

# Mult-layer perceptron

Back Propagation

# Steps of a **single epoch**

For each pattern

- Forward prop
  - Calculate  $net_j$  and  $o_j$  for all neurons (except input layer and bias neurons)
  - Calculate specific error (for single pattern)
- Back prop
  - Calculate  $\delta_j$  for all neurons (except input layer and bias neurons)
  - Calculate  $\Delta w_{i,j}$  for all variable weights including bias weights
  - $w_{i,j} := w_{i,j} + \Delta w_{i,j}$

# After end of epoch

- Calculate total error = sum of specific errors
- Check stopping condition
- Run another epoch if stopping condition is False

# Notes

Assuming layer K is before layer H, and either layer L is after layer H, or layer H is the output layer

- $net_h = \sum_{k \in K} w_{k,h} o_k$
- $o_h = f_{act}(net_h)$       Default is  $\frac{1}{1+e^{-net}}$
- $\delta_h =$ 
  - $f'_{act}(net_h) \cdot \sum_{l \in L} \delta_l w_{h,l}$       if h is hidden neuron
    - Default is:  $o_h(1 - o_h) \sum_{l \in L} \delta_l w_{h,l}$       for sigmoid activation function
  - $f'_{act}(net_h) \cdot -\frac{\partial Err_p}{\partial y_h}$       if h is output neuron
    - Default is:  $y_h(1 - y_h)(t_h - y_h)$       for sigmoid activation function  
and for  $Err_p = \frac{1}{2} \sum_{h \in H} (t_h - y_h)^2$

Warning! if different activation function or different error function is used, you must calculate the derivatives  $f'_{act}(net_h) \cdot -\frac{\partial Err_p}{\partial y_h}$

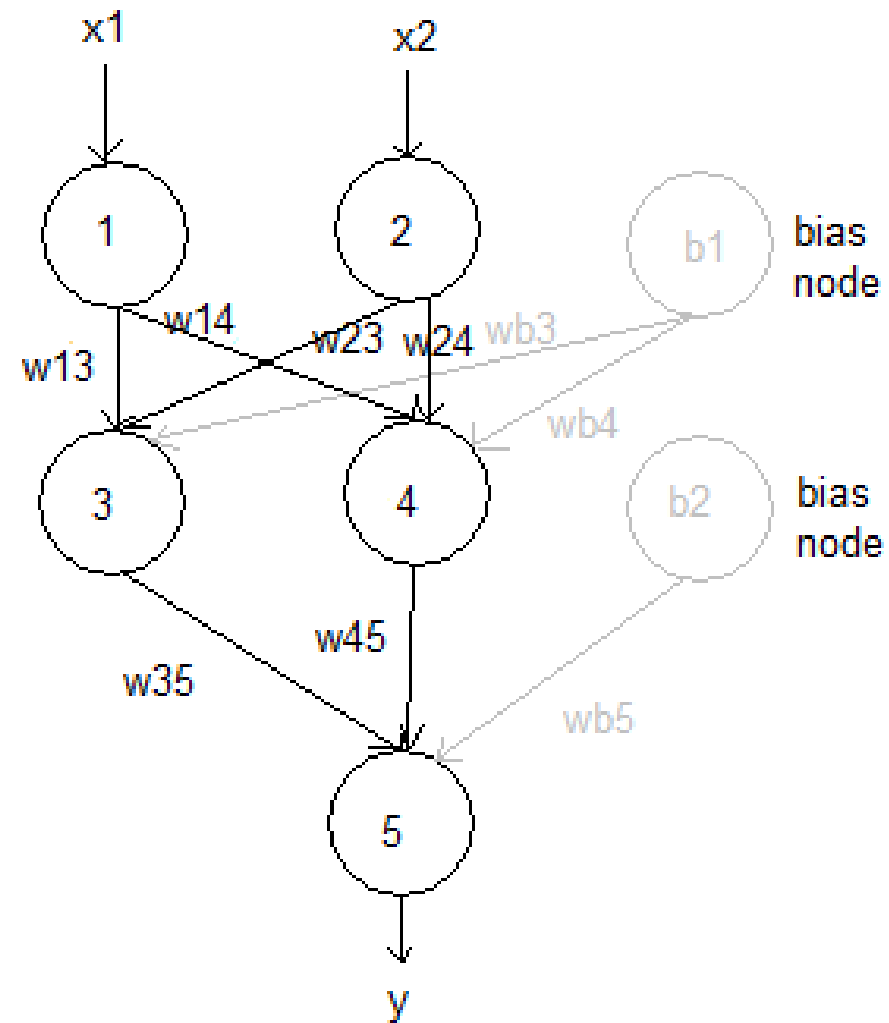
# Notes

- $\Delta w_{i,j} = \eta o_i \delta_j$
- $w_{i,j} := w_{i,j} + \eta o_i \delta_j$ 
  - If  $i$  is input neuron:  $o_i = x_i$
  - If  $i$  is bias neuron:  $o_i = 1$

# Example

x1	x2	t
0	0	0
0	1	1
1	0	1
1	1	0

Assume learning rate = 0.3



Epoch: 1

Pattern: 1:  $x_1 = 0, x_2 = 0, t = 0$

- Initial weights:

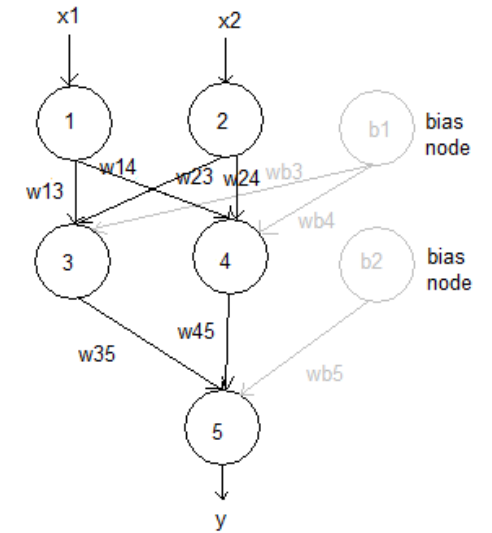
$w_{13} = 0.3$	$w_{23} = -0.1$	$w_{b3} = 0.2$
$w_{14} = -0.2$	$w_{24} = 0.2$	$w_{b4} = -0.3$
$w_{35} = 0.4$	$w_{45} = -0.2$	$w_{b5} = 0.4$

- Forward prop:

- $\text{net}_3 = w_{13} * x_1 + w_{23} * x_2 + w_{b3} = 0.3 * 0 - 0.1 * 0 + 0.2 = 0.2$
- $o_3 = 1/(1 + e^{-\text{net}_3}) = 1/(1 + e^{-0.2}) = 0.5498$
- $\text{net}_4 = w_{14} * x_1 + w_{24} * x_2 + w_{b4} = -0.2 * 0 + 0.2 * 0 - 0.3 = -0.3$
- $o_4 = 1/(1 + e^{-\text{net}_4}) = 1/(1 + e^{0.3}) = 0.4256$
- $\text{net}_5 = w_{35} * o_3 + w_{45} * o_4 + w_{b5} = 0.4 * 0.5498 - 0.2 * 0.4256 + 0.4 = 0.5348$
- $y = 1/(1 + e^{-\text{net}_5}) = 1/(1 + e^{-0.5348}) = 0.6306$

- Calculating error:

- $\text{Err}_{p1} = 0.5 * (0 - 0.6306)^2 = 0.1988$



Epoch: 1

Pattern: 1:  $x1 = 0, x2 = 0, t = 0$

Back Prop:

1) Finding delta

$$\delta_5 = y * (1 - y) * (t - y) = 0.6306 * (1 - 0.6306) * (0 - 0.6306) = -0.1469$$

$$\delta_3 = o_3(1 - o_3) * \delta_5 * w_{35} = 0.5498 * (1 - 0.5498) * -0.1469 * 0.4 = -0.0145$$

$$\delta_4 = o_4(1 - o_4) * \delta_5 * w_{45} = 0.4256 * (1 - 0.4256) * -0.1469 * -0.2 = 0.0072$$

2) Finding new weights

$$w_{35} := w_{35} + \eta * o_3 * \delta_5 = 0.4 + 0.3 * 0.5498 * -0.1469 = 0.3758$$

$$w_{45} := w_{45} + \eta * o_4 * \delta_5 = -0.2 + 0.3 * 0.4256 * -0.1469 = -0.2188$$

$$wb_5 := wb_5 + \eta * 1 * \delta_5 = 0.4 + 0.3 * 1 * -0.1469 = 0.3559$$

$$w_{14} := w_{14} + \eta * x_1 * \delta_4 = -0.2 + 0.3 * 0 * 0.0072 = -0.2$$

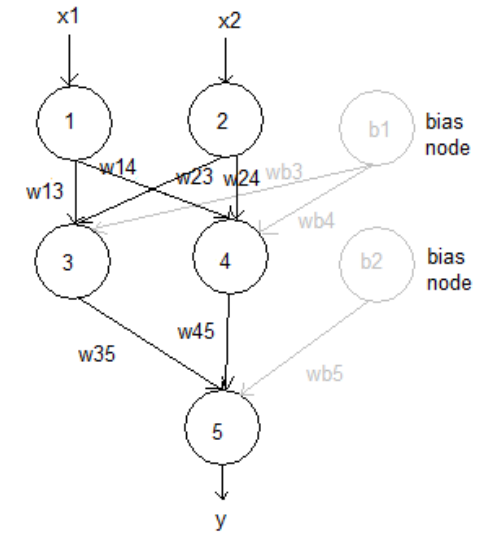
$$w_{24} := w_{24} + \eta * x_2 * \delta_4 = 0.2 + 0.3 * 0 * 0.0072 = 0.2$$

$$wb_4 := wb_4 + \eta * 1 * \delta_4 = -0.3 + 0.3 * 1 * 0.0072 = -0.2978$$

$$w_{13} := w_{13} + \eta * x_1 * \delta_3 = 0.3 + 0.3 * 0 * -0.0145 = 0.3$$

$$w_{23} := w_{23} + \eta * x_2 * \delta_3 = -0.1 + 0.3 * 0 * -0.0145 = -0.1$$

$$wb_3 := wb_3 + \eta * 1 * \delta_3 = 0.2 + 0.3 * 1 * -0.0145 = 0.1957$$



Epoch: 1

Pattern: 2:  $x_1 = 0, x_2 = 1, t = 1$

- weights:

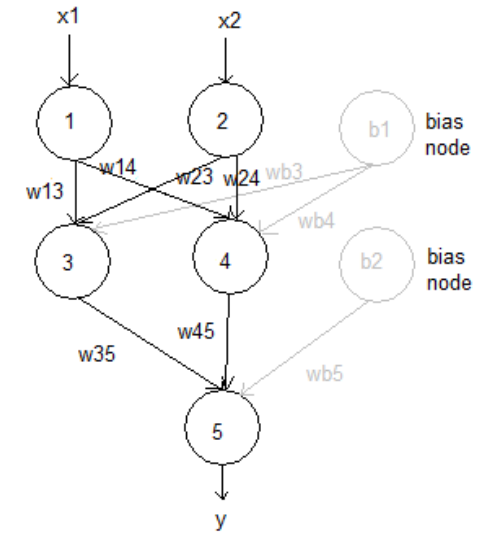
$w_{13} = 0.3$	$w_{23} = -0.1$	$w_{b3} = 0.1957$
$w_{14} = -0.2$	$w_{24} = 0.2$	$w_{b4} = -0.2978$
$w_{35} = 0.3758$	$w_{45} = -0.2188$	$w_{b5} = 0.3559$

- Forward prop:

- $\text{net}_3 = w_{13} * x_1 + w_{23} * x_2 + w_{b3} = \dots$
- $o_3 = 1 / (1 + e^{-\text{net}_3}) = \dots$
- $\text{net}_4 = w_{14} * x_1 + w_{24} * x_2 + w_{b4} = \dots$
- $o_4 = 1 / (1 + e^{-\text{net}_4}) = \dots$
- $\text{net}_5 = w_{35} * o_3 + w_{45} * o_4 + w_{b5} = \dots$
- $y = 1 / (1 + e^{-\text{net}_5}) = \dots$

- Calculating error:

- $\text{Err}_{p2} = 0.5 * (t - y)^2 = \dots$



Epoch: 1

Pattern: 2:       $x1 = 0, x2 = 1, t = 1$

Back Prop:

1) Finding delta

$$\delta_5 = y * (1 - y) * (t - y) = \dots$$

$$\delta_3 = o_3(1 - o_3) * \delta_5 * w_{35} = \dots$$

$$\delta_4 = o_4(1 - o_4) * \delta_5 * w_{45} = \dots$$

2) Finding new weights

$$w_{35} := w_{35} + \eta * o_3 * \delta_5 = \dots$$

$$w_{45} := w_{45} + \eta * o_4 * \delta_5 = \dots$$

$$wb_5 := wb_5 + \eta * 1 * \delta_5 = \dots$$

$$w_{14} := w_{14} + \eta * x_1 * \delta_4 = \dots$$

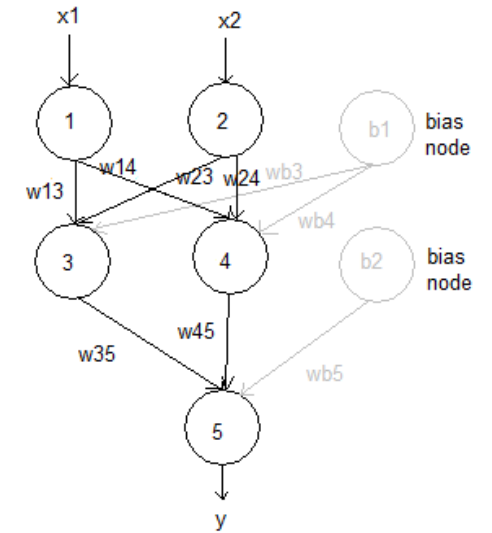
$$w_{24} := w_{24} + \eta * x_2 * \delta_4 = \dots$$

$$wb_4 := wb_4 + \eta * 1 * \delta_4 = \dots$$

$$w_{13} := w_{13} + \eta * x_1 * \delta_3 = \dots$$

$$w_{23} := w_{23} + \eta * x_2 * \delta_3 = \dots$$

$$wb_3 := wb_3 + \eta * 1 * \delta_3 = \dots$$



Epoch: 1

Pattern: 3:  $x_1 = 1, x_2 = 0, t = 1$

- weights:

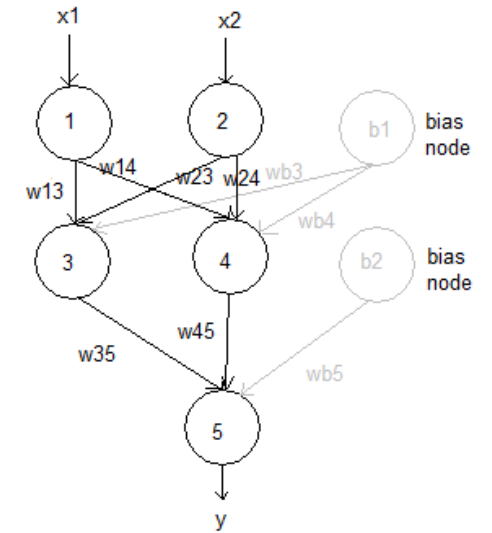
$w_{13} = ?$	$w_{23} = ?$	$w_{b3} = ?$
$w_{14} = ?$	$w_{24} = ?$	$w_{b4} = ?$
$w_{35} = ?$	$w_{45} = ?$	$w_{b5} = ?$

- Forward prop:

- $\text{net}_3 = w_{13} * x_1 + w_{23} * x_2 + w_{b3} = \dots$
- $o_3 = 1/(1 + e^{-\text{net}_3}) = \dots$
- $\text{net}_4 = w_{14} * x_1 + w_{24} * x_2 + w_{b4} = \dots$
- $o_4 = 1/(1 + e^{-\text{net}_4}) = \dots$
- $\text{net}_5 = w_{35} * o_3 + w_{45} * o_4 + w_{b5} = \dots$
- $y = 1/(1 + e^{-\text{net}_5}) = \dots$

- Calculating error:

- $\text{Err}_{p3} = 0.5 * (t - y)^2 = \dots$



Epoch: 1

Pattern: 3:  $x_1 = 1, x_2 = 0, t = 1$

Back Prop:

1) Finding delta

$$\delta_5 = y * (1 - y) * (t - y) = \dots$$

$$\delta_3 = o_3(1 - o_3) * \delta_5 * w_{35} = \dots$$

$$\delta_4 = o_4(1 - o_4) * \delta_5 * w_{45} = \dots$$

2) Finding new weights

$$w_{35} := w_{35} + \eta * o_3 * \delta_5 = \dots$$

$$w_{45} := w_{45} + \eta * o_4 * \delta_5 = \dots$$

$$w_{b5} := w_{b5} + \eta * 1 * \delta_5 = \dots$$

$$w_{14} := w_{14} + \eta * x_1 * \delta_4 = \dots$$

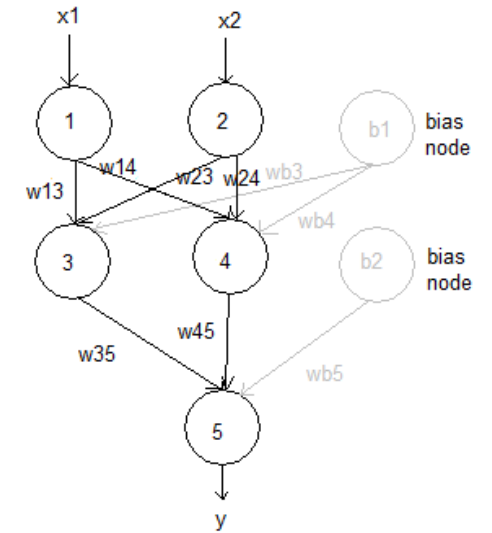
$$w_{24} := w_{24} + \eta * x_2 * \delta_4 = \dots$$

$$w_{b4} := w_{b4} + \eta * 1 * \delta_4 = \dots$$

$$w_{13} := w_{13} + \eta * x_1 * \delta_3 = \dots$$

$$w_{23} := w_{23} + \eta * x_2 * \delta_3 = \dots$$

$$w_{b3} := w_{b3} + \eta * 1 * \delta_3 = \dots$$



Epoch: 1

Pattern: 4:  $x_1 = 1, x_2 = 1, t = 0$

- weights:

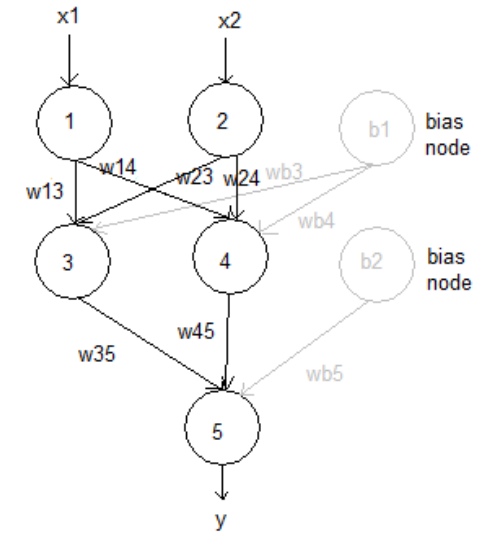
$w_{13} = ?$	$w_{23} = ?$	$w_{b3} = ?$
$w_{14} = ?$	$w_{24} = ?$	$w_{b4} = ?$
$w_{35} = ?$	$w_{45} = ?$	$w_{b5} = ?$

- Forward prop:

- $\text{net}_3 = w_{13} * x_1 + w_{23} * x_2 + w_{b3} = \dots$
- $o_3 = 1/(1 + e^{-\text{net}_3}) = \dots$
- $\text{net}_4 = w_{14} * x_1 + w_{24} * x_2 + w_{b4} = \dots$
- $o_4 = 1/(1 + e^{-\text{net}_4}) = \dots$
- $\text{net}_5 = w_{35} * o_3 + w_{45} * o_4 + w_{b5} = \dots$
- $y = 1/(1 + e^{-\text{net}_5}) = \dots$

- Calculating error:

- $\text{Err}_{p4} = 0.5 * (t - y)^2 = \dots$



Epoch: 1

Pattern: 4:  $x_1 = 1, x_2 = 1, t = 0$

Back Prop:

1) Finding delta

$$\delta_5 = y * (1 - y) * (t - y) = \dots$$

$$\delta_3 = o_3(1 - o_3) * \delta_5 * w_{35} = \dots$$

$$\delta_4 = o_4(1 - o_4) * \delta_5 * w_{45} = \dots$$

2) Finding new weights

$$w_{35} := w_{35} + \eta * o_3 * \delta_5 = \dots$$

$$w_{45} := w_{45} + \eta * o_4 * \delta_5 = \dots$$

$$wb_5 := wb_5 + \eta * 1 * \delta_5 = \dots$$

$$w_{14} := w_{14} + \eta * x_1 * \delta_4 = \dots$$

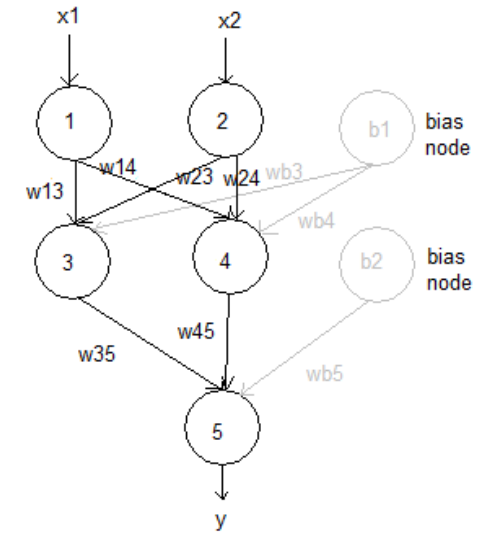
$$w_{24} := w_{24} + \eta * x_2 * \delta_4 = \dots$$

$$wb_4 := wb_4 + \eta * 1 * \delta_4 = \dots$$

$$w_{13} := w_{13} + \eta * x_1 * \delta_3 = \dots$$

$$w_{23} := w_{23} + \eta * x_2 * \delta_3 = \dots$$

$$wb_3 := wb_3 + \eta * 1 * \delta_3 = \dots$$



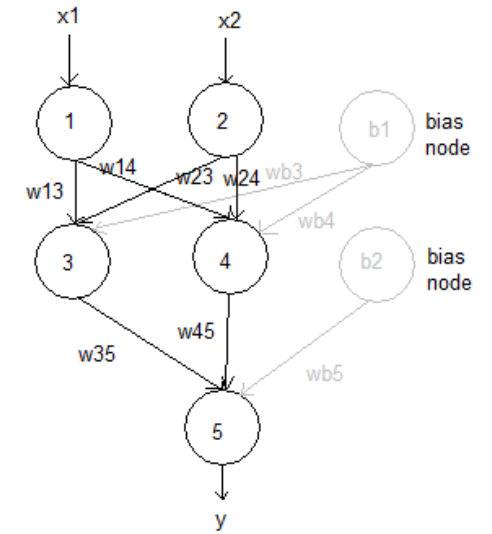
End of Epoch 1

Total error = Err\_p1 + Err\_p2 + Err\_p3 + Err\_p4 = 0.1988 + ...

If Total error <= tolerance (If given): Then stop training

If epoch number = max number of epochs (if given): Then stop training

Otherwise, run another epoch using last weights



<= Tolerance (0.022)

epoch 15