Feedforward

Y

[[1 0 0]
 [0 1 0]
 [0 0 1]]

O_{in}

[[0.49 -0.49 0.57]
 [0.43 -0.45 0.55]
 [0.37 -0.41 0.50]]

O_{out}

[[0.62 0.38 0.64]
 [0.61 0.39 0.63]
 [0.59 0.40 0.62]]

$W_{2\text{-update}}$

[[0.023 0.003 0.020]
 [0.013 0.003 0.019]]

W_2

[[0.0 -0.1 0.2]
 [0.6 -0.5 0.5]]

H_{in}

[[1.47 1.42]
 [3.42 0.91]
 [3.62 0.47]]

H_{out}

[[0.81 0.81]
 [0.97 0.71]
 [0.97 0.62]]

$W_{1\text{-update}}$

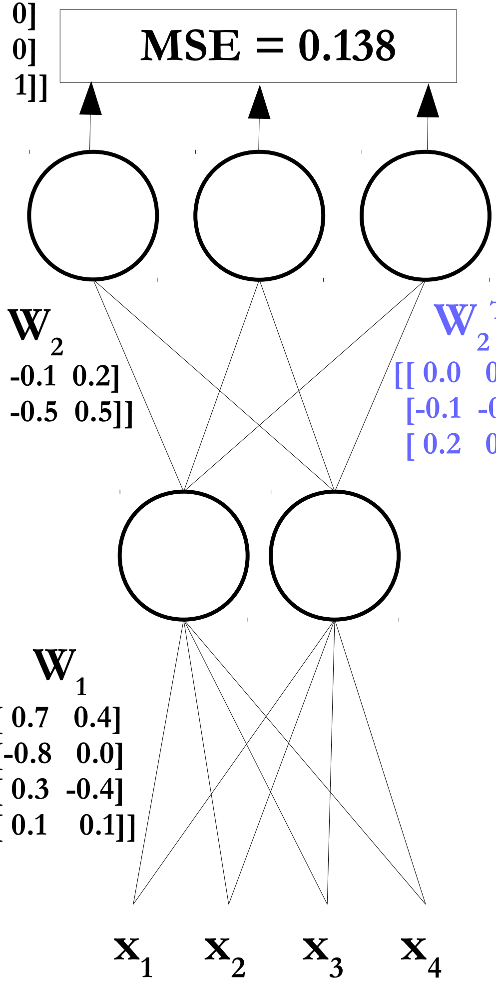
[[0.002 0.031]
 [0.001 0.015]
 [0.000 0.022]
 [0.000 0.007]]

W_1

[[0.7   0.4]
 [-0.8   0.0]
 [0.3 -0.4]
 [0.1   0.1]]

X

[[4.9 3.0 1.4 0.2]
 [6.4 3.2 4.5 1.5]
 [5.8 2.7 5.1 1.9]]



Backpropagation

H_{out}^T

[[0.81 0.97 0.97]
 [0.81 0.71 0.62]]

O_{delta}

[[-0.09   0.09 0.15]
 [0.15 -0.15   0.15]
 [0.14   0.10 -0.09]]

O_{error}

[[-0.38 0.38   0.64]
 [0.61 -0.61   0.63]
 [0.59 0.40 -0.38]]

W_2^T

[[0.0   0.6]
 [-0.1 -0.5]
 [0.2   0.5]]

X^T

[[4.9 6.4 5.8]
 [3.0 3.2 2.7]
 [1.4 4.5 5.1]
 [0.2 1.5 1.9]]

H_{delta}

[[0.003 -0.004]
 [0.001   0.049]
 [-0.001 -0.003]]

H_{error}

[[0.021 -0.024]
 [0.045   0.240]
 [-0.028 -0.011]]

$\frac{\partial MSE}{\partial W_2}(W_2)$:

$$O_{error} = O_{out} - Y$$
$$O_{delta} = O_{error} \odot O_{out} \odot (1 - O_{out})$$
$$W_{2\text{-update}} = \frac{1}{N}(H_{out}^T \cdot O_{delta})$$

$\frac{\partial MSE}{\partial W_1}(W_1)$:

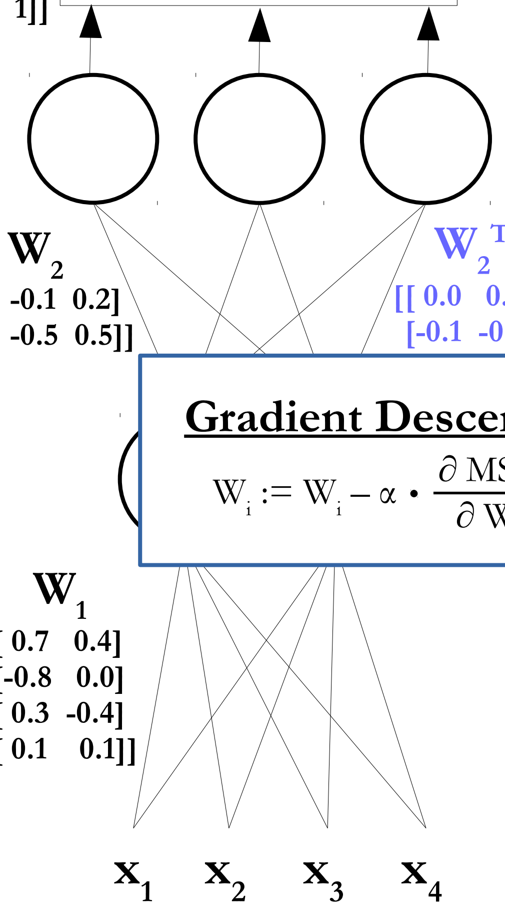
$$H_{error} = O_{delta} \cdot W_2^T$$
$$H_{delta} = H_{error} \odot H_{out} \odot (1 - H_{out})$$
$$W_{1\text{-update}} = \frac{1}{N}(X^T \cdot H_{delta})$$

Feedforward

Y

$$Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

MSE = 0.138



O_{in}

[[0.49 -0.49 0.57]
[0.43 -0.45 0.55]
[0.37 -0.41 0.50]]

O_{out}

[[0.62 0.38 0.64]
[0.61 0.39 0.63]
[0.59 0.40 0.62]]

$W_{2\text{-update}}$

[[0.023 0.003 0.020]
[0.013 0.003 0.019]]

W_2

[[0.0 -0.1 0.2]
[0.6 -0.5 0.5]]

H_{in}

[[1.47 1.42]
[3.42 0.91]
[3.62 0.47]]

H_{out}

[[0.81 0.81]
[0.97 0.71]
[0.97 0.62]]

$W_{1\text{-update}}$

[[0.002 0.031]
[0.001 0.015]
[0.000 0.022]
[0.000 0.007]]

W_1

[[0.7 0.4]
[-0.8 0.0]
[0.3 -0.4]
[0.1 0.1]]

X

[[4.9 3.0 1.4 0.2]
[6.4 3.2 4.5 1.5]
[5.8 2.7 5.1 1.9]]

Gradient Descent:

$$W_i := W_i - \alpha \cdot \frac{\partial\, MSE}{\partial\, W_i}(W_i)$$

Backpropagation

H_{out}^T

[[0.81 0.97 0.97]
[0.81 0.71 0.62]]

O_{delta}

[[-0.09 0.09 0.15]
[0.15 -0.15 0.15]
[0.14 0.10 -0.09]]

O_{error}

[[-0.38 0.38 0.64]
[0.61 -0.61 0.63]
[0.59 0.40 -0.38]]

W_2^T

[[0.0 0.6]
[-0.1 -0.5]]

X^T

.9 6.4 5.8]
.0 3.2 2.7]
.4 4.5 5.1]
.2 1.5 1.9]]

H_{delta}

[[0.003 -0.004]
[0.001 0.049]
[-0.001 -0.003]]

H_{error}

[[0.021 -0.024]
[0.045 0.240]
[-0.028 -0.011]]

$\frac{\partial MSE}{\partial W_2}(W_2)$:

$$O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \odot O_{out} \odot (1 - O_{out})$$

$$W_{2\text{-update}} = \frac{1}{N}(H_{out}^T \cdot O_{delta})$$

$\frac{\partial MSE}{\partial W_1}(W_1)$:

$$H_{error} = O_{delta} \cdot W_2^T$$

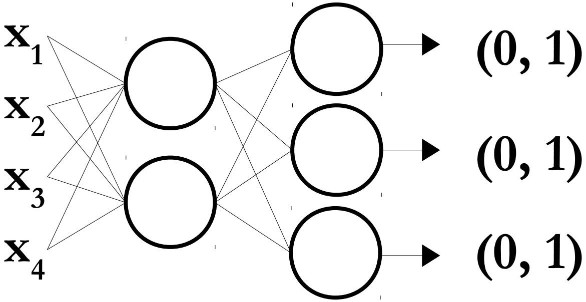
$$H_{delta} = H_{error} \odot H_{out} \odot (1 - H_{out})$$

$$W_{1\text{-update}} = \frac{1}{N}(X^T \cdot H_{delta})$$

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = \text{sigmoid}(H_{in})$$

$$O_{in} = H_{out}W_2$$

$$O_{out} = \text{sigmoid}(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

$$\text{Mean Squared Error} = \frac{1}{2N}\sum_{e}\sum_{n}(O_{out\,e,n} - Y_{e,n})^2$$

Gradient Descent:

$$W_i := W_i - \alpha \cdot \frac{\partial\, MSE}{\partial\, W_i}(W_i)$$

Chain Rule:

$$\frac{\partial Z}{\partial X} = \frac{\partial Z}{\partial Y} \cdot \frac{\partial Y}{\partial X}$$

Backpropagation

$$\frac{\partial\, MSE}{\partial\, W_1}(W_1) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, H_{out}} \frac{\partial\, H_{out}}{\partial\, H_{in}} \frac{\partial\, H_{in}}{\partial\, W_1}$$

$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, W_2}$$

N: e * n