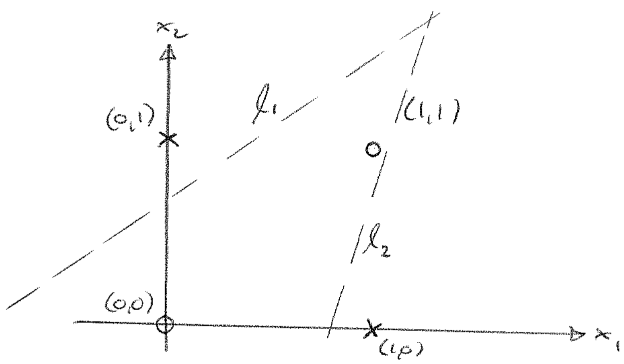
COMS 4030A

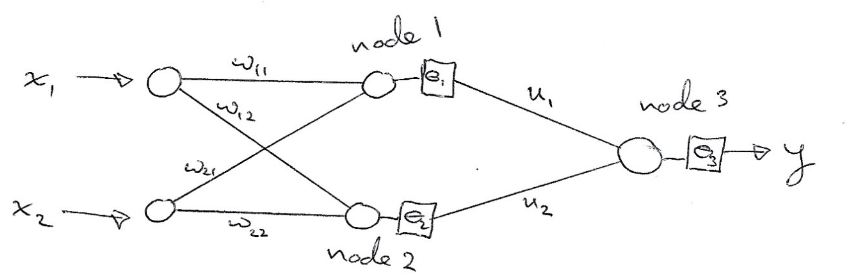
Adaptive Computation and Machine Learning

The dataset $D = \{((0,0),0),((0,1),1),((1,0),1),((1,1),0)\}$ is not linearly separable, i.e., it is impossible to draw a straight line in $\mathbb{R}^2$ so that all the 1's are on one side of the line and all the 0's are on the other. However, it is possible to separate the dataset using **two** straight lines, as shown below, where a $\circ$ represents a target of 0 and a $\times$ represents a target of 1.



If a point lies above line ℓ_1 and above line ℓ_2 then it is a $\times$. If it lies below ℓ_1 and above ℓ_2 then it is a $\circ$; and if it lies below line ℓ_1 and below line ℓ_2 then it is a $\times$.

We can use a 2-layer perceptron to model the separation by two straight lines as follows:



EXERCISE: Try to find suitable values for w_{11}, w_{12}, w_{21}, w_{22}, u_1, u_2, θ_1, θ_2 and θ_3 so that the above 2-layer perceptron correctly outputs the data in D.

Solution:

There are many possible solutions. For example,

$w_{11} = 1$, $w_{12} = 1$, $w_{21} = -1$, $w_{22} = -1$, $u_1 = 2$, $u_2 = -1$, $\theta_1 = \frac{1}{2}$, $\theta_2 = -\frac{1}{2}$ and $\theta_3 = -\frac{1}{2}$

(1) Consider a network with 4 input nodes, one hidden layer with 4 nodes and 2 output nodes. The weights between layer 0 and layer 1, and the biases at layer 1 are given by:

$$W_1 = \begin{bmatrix} 4 & -5 & 0 & 1 \\ -3 & 6 & -1 & 2 \\ 0 & 1 & 1 & -2 \\ 2 & 0 & -3 & 4 \end{bmatrix}, \qquad \boldsymbol{b}_1 = (1, -2, 0, -1).$$

The weights between layer 1 and layer 2, and the biases at layer 2 are given by:

$$W_2 = \begin{bmatrix} -2 & 1 \\ -1 & 0 \\ 1 & -3 \\ 5 & -1 \end{bmatrix}, \qquad \boldsymbol{b}_2 = (-2, 2).$$

(a) Using the sigmoid activation function σ at each node, calculate the output from the network for the following inputs:

(i) $(1, 0, 2, -3)$

(ii) $(-2, 3, 1, 0)$

(iii) $(1, 0, 1, 1)$

(b) Using the activation function *relu* at each node, calculate the output from the network with same inputs as above.

(c) Using *relu* at the hidden layer and σ at the output layer, calculate the output from the network with same inputs as above.

Solutions: (a) (i)

$$(1, 0, 2, -3)W_1 = (1, 0, 2, -3) \begin{bmatrix} 4 & -5 & 0 & 1 \\ -3 & 6 & -1 & 2 \\ 0 & 1 & 1 & -2 \\ 2 & 0 & -3 & 4 \end{bmatrix} = (-2, -3, 11, -15)$$

$(-2, -3, 11, -15) + \boldsymbol{b}_1 = (-2, -3, 11, -15) + (1, -2, 0, -1) = (-1, -5, 11, -16).$

$$\sigma(-1, -5, 11, -16) = (0.269, 0.001, 1.000, 0.000)$$

$$(0.269, 0.001, 1.000, 0.000)W_2 = (0.269, 0.001, 1.000, 0.000)\begin{bmatrix} -2 & 1 \\ -1 & 0 \\ 1 & -3 \\ 5 & -1 \end{bmatrix} = (0.461, -2.731)$$

$$(0.461, -2.731) + \boldsymbol{b}_2 = (0.461, -2.731) + (-2, 2) = (-1.539, -0.731).$$

$$\sigma(-1.539, -0.731) = (0.177, 0.325).$$

(b) (i)

$$(1, 0, 2, -3)W_1 = (1, 0, 2, -3)\begin{bmatrix} 4 & -5 & 0 & 1 \\ -3 & 6 & -1 & 2 \\ 0 & 1 & 1 & -2 \\ 2 & 0 & -3 & 4 \end{bmatrix} = (-2, -3, 11, -15)$$

$$(-2, -3, 11, -15) + \boldsymbol{b}_1 = (-2, -3, 11, -15) + (1, -2, 0, -1) = (-1, -5, 11, -16).$$

$$relu(-1, -5, 11, -16) = (0, 0, 11, 0)$$

$$(0, 0, 11, 0)W_2 = (0, 0, 11, 0)\begin{bmatrix} -2 & 1 \\ -1 & 0 \\ 1 & -3 \\ 5 & -1 \end{bmatrix} = (11, -33)$$

$$(11, -33) + \boldsymbol{b}_2 = (11, -33) + (-2, 2) = (9, -31).$$

$$relu(9, -31) = (9, 0).$$

(c) (i)

$$(1, 0, 2, -3)W_1 = (1, 0, 2, -3)\begin{bmatrix} 4 & -5 & 0 & 1 \\ -3 & 6 & -1 & 2 \\ 0 & 1 & 1 & -2 \\ 2 & 0 & -3 & 4 \end{bmatrix} = (-2, -3, 11, -15)$$

$$(-2, -3, 11, -15) + \boldsymbol{b}_1 = (-2, -3, 11, -15) + (1, -2, 0, -1) = (-1, -5, 11, -16).$$

$$relu(-1, -5, 11, -16) = (0, 0, 11, 0)$$

$$(0,0,11,0)W_2 = (0,0,11,0)\begin{bmatrix} -2 & 1 \\ -1 & 0 \\ 1 & -3 \\ 5 & -1 \end{bmatrix} = (11,-33)$$

$$(11,-33) + \boldsymbol{b}_2 = (11,-33) + (-2,2) = (9,-31).$$

$$\sigma(9,-31) = (1,0).$$

(2) Consider a network with 2 input nodes, 3 nodes in the first hidden layer, 3 nodes in the second hidden layer and 2 output nodes. The weights between the respective layers are given by the following matrices:

$$W_1 = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 4 & -1 \end{bmatrix} \quad W_2 = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & -3 \\ 5 & 0 & 0 \end{bmatrix} \quad W_3 = \begin{bmatrix} 0 & 2 \\ -1 & -1 \\ 4 & -3 \end{bmatrix}$$

The bias values at the two hidden layers and output layer are given by:

$$\boldsymbol{b}_1 = (0, -3, 2), \quad \boldsymbol{b}_2 = (3, 3, -1), \quad \boldsymbol{b}_3 = (0, 1).$$

(a) Using the *relu* activation function at each node, calculate the output from the network for the following inputs:

(i) $(1, -1)$

(ii) $(3, -2)$

(b) Using the sigmoid activation function σ at each node, calculate the output from the network with same inputs as above.

Solutions:

(a) (i)

$$(1, -1)W_1 = (1, -1) \begin{bmatrix} 1 & -1 & 0 \\ -2 & 4 & -1 \end{bmatrix} = (3, -5, 1)$$

$$(3, -5, 1) + \boldsymbol{b}_1 = (3, -5, 1) + (0, -3, 2) = (3, -8, 3)$$

$$relu(3, -8, 3) = (3, 0, 3)$$

$$(3, 0, 3)W_2 = (3, 0, 3) \begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & -3 \\ 5 & 0 & 0 \end{bmatrix} = (21, 0, -3)$$

$$(21, 0, -3) + \boldsymbol{b}_2 = (21, 0, 3) + (3, 3, -1) = (24, 3, 2)$$

$$relu(24, 3, 2) = (24, 3, 2)$$

$$(24, 3, 2)W_3 = (24, 3, 2) \begin{bmatrix} 0 & 2 \\ -1 & -1 \\ 4 & -3 \end{bmatrix} = (5, -9)$$

$$(5, -9) + \boldsymbol{b}_3 = (5, -9) + (0, 1) = (5, -8)$$

$$relu(5, -8) = (5, 0).$$

(b) (i)

$$(1,-1)W_1 = (1,-1)\begin{bmatrix} 1 & -1 & 0 \\ -2 & 4 & -1 \end{bmatrix} = (3,-5,1)$$

$$(3,-5,1) + \boldsymbol{b}_1 = (3,-5,1) + (0,-3,2) = (3,-8,3)$$

$$\sigma(3,-8,3) = (0.953, 0.000, 0.953)$$

$$(0.953, 0.000, 0.953)W_2 = (0.953, 0.000, 0.953)\begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & -3 \\ 5 & 0 & 0 \end{bmatrix} = (6.671, 0.000, -0.953)$$

$$(6.671, 0.000, -0.953) + \boldsymbol{b}_2 = (6.671, 0.000, -0.953) + (3,3,-1) = (9.671, 3.000, -1.953)$$

$$\sigma(9.671, 3.000, -1.953) = (1.000, 0.953, 0.124)$$

$$(1.000, 0.953, 0.124)W_3 = (1.000, 0.953, 0.124)\begin{bmatrix} 0 & 2 \\ -1 & -1 \\ 4 & -3 \end{bmatrix} = (-0.457, 0.675)$$

$$(-0.457, 0.675) + \boldsymbol{b}_3 = (-0.457, 0.675) + (0,1) = (-0.457, 1.675)$$

$$\sigma(-0.457, 1.675) = (0.388, 0.842)$$