Homework-2 Solutions

Question 1

Consider the following methods for evaluating the performance of a learning algorithm in a particular setting.

Method 1. 20 examples are drawn at random. The method of 5-fold cross validation is applied, giving the error estimate e_1.

Method 2. 16 examples are drawn at random. The classifier is trained using these examples. Another set of 1000 examples are drawn at random, and the classifier is tested on these examples giving the error estimate e_2.

Method 3. 1000 examples are drawn at random. The method of 5-fold cross validation is applied, giving the error estimate e_3.

What is the *likely* relation between the error estimates? Mark all answers that you believe will hold with probability greater than 0.5.

1. $e_1 < e_2$.

2. $e_1 > e_2$.

3. $e_1 < e_3$.

$\boxed{\textbf{4.}}$ $e_1 > e_3$.

5. $e_2 < e_3$.

$\boxed{\textbf{6.}}$ $e_2 > e_3$.

7. None of the above will hold with probability greater than 0.5.

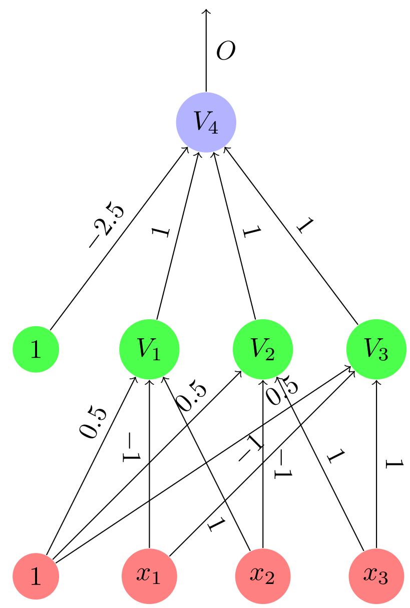
Question 2

We would like to build a neural net to compute the following Boolean function:

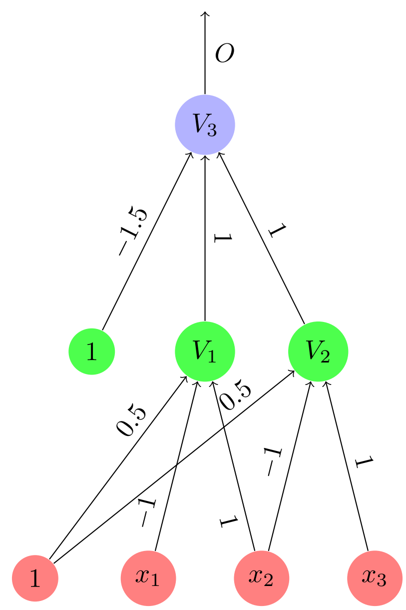
$$(\bar{x_1} \vee x_2) \wedge (\bar{x_2} \vee x_3) \wedge (\bar{x_1} \vee x_3)$$

The input to the network is the three variables x_1, x_2, x_3, and a constant bias. Use additional bias nodes as necessary. Draw the network and specify the values of all the weights.

Answer



Another right answer:



Question 3

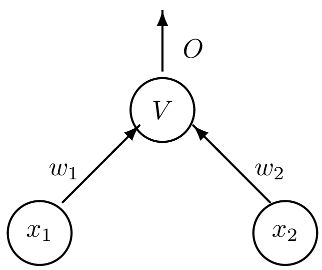
$$g(h) = \frac{1}{1+|h|} = \begin{cases} \frac{1}{1+h} & \text{if } h \geq 0 \\ \frac{1}{1-h} & \text{if } h < 0 \end{cases}$$

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

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 4



	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.

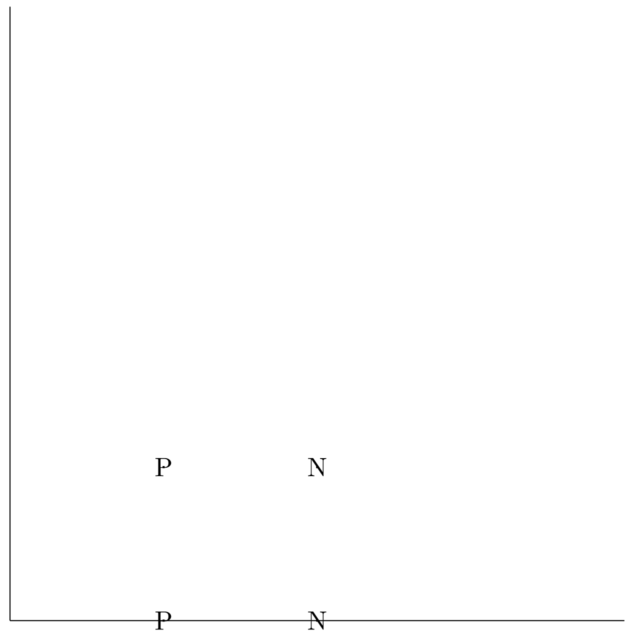
$$\begin{aligned}
&h = w_1x_1 + w_2x_2 = 0 \cdot 2 + 1 \cdot 0 = 0 \\
&O = V = S(0) = \frac{1}{2} \\
&\delta = V(1-V)(0-V) = -V^2(1-V) = -\frac{1}{8} \\
&\text{new } w_1 = w_1 + 0.1 \cdot 2 \cdot \delta = -0.025 \\
&\text{new } w_2 = w_2 + 0 = 1
\end{aligned}$$

Part 2

a. If you could choose ϵ 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.