The Algorithm of Backpropagation Neural Network

September 19, 2013

Recall of perceptron algorithm

Generate output using the perceptron

Find the error

Use the error to update

The Structure of the ANN

A neural network is a network of multiple neurons

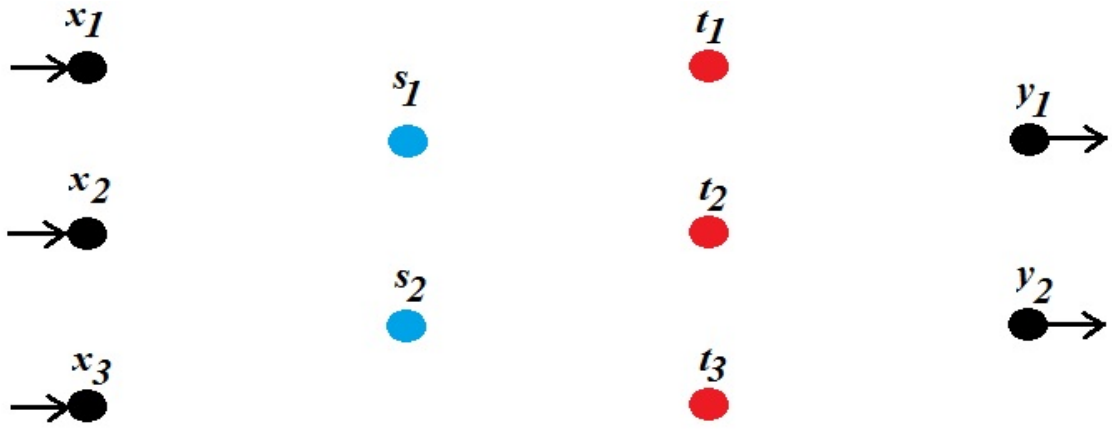
There are 3 layer, the input layer (denote by x) , the output layer (denote by y) and the hidden layer

There will be only one layer of input and output, but can have multiple layer of hidden layer

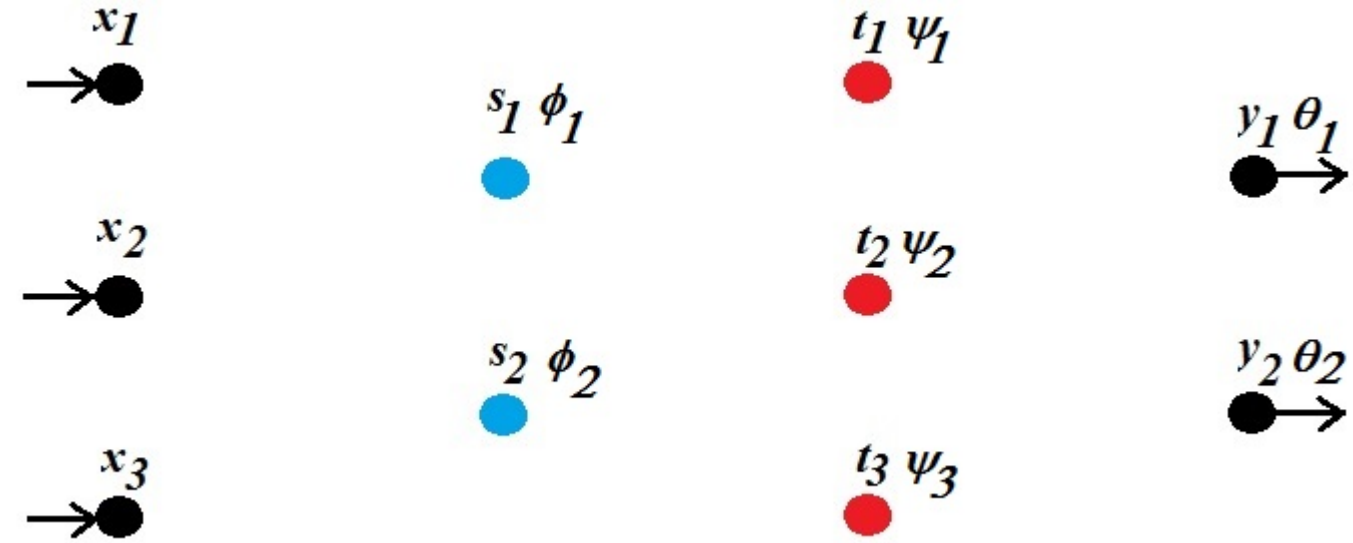
The number of neurons in input and output layer are problem-specific, but there is no general rule on number of layers and neurons in hidden layer.

Consider an example of a 2 hidden layer neural network

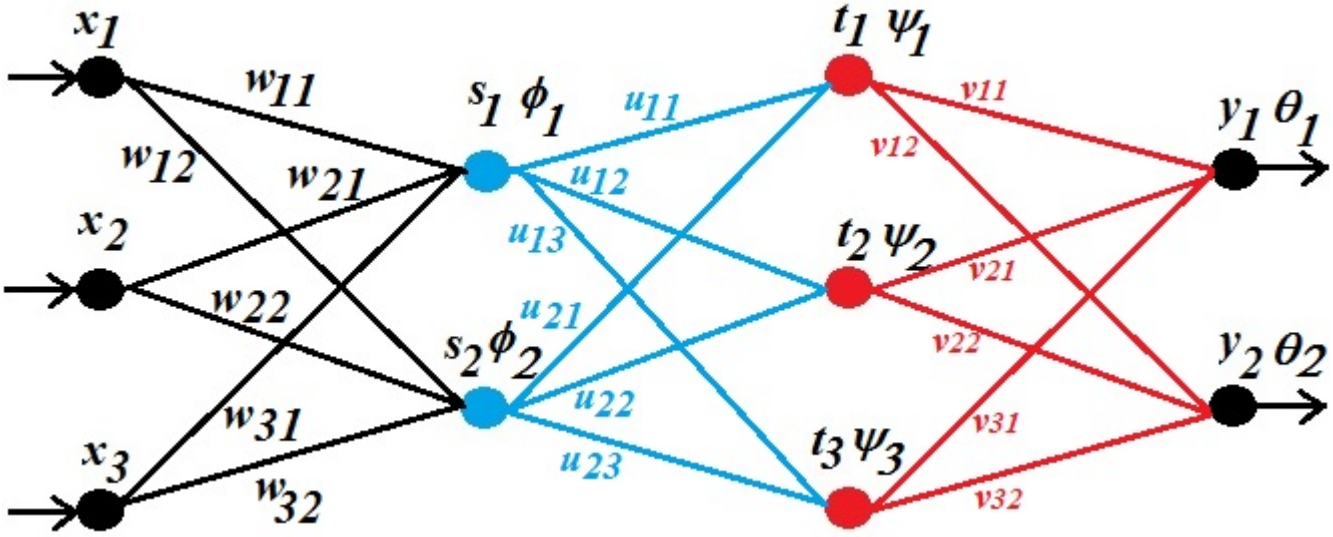
There are 3 input in the input layer x_1, x_2 , x_3
There are 2 output in the output layer y_1 , y_2
There are 2 hidden layer s and t that 2 neuron in the first hidden layer s_1 , s_2 and 3 for the second hidden layer t_1 , t_2 , t_3



Recall that a node is actually a perceptron, thus all the nodes (except input neuron) have threshold



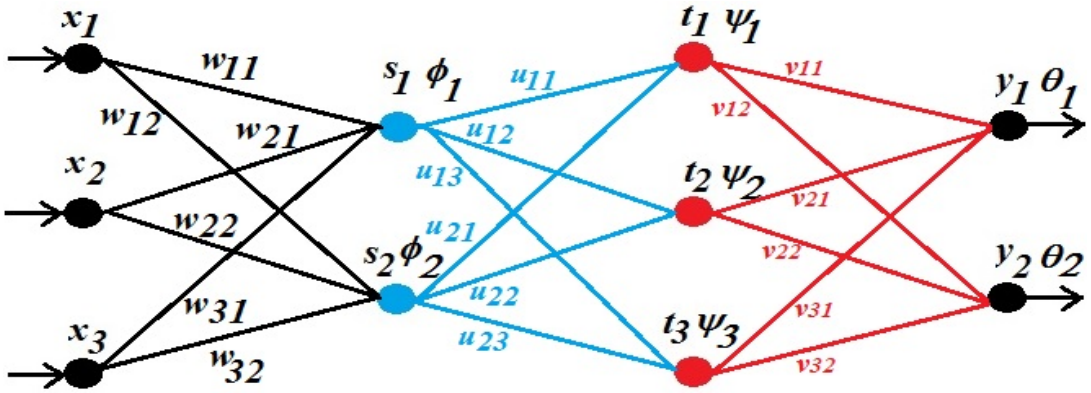
All the nodes are connected using weighting



Then we need to select the activation function of the layer, for example, *sigmoid* for s , *sigmoid* for t and linear for y

And the values for updating

w_{11}	ϕ_1	u_{11}	ψ_1	v_{11}	θ_1
w_{12}	ϕ_2	u_{12}	ψ_2	v_{12}	θ_2
w_{21}		u_{13}	ψ_3	v_{21}	
w_{22}		u_{21}		v_{22}	
w_{31}		u_{22}		v_{31}	
w_{32}		u_{23}		v_{32}	



The node input output equations

$$s_1 = \text{sigmoid}\left(x_1 w_{11} + x_2 w_{21} + x_3 w_{31} + \phi_1\right) = \frac{1}{1 + e^{-(x_1 w_{11} + x_2 w_{21} + x_3 w_{31} + \phi_1)}}$$

$$s_2 = \text{sigmoid}\left(x_1 w_{12} + x_2 w_{22} + x_3 w_{32} + \phi_2\right) = \frac{1}{1 + e^{-(x_1 w_{12} + x_2 w_{22} + x_3 w_{32} + \phi_2)}}$$

$$t_1 = \text{sigmoid}\left(s_1 u_{11} + s_2 u_{21} + \psi_1\right) = \frac{1}{1 + e^{-(s_1 u_{11} + s_2 u_{21} + \psi_1)}}$$

$$t_2 = \text{sigmoid}\left(s_1 u_{12} + s_2 u_{22} + \psi_2\right) = \frac{1}{1 + e^{-(s_1 u_{12} + s_2 u_{22} + \psi_2)}}$$

$$t_3 = \text{sigmoid}\left(s_1 u_{13} + s_2 u_{23} + \psi_3\right) = \frac{1}{1 + e^{-(s_1 u_{13} + s_2 u_{23} + \psi_3)}}$$

$$y_1 = \text{linear}\left(t_1 v_{11} + t_2 v_{21} + t_3 v_{31} + \theta_1\right) = t_1 v_{11} + t_2 v_{21} + t_3 v_{31} + \theta_1$$

$$y_2 = \text{linear}\left(t_1 v_{12} + t_2 v_{22} + t_3 v_{32} + \theta_2\right) = t_1 v_{12} + t_2 v_{22} + t_3 v_{32} + \theta_2$$

The Error Functions (Backward propagation)

The error of the output layer

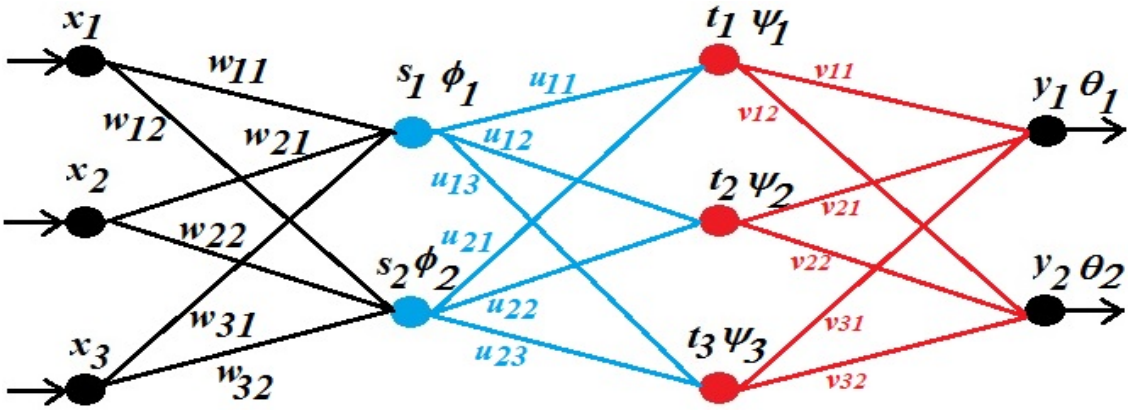
$$e_1^{output} = y_1^d - y_1$$
$$e_2^{output} = y_2^d - y_2$$

The error of the second hidden layer

$$e_1^{2hid} = t_1(1 - t_1)\left(v_{11} e_1^{output} + v_{12} e_2^{output}\right)$$
$$e_2^{2hid} = t_2(1 - t_2)\left(v_{21} e_1^{output} + v_{22} e_2^{output}\right)$$
$$e_3^{2hid} = t_3(1 - t_3)\left(v_{31} e_1^{output} + v_{32} e_2^{output}\right)$$

The error of the first hidden layer

$$e_1^{1hid} = s_1(1 - s_1)\left(u_{11} e_1^{2hid} + u_{12} e_2^{2hid} + u_{13} e_3^{2hid}\right)$$
$$e_2^{1hid} = s_2(1 - s_2)\left(u_{21} e_1^{2hid} + u_{22} e_2^{2hid} + u_{23} e_3^{2hid}\right)$$



Update the weight

General Idea

New Weight = Old Weight
+ learning rate $\times$ input linked to that node $\times$ error linked to that node

The 2nd-hidden layer weighting update

$$v_{11}^{new} = v_{11}^{old} + \alpha t_1 e_1^{output}$$
$$v_{12}^{new} = w_{12}^{old} + \alpha t_1 e_2^{output}$$
$$v_{21}^{new} = w_{21}^{old} + \alpha t_2 e_1^{output}$$
$$w_{21}^{new} = w_{22}^{old} + \alpha t_2 e_2^{output}$$
$$w_{31}^{new} = w_{31}^{old} + \alpha t_3 e_1^{output}$$
$$w_{32}^{new} = w_{31}^{old} + \alpha t_3 e_2^{output}$$

The output layer threshold update

$$\theta_1^{New} = \theta_1^{Old} + \alpha e_1^{output}$$
$$\theta_2^{New} = \theta_2^{Old} + \alpha e_2^{output}$$

The 1st-hidden layer weighting update

$$u_{11}^{new} = u_{11}^{old} + \alpha s_1 e_1^{2hid}$$
$$u_{12}^{new} = u_{12}^{old} + \alpha s_1 e_2^{2hid}$$
$$u_{13}^{new} = u_{13}^{old} + \alpha s_1 e_3^{2hid}$$
$$u_{21}^{new} = u_{21}^{old} + \alpha s_2 e_1^{2hid}$$
$$u_{22}^{new} = u_{22}^{old} + \alpha s_2 e_2^{2hid}$$
$$u_{23}^{new} = u_{23}^{old} + \alpha s_2 e_3^{2hid}$$

The 2nd-hidden layer threshold update

$$\psi_1^{New} = \psi_1^{Old} + \alpha e_1^{2hid}$$
$$\psi_2^{New} = \psi_2^{Old} + \alpha e_2^{2hid}$$
$$\psi_3^{New} = \psi_3^{Old} + \alpha e_2^{2hid}$$

The input layer weighting update

$$w_{11}^{new} = w_{11}^{old} + \alpha x_1 e_1^{1hid}$$
$$w_{12}^{new} = w_{12}^{old} + \alpha x_1 e_2^{1hid}$$
$$w_{21}^{new} = w_{21}^{old} + \alpha x_2 e_1^{1hid}$$
$$w_{22}^{new} = w_{22}^{old} + \alpha x_2 e_2^{1hid}$$
$$w_{31}^{new} = w_{31}^{old} + \alpha x_3 e_1^{1hid}$$
$$w_{32}^{new} = w_{32}^{old} + \alpha x_3 e_2^{1hid}$$

The 1st-hidden layer threshold update

$$\phi_1^{New} = \phi_1^{Old} + \alpha e_1^{1hid}$$
$$\phi_2^{New} = \phi_2^{Old} + \alpha e_2^{1hid}$$

The equations above are for *ONE LOOP* a set of input and desired output
The eqautions above will be used N time for a training set with N input-output pairs.

−END−