}
net_{o1} &= \omega_5 out_{h1} + \omega_6 out_{h2} + b_2 \\
&= 0.373(0.623) + 0.423(0.646) + 0.6 \\
&= 1.106
\end{aligned}
\tag{118}
$$

$$out_{o1} = \frac{1}{1 + e^{-net_{o1}}}$$
$$= \frac{1}{1 + e^{-1.106}} \tag{119}$$
$$= 0.751$$

$$net_{02} = \omega_7 out_{h1} + \omega_8 out_{h2} + b_2$$
$$= 0.471(0.623) + 0.519(0.646) + 0.6 \tag{120}$$
$$= 1.229$$

$$out_{o1} = \frac{1}{1 + e^{-net_{h1}}}$$
$$= \frac{1}{1 + e^{-1.229}} \tag{121}$$
$$= 0.774$$

Now we need to calculate the error

$$E_{total} = E_{o1} + E_{02}$$
$$= \frac{1}{2}(0.9 - 0.751)^2 + \frac{1}{2}(0.1 - 0.774)^2 \tag{122}$$
$$= 0.238$$

$$\delta_{o1} = -(target_{o1} - out_{o1})(out_{o1})(1 - out_{o1})$$
$$= -(0.9 - 0.751)(0.751)(1 - 0.751) \tag{123}$$
$$= -0.028$$

$$\frac{\partial E_{total}}{\partial \omega_5} = \delta_{o1} out_{h1}$$
$$= -0.028(0.623) \tag{124}$$
$$= -0.017$$

We calculate the new value for ω_5

$$\omega_5^3 = \omega_5^2 - \eta \frac{\partial E_{total}}{\partial \omega_5}$$
$$= 0.373 - 0.5(-0.017) \tag{125}$$
$$= 0.382$$

$$\frac{\partial E_{total}}{\partial \omega_6} = \delta_{o1} out_{h2}$$
$$= -0.028(0.646) \tag{126}$$
$$= -0.018$$

We calculate the new value for ω_6

$$\omega_6^3 = \omega_6^2 - \eta \frac{\partial E_{total}}{\partial \omega_6}$$
$$= 0.423 - 0.5(-0.018) \tag{127}$$
$$= 0.432$$

$$\begin{aligned} \delta_{o2} &= -(target_{o2} - out_{o2})(out_{o2})(1 - out_{o2}) \\ &= -(0.1 - 0.774)(0.774)(1 - 0.774) \\ &= 0.118 \end{aligned} \tag{128}$$

$$\begin{aligned} \frac{\partial E_{total}}{\partial \omega_7} &= \delta_{o2} out_{h1} \\ &= 0.118(0.623) \\ &= 0.074 \end{aligned} \tag{129}$$

We calculate the new value for ω_7

$$\begin{aligned} \omega_7^3 &= \omega_7^2 - \eta \frac{\partial E_{total}}{\partial \omega_7} \\ &= 0.471 - 0.5(0.074) \\ &= 0.434 \end{aligned} \tag{130}$$

$$\begin{aligned} \frac{\partial E_{total}}{\partial \omega_8} &= \delta_{o2} out_{h2} \\ &= 0.118(0.646) \\ &= 0.077 \end{aligned} \tag{131}$$

We calculate the new value for ω_8

$$\begin{aligned} \omega_8^3 &= \omega_8^2 - \eta \frac{\partial E_{total}}{\partial \omega_8} \\ &= 0.519 - 0.5(0.077) \\ &= 0.480 \end{aligned} \tag{132}$$

Now we need to pass the error back to the hidden units

$$\begin{aligned} \delta_{h1} &= (\delta_{o1}\omega_5^2 + \delta_{o2}\omega_7^2)(out_{h1})(1 - out_{h1}) \\ &= (-0.028(0.373) + 0.118(0.471))(0.623)(1 - 0.623) \\ &= 0.011 \end{aligned} \tag{133}$$

$$\begin{aligned} \frac{\partial E_{total}}{\partial \omega_1} &= \delta_{h1} x_1 \\ &= 0.011(0.9) \\ &= 0.010 \end{aligned} \tag{134}$$

$$\begin{aligned} \omega_1^3 &= \omega_1^2 - \eta \frac{\partial E_{total}}{\partial \omega_1} \\ &= 0.149 - 0.5(0.010) \\ &= 0.144 \end{aligned} \tag{135}$$

$$\begin{aligned} \frac{\partial E_{total}}{\partial \omega_2} &= \delta_{h1} x_2 \\ &= 0.011(0.1) \\ &= 0.001 \end{aligned} \tag{136}$$

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	b_1	b_2
0.144	0.194	0.244	0.294	0.382	0.432	0.434	0.480	0.35	0.6

Table 5: Weights after applying pattern 3

$$
\begin{aligned}
\omega_2^3 &= \omega_2^2 - \eta \frac{\partial E_{total}}{\partial \omega_2} \\
&= 0.194 - 0.5(0.001) \\
&= 0.194
\end{aligned}
\tag{137}
$$

$$
\begin{aligned}
\delta_{h2} &= (\delta_{o1}\omega_6^2 + \delta_{o2}\omega_8^2)(out_{h2})(1 - out_{h2}) \\
&= (-0.028(0.423) + 0.118(0.519))(0.646)(1 - 0.646) \\
&= 0.011
\end{aligned}
\tag{138}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_3} &= \delta_{h2}x_1 \\
&= 0.011(0.9) \\
&= 0.010
\end{aligned}
\tag{139}
$$

$$
\begin{aligned}
\omega_3^3 &= \omega_3^2 - \eta \frac{\partial E_{total}}{\partial \omega_3} \\
&= 0.249 - 0.5(0.010) \\
&= 0.244
\end{aligned}
\tag{140}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_4} &= \delta_{h2}x_2 \\
&= 0.011(0.1) \\
&= 0.001
\end{aligned}
\tag{141}
$$

$$
\begin{aligned}
\omega_4^3 &= \omega_4^2 - \eta \frac{\partial E_{total}}{\partial \omega_4} \\
&= 0.294 - 0.5(0.001) \\
&= 0.294
\end{aligned}
\tag{142}
$$

3.1.4 Pattern 4

Continuing with the forward pass for pattern 4

$$
\begin{aligned}
net_{h1} &= \omega_1 x_1 + \omega_2 x_2 + b_1 \\
&= 0.144(0.9) + 0.194(0.9) + 0.35 \\
&= 0.654
\end{aligned}
\tag{143}
$$

$$
\begin{aligned}
out_{h1} &= \frac{1}{1 + e^{-net_{h1}}} \\
&= \frac{1}{1 + e^{-0.654}} \\
&= 0.658
\end{aligned}
\tag{144}
$$

26

$$
\begin{aligned}
net_{h2} &= \omega_3 x_1 + \omega_4 x_2 + b_1 \\
&= 0.244(0.9) + 0.294(0.9) + 0.35 \\
&= 0.834
\end{aligned}
\tag{145}
$$

$$
\begin{aligned}
out_{h2} &= \frac{1}{1 + e^{-net_{h2}}} \\
&= \frac{1}{1 + e^{-0.834}} \\
&= 0.697
\end{aligned}
\tag{146}
$$

Continuing with the outputs of the network

$$
\begin{aligned}
net_{o1} &= \omega_5 out_{h1} + \omega_6 out_{h2} + b_2 \\
&= 0.382(0.658) + 0.432(0.697) + 0.6 \\
&= 1.152
\end{aligned}
\tag{147}
$$

$$
\begin{aligned}
out_{o1} &= \frac{1}{1 + e^{-net_{o1}}} \\
&= \frac{1}{1 + e^{-1.152}} \\
&= 0.760
\end{aligned}
\tag{148}
$$

$$
\begin{aligned}
net_{02} &= \omega_7 out_{h1} + \omega_8 out_{h2} + b_2 \\
&= 0.434(0.658) + 0.480(0.697) + 0.6 \\
&= 1.220
\end{aligned}
\tag{149}
$$

$$
\begin{aligned}
out_{o1} &= \frac{1}{1 + e^{-net_{h1}}} \\
&= \frac{1}{1 + e^{-1.220}} \\
&= 0.772
\end{aligned}
\tag{150}
$$

Now we need to calculate the error

$$
\begin{aligned}
E_{total} &= E_{o1} + E_{02} \\
&= \frac{1}{2}(0.1 - 0.760)^2 + \frac{1}{2}(0.9 - 0.772)^2 \\
&= 0.226
\end{aligned}
\tag{151}
$$

$$
\begin{aligned}
\delta_{o1} &= -(target_{o1} - out_{o1})(out_{o1})(1 - out_{o1}) \\
&= -(0.1 - 0.760)(0.760)(1 - 0.760) \\
&= 0.120
\end{aligned}
\tag{152}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_5} &= \delta_{o1} out_{h1} \\
&= 0.120(0.658) \\
&= 0.079
\end{aligned}
\tag{153}
$$

We calculate the new value for ω_5

$$
\begin{aligned}
\omega_5^4 &= \omega_5^3 - \eta \frac{\partial E_{total}}{\partial \omega_5} \\
&= 0.382 - 0.5(0.079) \\
&= 0.342
\end{aligned}
\tag{154}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_6} &= \delta_{o1} out_{h2} \\
&= 0.120(0.697) \\
&= 0.084
\end{aligned}
\tag{155}
$$

We calculate the new value for ω_6

$$
\begin{aligned}
\omega_6^4 &= \omega_6^3 - \eta \frac{\partial E_{total}}{\partial \omega_6} \\
&= 0.432 - 0.5(0.084) \\
&= 0.390
\end{aligned}
\tag{156}
$$

$$
\begin{aligned}
\delta_{o2} &= -(target_{o2} - out_{o2})(out_{o2})(1 - out_{o2}) \\
&= -(0.9 - 0.772)(0.772)(1 - 0.772) \\
&= -0.023
\end{aligned}
\tag{157}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_7} &= \delta_{o2} out_{h1} \\
&= -0.023(0.658) \\
&= -0.015
\end{aligned}
\tag{158}
$$

We calculate the new value for ω_7

$$
\begin{aligned}
\omega_7^4 &= \omega_7^3 - \eta \frac{\partial E_{total}}{\partial \omega_7} \\
&= 0.434 - 0.5(-0.015) \\
&= 0.441
\end{aligned}
\tag{159}
$$

$$
\begin{aligned}
\frac{\partial E_{total}}{\partial \omega_8} &= \delta_{o2} out_{h2} \\
&= -0.023(0.697) \\
&= -0.016
\end{aligned}
\tag{160}
$$

We calculate the new value for ω_8

$$
\begin{aligned}
\omega_8^4 &= \omega_8^3 - \eta \frac{\partial E_{total}}{\partial \omega_8} \\
&= 0.480 - 0.5(-0.016) \\
&= 0.488
\end{aligned}
\tag{161}
$$

Now we need to pass the error back to the hidden units

$$
\begin{aligned}
\delta_{h1} &= (\delta_{o1}\omega_5^3 + \delta_{o2}\omega_7^3)(out_{h1})(1 - out_{h1}) \\
&= (0.120(0.382) - 0.023(0.434))(0.658)(1 - 0.658) \\
&= 0.008
\end{aligned}
\tag{162}
$$

$$\frac{\partial E_{total}}{\partial \omega_1} = \delta_{h1} x_1$$
$$= 0.008(0.9)$$
$$= 0.007$$
(163)

$$\omega_1^4 = \omega_1^3 - \eta \frac{\partial E_{total}}{\partial \omega_1}$$
$$= 0.144 - 0.5(0.007)$$
$$= 0.141$$
(164)

$$\frac{\partial E_{total}}{\partial \omega_2} = \delta_{h1} x_2$$
$$= 0.008(0.9)$$
$$= 0.007$$
(165)

$$\omega_2^4 = \omega_2^3 - \eta \frac{\partial E_{total}}{\partial \omega_2}$$
$$= 0.194 - 0.5(0.007)$$
$$= 0.191$$
(166)

$$\delta_{h2} = (\delta_{o1} \omega_6^3 + \delta_{o2} \omega_8^3)(out_{h2})(1 - out_{h2})$$
$$= (0.120(0.432) - 0.023(0.480))(0.697)(1 - 0.697)$$
$$= 0.009$$
(167)

$$\frac{\partial E_{total}}{\partial \omega_3} = \delta_{h2} x_1$$
$$= 0.009(0.9)$$
$$= 0.008$$
(168)

$$\omega_3^4 = \omega_3^3 - \eta \frac{\partial E_{total}}{\partial \omega_3}$$
$$= 0.244 - 0.5(0.008)$$
$$= 0.240$$
(169)

$$\frac{\partial E_{total}}{\partial \omega_4} = \delta_{h2} x_2$$
$$= 0.009(0.9)$$
$$= 0.008$$
(170)

$$\omega_4^3 = \omega_4^2 - \eta \frac{\partial E_{total}}{\partial \omega_4}$$
$$= 0.294 - 0.5(0.008)$$
$$= 0.290$$
(171)

The resultant weights after all 4 patters can be found in Table 6

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	b_1	b_2
0.141	0.191	0.240	0.290	0.342	0.390	0.441	0.488	0.35	0.6

Table 6: Weights after applying pattern 4

3.2   Question 3(b)

According to Mitchell (1997) the following functions can be modeled by feed forward networks. Firstly any boolean function can be represented by using 2 layers of units (not counting the input layer). But, in the worst case, the number of hidden unit will grow exponentially with the number of inputs. Secondly any Continuous bounded function can be modeled to an arbitrary level of accuracy with a network with one hidden layer. Thirdly any arbitrary function can be approximated with a network containing 2 layers of hidden units.

3.3   Question 3(c)

The error surface for multi-layered networks contains multiple local minima. The gradient descent method used can only converge to a local minima which is not necessarily the global minimum. When the gradient decent process starts close to a local minimum the algorithm will follow the gradient to the local minimum (Mitchell, 1997).

Networks with higher dimensions may suffer less from local minima because when the network reaches a local minimum in a certain dimension, the other dimensions are not necessarily at a minimum. Their weights can get the network out of some local minima (Mitchell, 1997).

The problem of local minima can also lessened by using a momentum term. As in

$$\omega_i^{j+1} = \omega_i^j - \eta \frac{\partial E_{total}}{\partial \omega_i^j} + \alpha(\omega_i^j - \omega_i^{j-1}) \tag{172}$$

where alpha is the momentum constant and this depends on the previous change in the weight. This way the algorithm can get out of a relatively small local minimum (Mitchell, 1997).

One technique is so called "on-line training". Here the weights are updated after each pattern is processed rather than batch or off-line where the weights are updated after all the patterns are processed. On-line training can add noise to the gradient direction and thereby avoiding some shallow local minima (Rojas, 1996).

Another method is simulated annealing (Rojas, 1996);

3.4   Question 3(d)

A discussion of the identity function can be found in (Mitchell, 1997, p106) at $http://www.cs.ubbcluj.ro/\sim gabis/ml/ml-books/McGrawHill-MachineLearning-TomMitchell.pdf$

Using a 8 X 3 X 8 network we want to represent the identity function. In other words when any input node is given an high input we want the corresponding output node to be high. Thus if input 2 is high and all the other nodes are low (01000000), output 2 needs to be high and all other outputs low (01000000). The problem is that there is not enough nodes in the hidden layer to simply forward the value from the input node to the corresponding output node.

When we attempt to train such a network the network finds a representation in the hidden layer to represent all 8 inputs. The network uses a binary representation where the input (00100000) will be represented as (011) in the hidden units. This will then be output as (00100000). This way the network learn a new representation that it was not explicitly told.

The values from the book can be found in table 7

3.5   Question 3(e)

The sigmoid function is defined as

$$\sigma(y) = \frac{1}{1 + e^{-y}} \tag{173}$$

Input	h_1	h_2	h_3	Output
10000000	0.89	0.04	0.08	10000000
01000000	0.15	0.99	0.99	01000000
00100000	0.01	0.97	0.27	00100000
00010000	0.99	0.97	0.71	00010000
00001000	0.03	0.05	0.02	00001000
00000100	0.01	0.11	0.88	00000100
00000010	0.80	0.01	0.98	00000010
00000001	0.60	0.94	0.01	00000001

Table 7: Hidden unit representation

Firstly the output of this activation function is continuous. This makes it easier to connect multiple neurons together and weight different inputs form different neurons. The function is also easily differentiable. Secondly because the function is very steep it tends to force values to be high or low. Which makes it suitable as a classifier. Thirdly because the output of the sigmoid is always bound, our function wont grow out of bounds (Avinash Sharma V, 2017).

Another possibility is the *tanh* function (Avinash Sharma V, 2017; C.M. Bishop, 2006).

$$f(x) = tanh(x)$$
$$= \frac{2}{1 - e^{-2x}} - 1 \tag{174}$$

As we can see the *tanh* function is very similar to the sigmoid function and has thus most of the same advantages. But the angle of *tanh* is steeper than the sigmoid function.

Another possibility is the linear function.

$$f(x) = mx + c \tag{175}$$

The problem with the linear function is that stacking multiple neurons is always equivalent to 1 linear function. Thus we lose the ability to classify with multiple neurons. Also the derivative of a linear function is not a function of x. Thus the backpropagation step will always make a constant step. This is not desirable.

References

Avinash Sharma V. (2017). *Understanding Activation Functions in Neural Networks.* Retrieved 2018-07-15, from https://medium.com/the-theory-of-everything/understanding-activation-functions -in-neural-networks-9491262884e0

C.M. Bishop. (2006). *Pattern Recognition and Machine Learning.* Retrieved from http://users.isr.ist.utl.pt/ wurmd/Livros/school/Bishop - Pattern Recognition And Machine Learning - Springer 2006.pdf

Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill.

Nils J Nilsson. (1998). *Introduction to Machine Learning.* Retrieved from http://robotics.stanford.edu/people/nilsson/MLBOOK.pdf

Rojas, R. (1996). *Neural Networks A Systematic Introduction.* Springer-Verlag Springer-Verlag. Retrieved from https://page.mi.fu-berlin.de/rojas/neural/neuron.pdf

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. Retrieved from http://www.cs.toronto.edu/hinton/absps/naturebp.pdf doi: 10.1038/323533a0