## Homework-2 Solutions

# Question 1

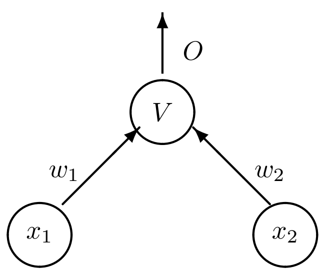
$$g(h) = \frac{1}{1+|h|} = \begin{cases} \frac{1}{1+h} & \text{if } h \geq 0 \\ \frac{1}{1-h} & \text{if } h < 0 \end{cases}$$

$$g'(h) = \begin{cases} \frac{-1}{(1+h)^2} = -O^2 & \text{if } h \geq 0 \\ \frac{1}{(1-h)^2} = O^2 & \text{if } h < 0 \end{cases}$$

As was shown in class, delta is given by:

$$\delta = g'(h)(y - O) = \begin{cases} -O^2(y - O) & \text{if } h \geq 0 \\ O^2(y - O) & \text{if } h < 0 \end{cases}$$

# Question 2




| | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| $e_1$ | 1 | 0 | 1 |
| $e_2$ | 1 | 1 | 1 |
| $e_3$ | 2 | 0 | 0 |
| $e_4$ | 2 | 1 | 0 |

You are given a perceptron implemented with a sigmoid, with $\beta = 1$. There are NO bias connections. The initial values of the weights are $w_1 = 0$, $w_2 = 1$.

## Part 1

Give explicit expressions to the way the weights change if the network is given the example $e_3$. Use $\epsilon = 0.1$. You may use temporary variables in your answer, but make sure that they are all specified in terms of the given values. You may use the notation $S(.)$ instead of explicitly computing sigmoid values.

$$h = w_1 x_1 + w_2 x_2 = 0 \cdot 2 + 1 \cdot 0 = 0$$

$$O = V = S(0) = \frac{1}{2}$$

$$\delta = V(1 - V)(0 - V) = -V^2(1 - V) = -\frac{1}{8}$$

$$\text{new } w_1 = w_1 + 0.1 \cdot 2 \cdot \delta = -0.025$$

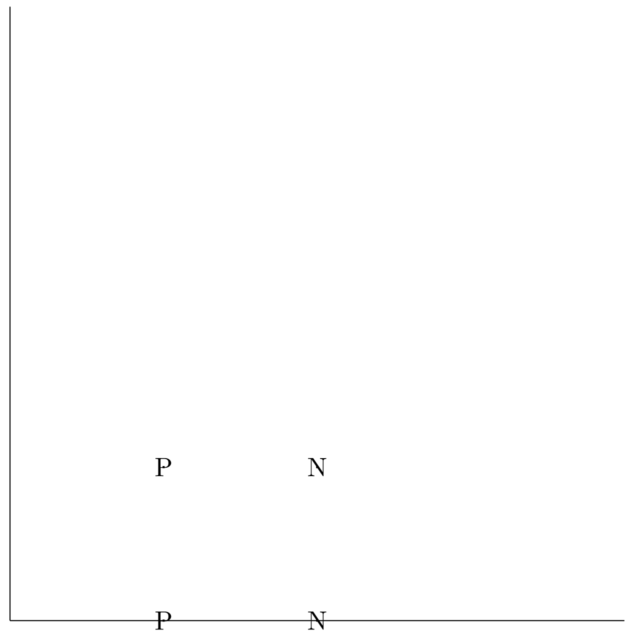
$$\text{new } w_2 = w_2 + 0 = 1$$

## Part 2

**a.** If you could choose $\epsilon$ to be as small as you like, and run back propagation as many epochs as you like with the four examples $e_1$, $e_2$, $e_3$, $e_4$, do you expect the computed output of the perceptron to be within 0.001 of the desired output (the value of $y$) for all four examples?

**Answer:** NO

Here is a picture of the examples in the $x_1, x_2$ plane:




In order for the perceptron to correctly classify all the examples there should be a separating line passing through the origin. But there is no such line.

**b.** Will your answer to a. remain the same if bias connection is added (everything else stays as in a.)?

**Answer:**

- NO. My answer changes. With this new condition
  - my new answer to a. is YES.

With bias, the separation is with an arbitrary line. And such line clearly exists.

**c.** Will your answer to a. remain the same if the training is performed with back propagation with momentum? (everything else stays as in a.)?

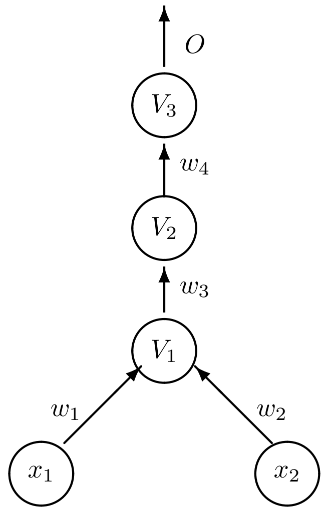
**Answer:**

- YES. My answer is exactly the same as in Part a.

Momentum only affects the speed of convergence.

# Question 3

The following neural network has 3 layers with nodes that compute the sigmoid function. There are no bias connections.



Write an explicit expression to how back propagation (applied to minimize the least squares error function) changes the values of $w_1, \ldots, w_4$ when the algorithm is given the example $x_1 = 1, x_2 = -1$, with the desired response $y = 0$. Assume that $\epsilon = 0.1$, $\beta = 1$. You may use $S(.)$ instead of explicitly computing sigmoid values.

## Answer

Propagate forward:

$$\begin{aligned} h_1 &= w_1 - w_2 & V_1 &= S(h_1) \\ h_2 &= w_3 V_1 & V_2 &= S(h_2) \\ h_3 &= w_4 V_2 & V_3 &= S(h_3) \end{aligned}$$

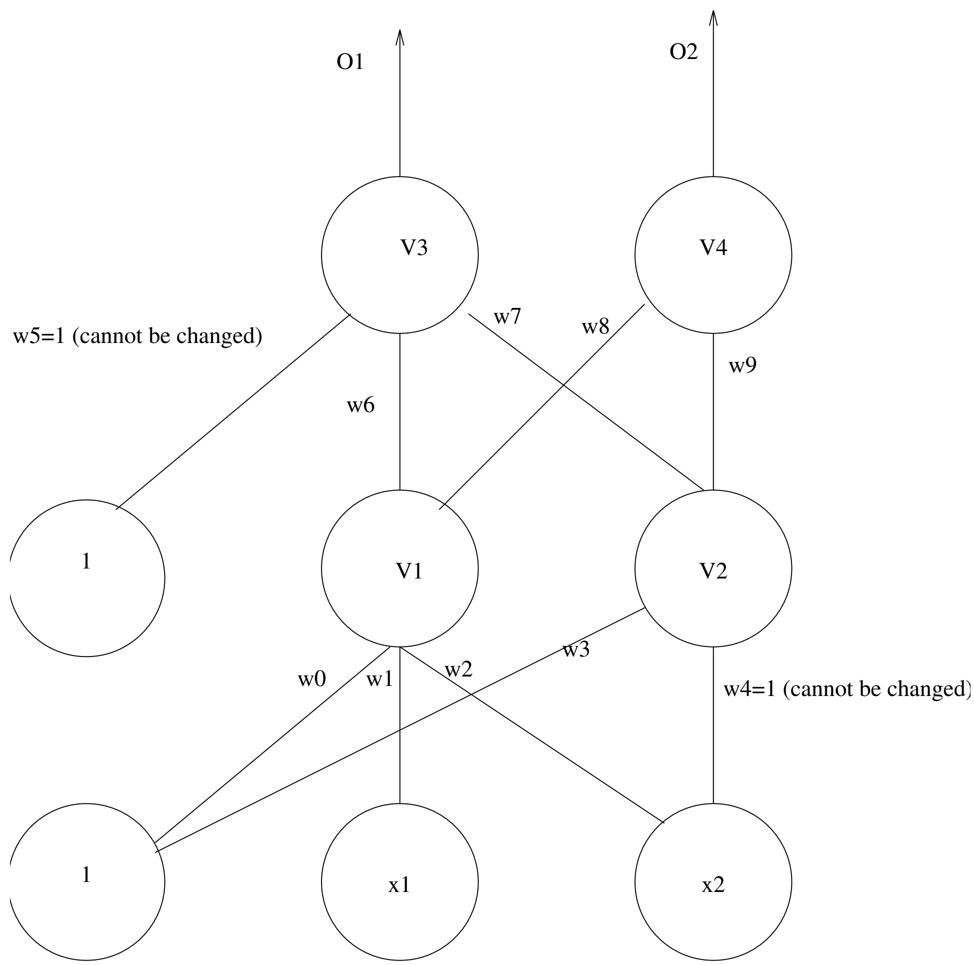
$$O = V_3$$

Back propagation:

$$\begin{aligned} &\delta_3 = 2O(1-O)(y-O) = -2O^2(1-O) \\ &\delta_2 = 2V_2(1-V_2)w_4\delta_3 \\ &\delta_1 = 2V_1(1-V_1)w_3\delta_2 \end{aligned}$$

$$\begin{aligned} &\text{new } w_4 \leftarrow w_4 + \epsilon\delta_3 V_2 \\ &\text{new } w_3 \leftarrow w_3 + \epsilon\delta_2 V_1 \\ &\text{new } w_2 \leftarrow w_2 + \epsilon\delta_1 x_2 \\ &\text{new } w_1 \leftarrow w_1 + \epsilon\delta_1 x_1 \end{aligned}$$

## Question 4




The above neural network has two layers (one hidden layer), two inputs, and two outputs. There is NO bias connection for $V_4$, and there is NO connection between $V_2$ and $x_1$. The weights $w_4$ and $w_5$ have the value 1 and cannot be changed. All nodes compute the sigmoid function with $\beta = 1$.

**A.1** Give explicit expressions to the values of all nodes in forward propagation when the network is given the input $x_1 = 2$, $x_2 = 3$, with the desired output $y_1 = 0$, $y_2 = 2$ Your answer should be in terms of the old weights $w_0$, $w_1$, $w_2$, $w_3$, $w_6$, $w_7$, $w_8$, $w_9$. Do not use any other temporary variables (such as $h_i$). If needed, you may use the notation $S(.)$ instead of explicitly computing sigmoid values.

**Answer**

$$
\begin{aligned}
V_1 &= S(w_0 + 2w_1 + 3w_2) \\
V_2 &= S(w_3 + 3) \\
V_3 &= S(1 + w_6 V_1 + w_7 V_2) = S(1 + w_6 S(w_0 + 2w_1 + 3w_2) + w_7 S(w_3 + 3)) \\
V_4 &= S(w_8 V_1 + w_9 V_2) = S(w_8 S(w_0 + 2w_1 + 3w_2) + w_9 S(w_3 + 3))
\end{aligned}
$$

**A.2** Give explicit expressions to the output produces by the network when given this example. Your answer should be in terms of the old weights $w_0$, $w_1$, $w_2$, $w_3$, $w_6$, $w_7$, $w_8$, $w_9$, and the values $V_1$, $V_2$, $V_3$, $V_4$ computed in A.1. Do not use any other temporary variables (such as $h_i$). If needed, you may use the notation $S(.)$ instead of explicitly computing sigmoid values.

**Answer**

$$\begin{aligned} O_1 &= V_3 \\ O_2 &= V_4 \end{aligned}$$

**A.3** Give explicit expressions to how the weights change by back propagation when the network is given the same example as above. Use $\epsilon = 0.1$.

Your answer should be in terms of the old weights $w_0$, $w_1$, $w_2$, $w_3$, $w_6$, $w_7$, $w_8$, $w_9$. and the values $V_1$, $V_2$, $V_3$, $V_4$ computed in A.1. and the values $O_1$, $O_2$ computed in A.2. If needed, you may use the notation $S(.)$ instead of explicitly computing sigmoid values. You may use temporary variables in your answer, but make sure that they are defined in terms of the above variables.

**Answer** I am using the following temporary variables in my answer:

$$\begin{aligned} \delta_4 &= 2O_2(1-O_2)(2-O_2) \\ \delta_3 &= -2O_1^2(1-O_1) \\ \delta_2 &= 2V_2(1-V_2)(w_7\delta_3 + w_9\delta_4) \\ \delta_1 &= 2V_1(1-V_1)(w_6\delta_3 + w_8\delta_4) \end{aligned}$$

The expressions for the new weights are:

$$\begin{aligned} \text{new } w_0 &= w_0 + \epsilon\delta_1 = w_0 + 0.1\delta_1 \\ \text{new } w_1 &= w_1 + \epsilon\delta_1 x_1 = w_1 + 0.2\delta_1 \\ \text{new } w_2 &= w_2 + \epsilon\delta_1 x_2 = w_2 + 0.3\delta_1 \\ \text{new } w_3 &= w_3 + \epsilon\delta_2 = w_3 + 0.1\delta_2 \\ \text{new } w_6 &= w_6 + 0.1\delta_3 V_1 \\ \text{new } w_7 &= w_7 + 0.1\delta_3 V_2 \\ \text{new } w_8 &= w_8 + 0.1\delta_4 V_1 \\ \text{new } w_9 &= w_9 + 0.1\delta_4 V_2 \end{aligned}$$