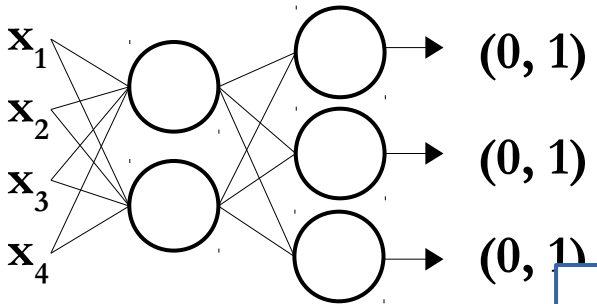
Deep Learning

How does the algorithm make a decision?

Feedforward



$$H_{in} = XW_1$$

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

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

$$W_2: \quad \frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_e (o_{out\,n}^{(e)} - y_n^{(e)}) \cdot o_{out\,n}^{(e)} (1 - o_{out\,n}^{(e)}) \cdot h_{out\,f}^{(e)}$$

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:

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

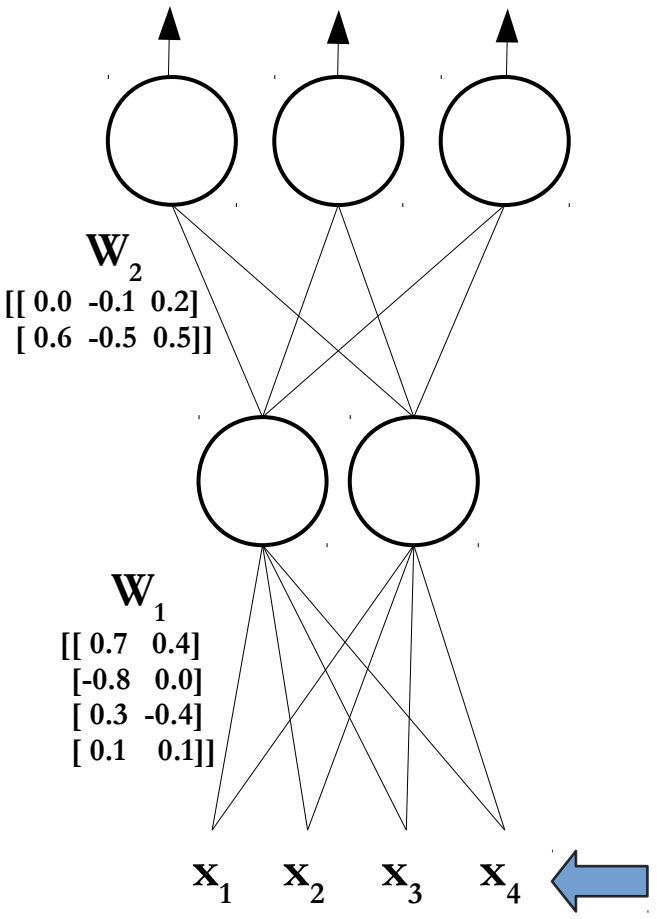
Backpropagation

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

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

N: e * n

Feedforward

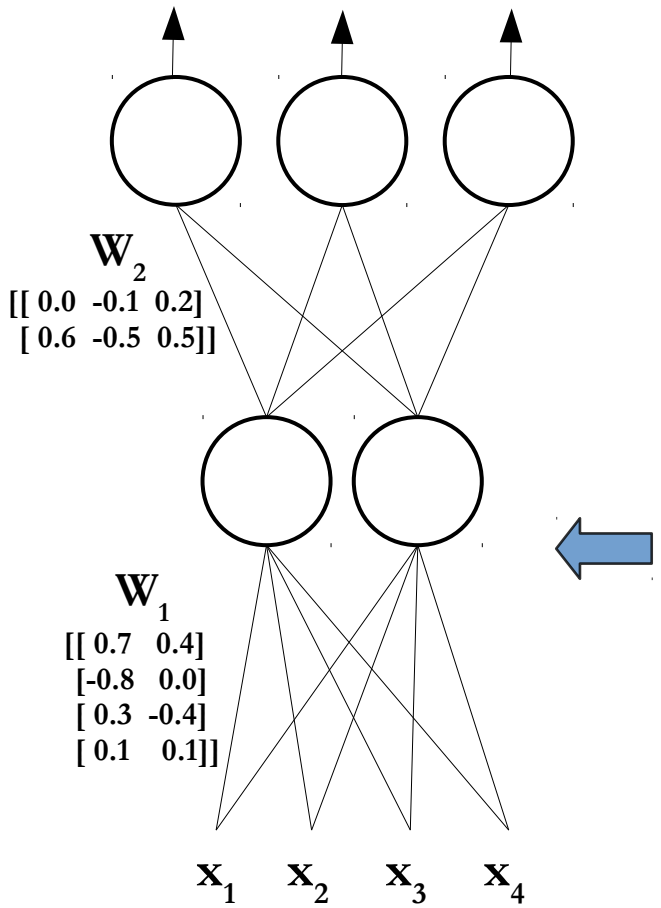


X

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

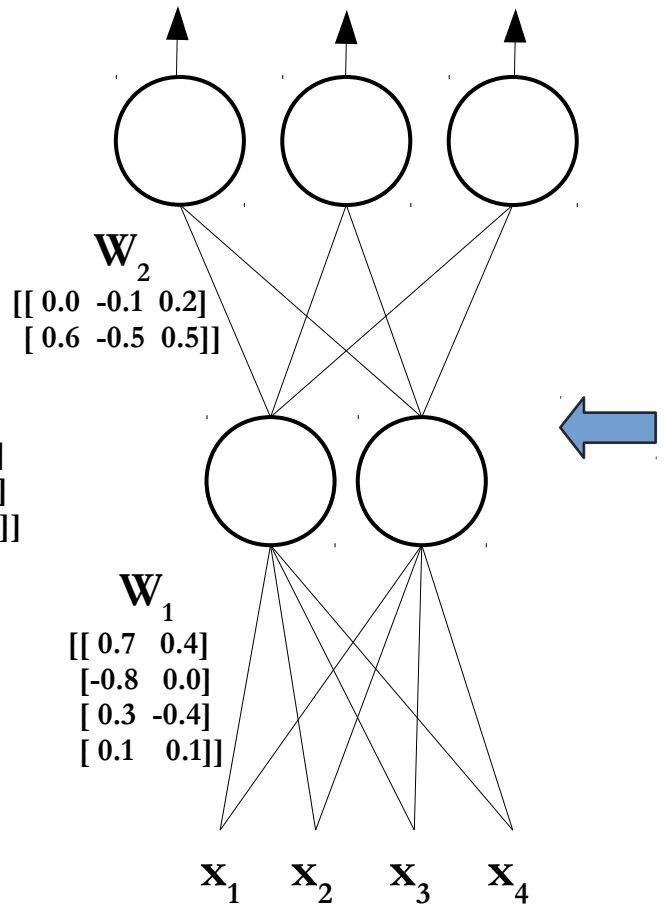
Feedforward

$$H_{in}$$
[[1.47 1.42]
 [3.42 0.91]
 [3.62 0.47]]



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

Feedforward



H_{in}
[[1.47 1.42]
[3.42 0.91]
[3.62 0.47]]

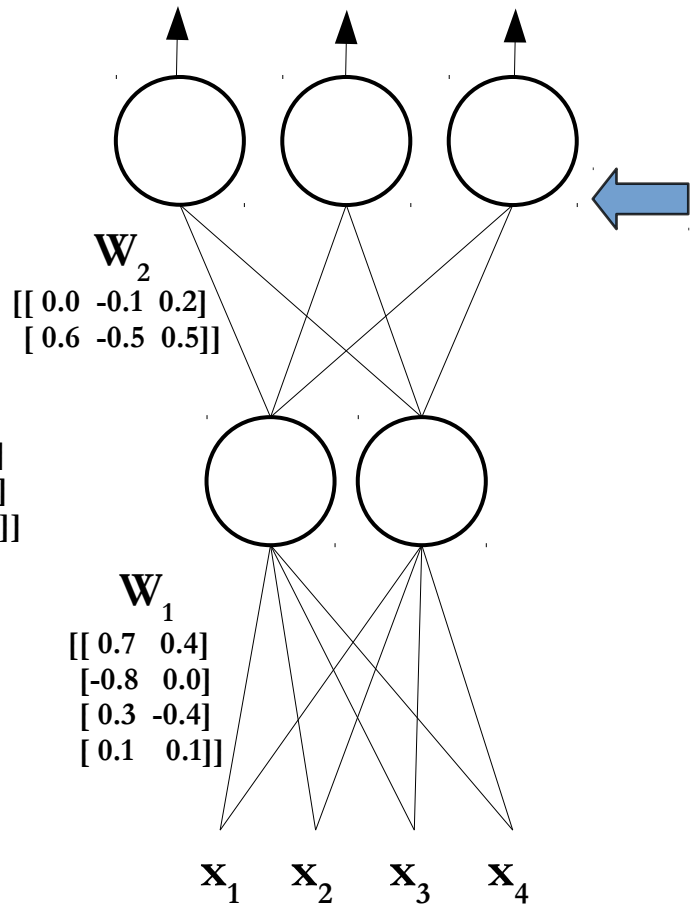
H_{out}
[[0.81 0.81]
[0.97 0.71]
[0.97 0.62]]

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]]

Feedforward

O_{in}

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



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]]
```

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]]
```

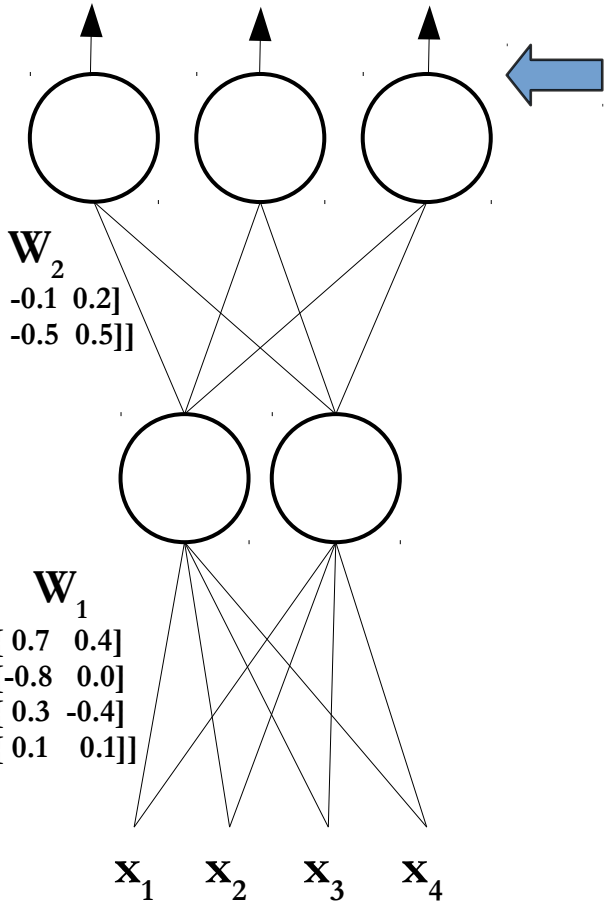
Feedforward

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]]



H_{in}

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

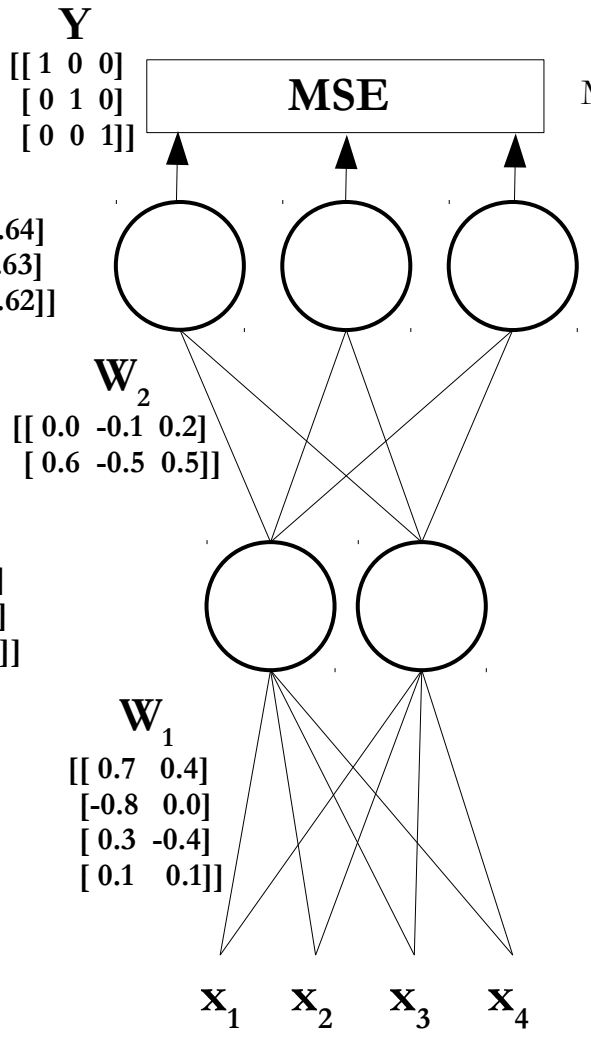
H_{out}

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

X

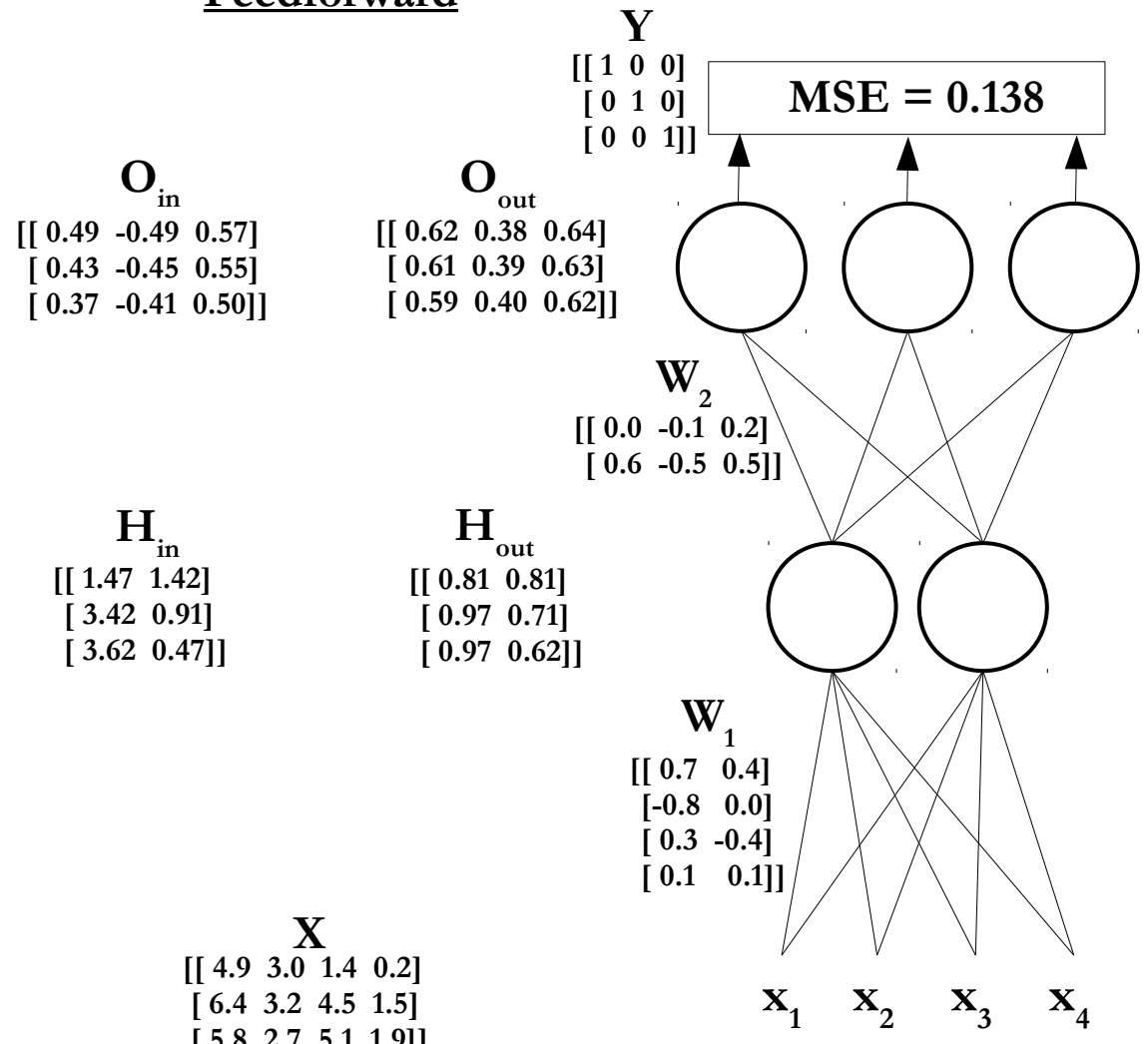
[[4.9 3.0 1.4 0.2]
[6.4 3.2 4.5 1.5]
[5.8 2.7 5.1 1.9]]

Feedforward



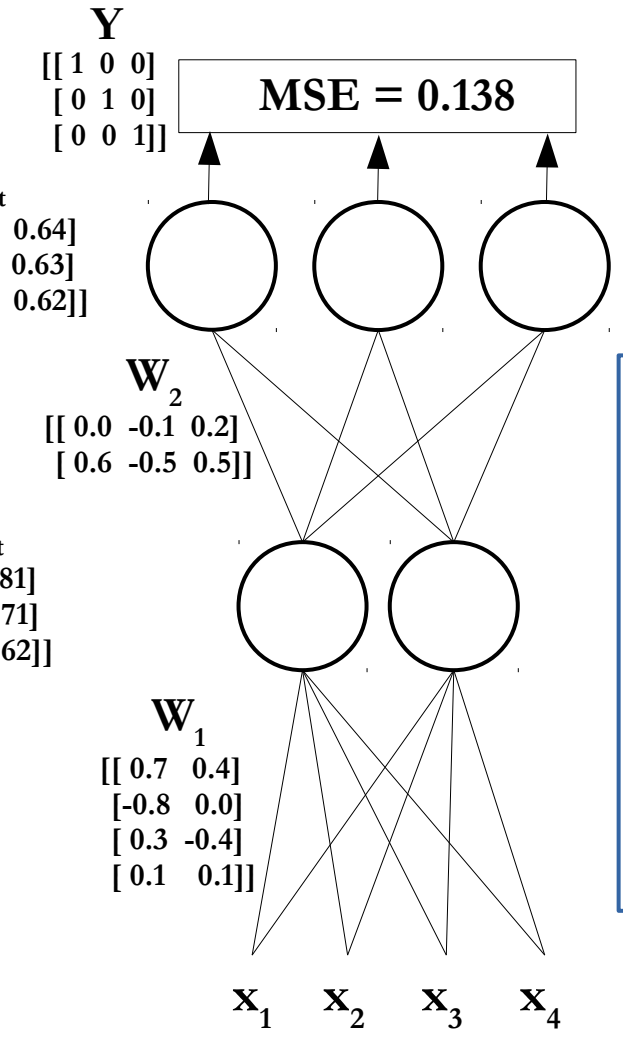
$$\text{Mean Squared Error} = \frac{1}{2N} \sum_e \sum_n (O_{\text{out } e,n} - Y_{e,n})^2$$

Feedforward



Feedforward

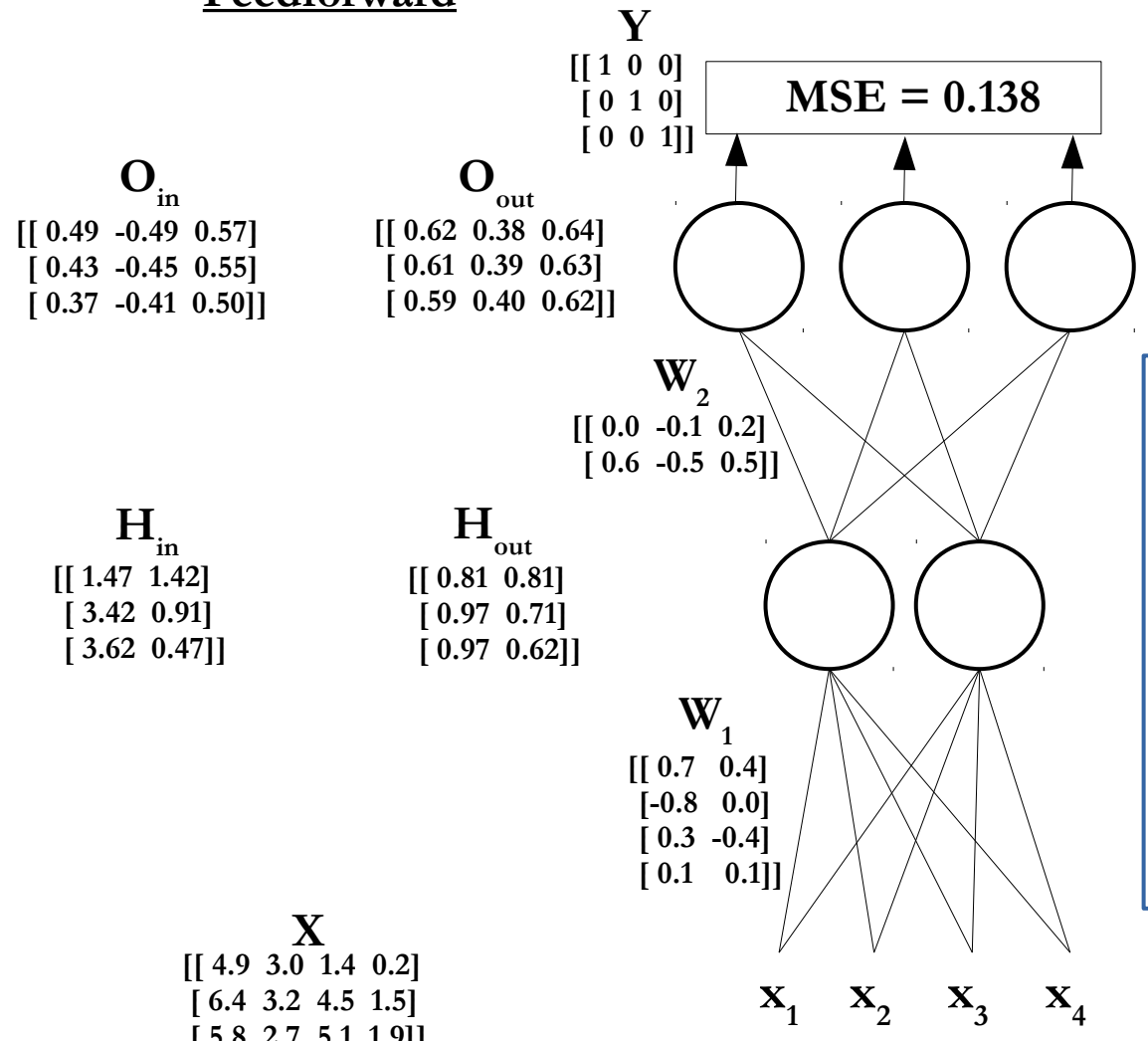
Backpropagation



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

Feedforward

Backpropagation

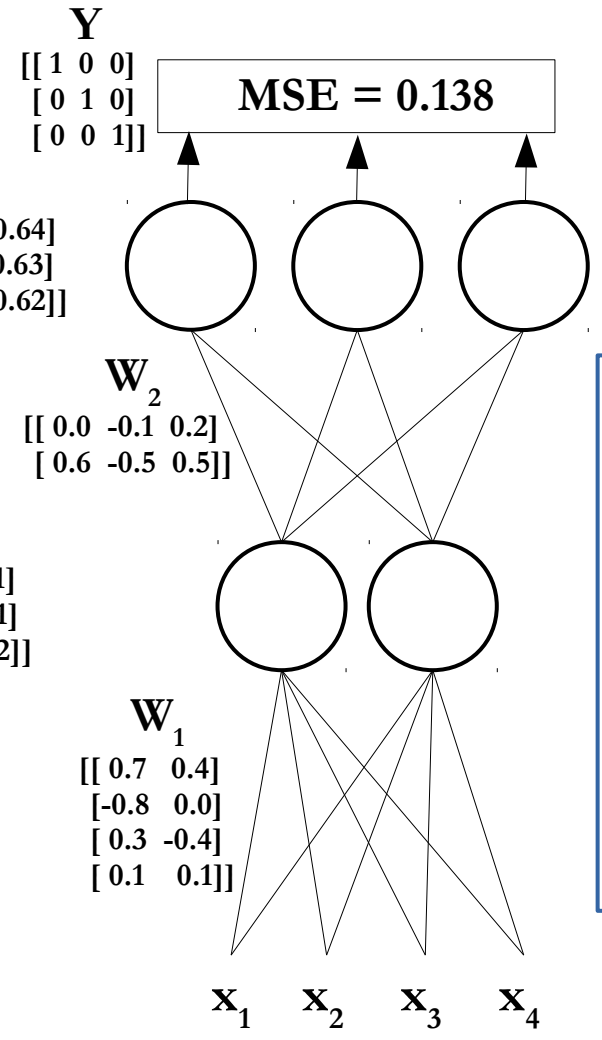


$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} (o_{out\,n}{}^{(e)} - y_n{}^{(e)}) \bullet o_{out\,n}{}^{(e)} (1 - o_{out\,n}{}^{(e)}) \bullet h_{out\,f}{}^{(e)}$$

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

Feedforward

Backpropagation

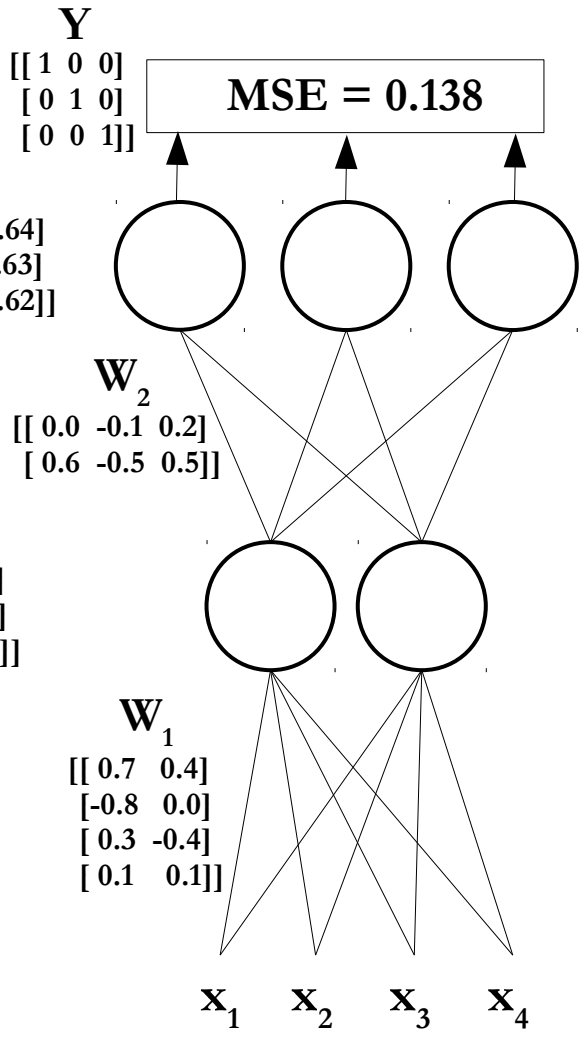


$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}\,(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

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

Feedforward

Backpropagation



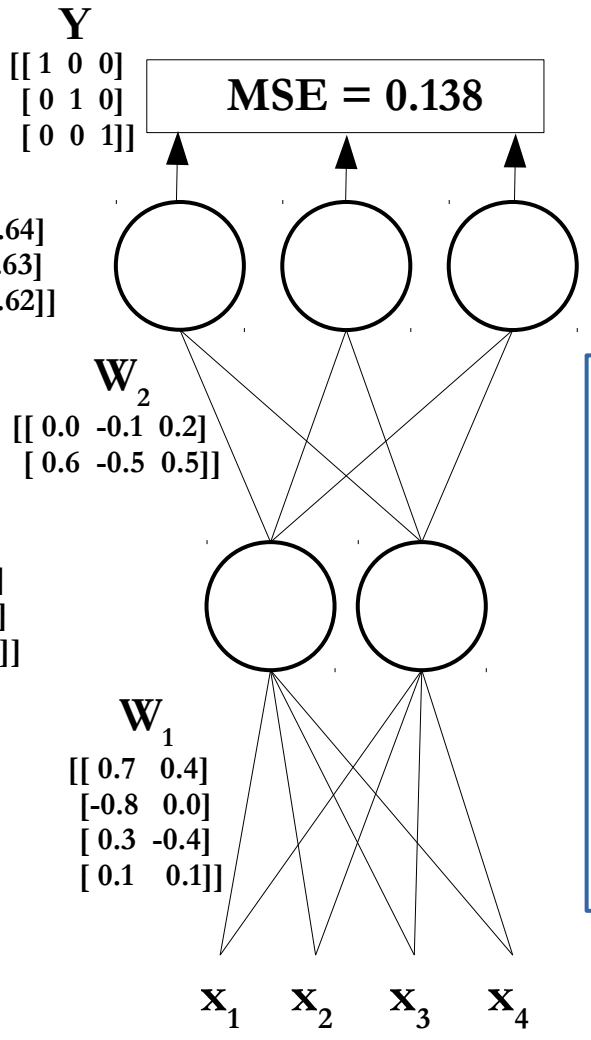
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}\,(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

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

$$\frac{\partial MSE}{\partial W_2}(W_2): \qquad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

Feedforward

Backpropagation



$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}\,(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

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

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

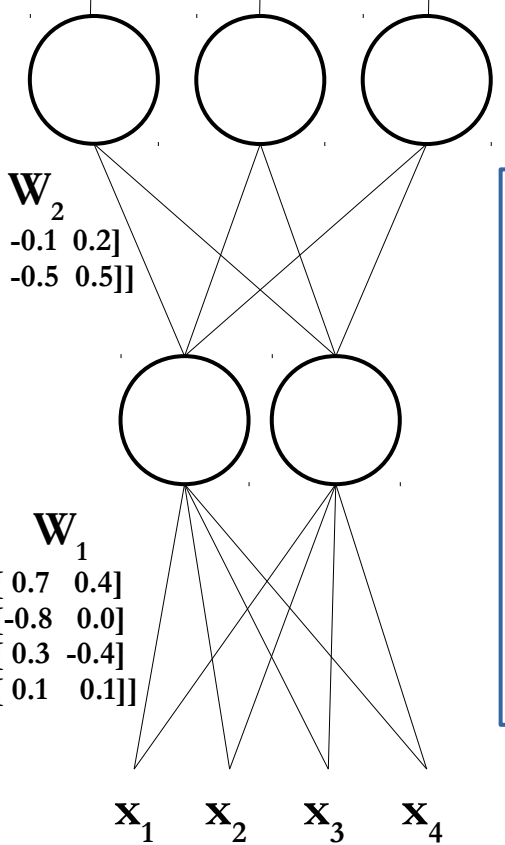
$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$



$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} (o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)} (1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

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

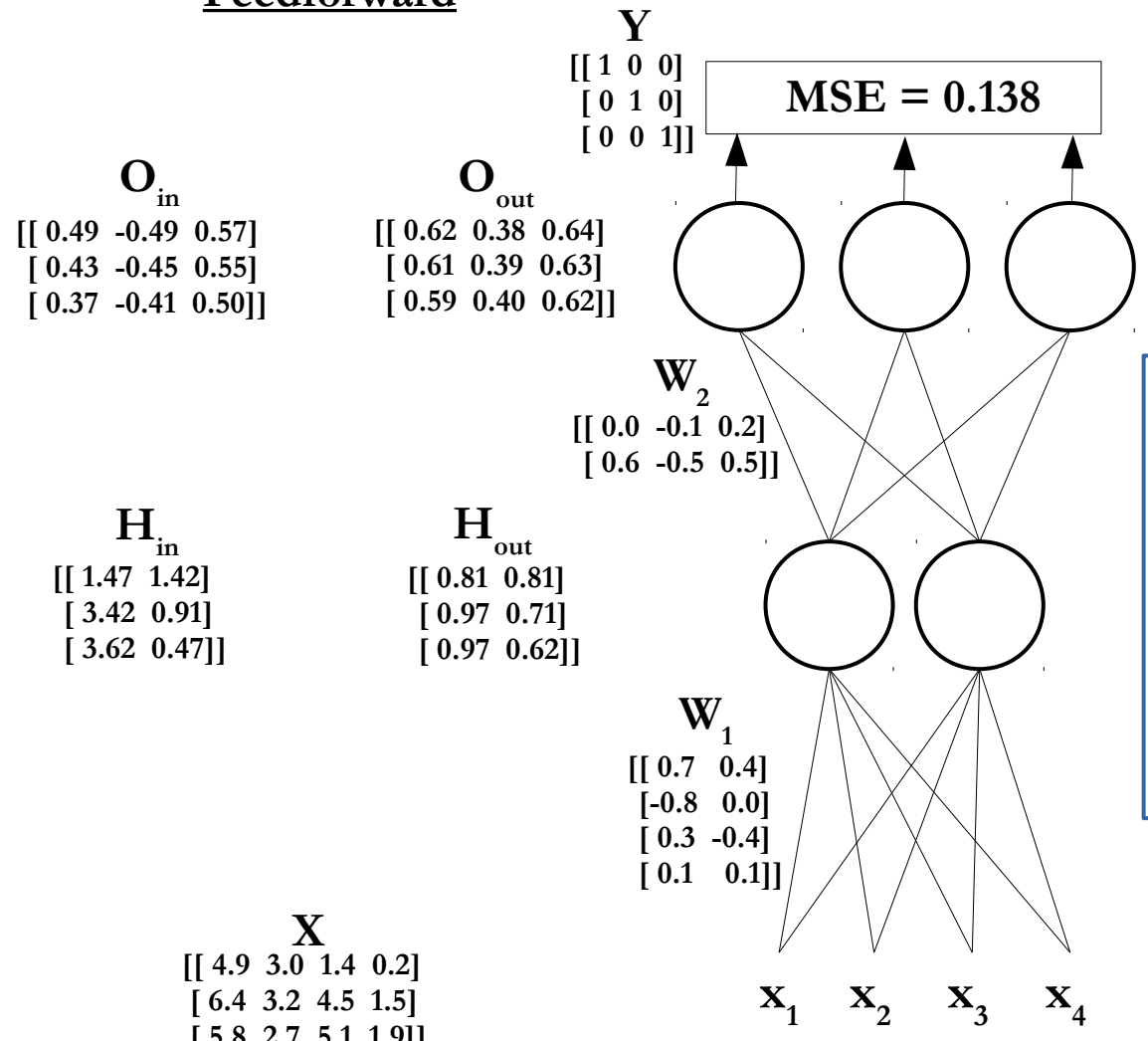
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}\,(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

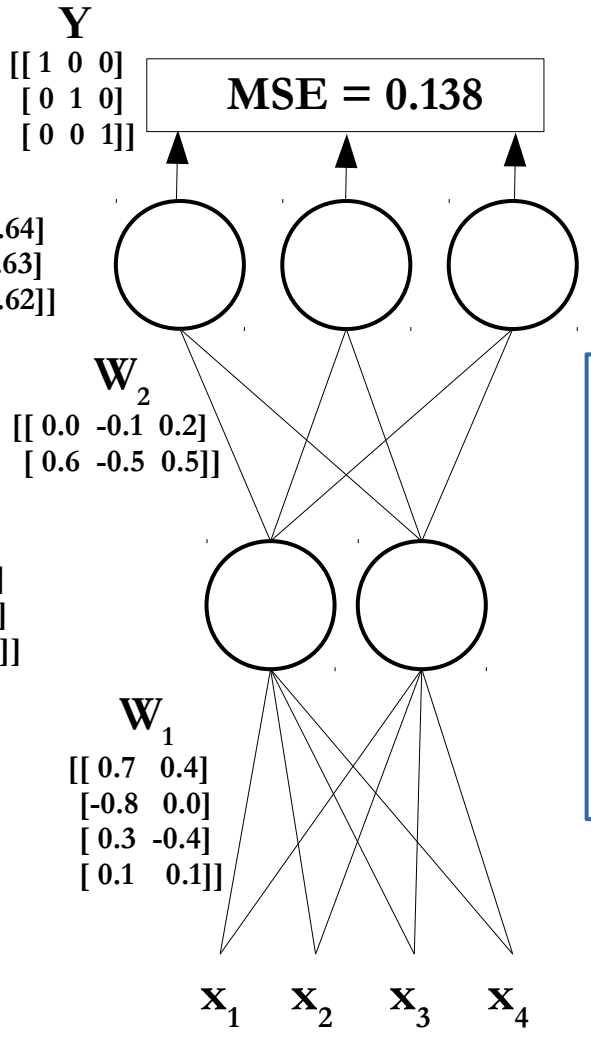
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]
```

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]]
```

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]]
```

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

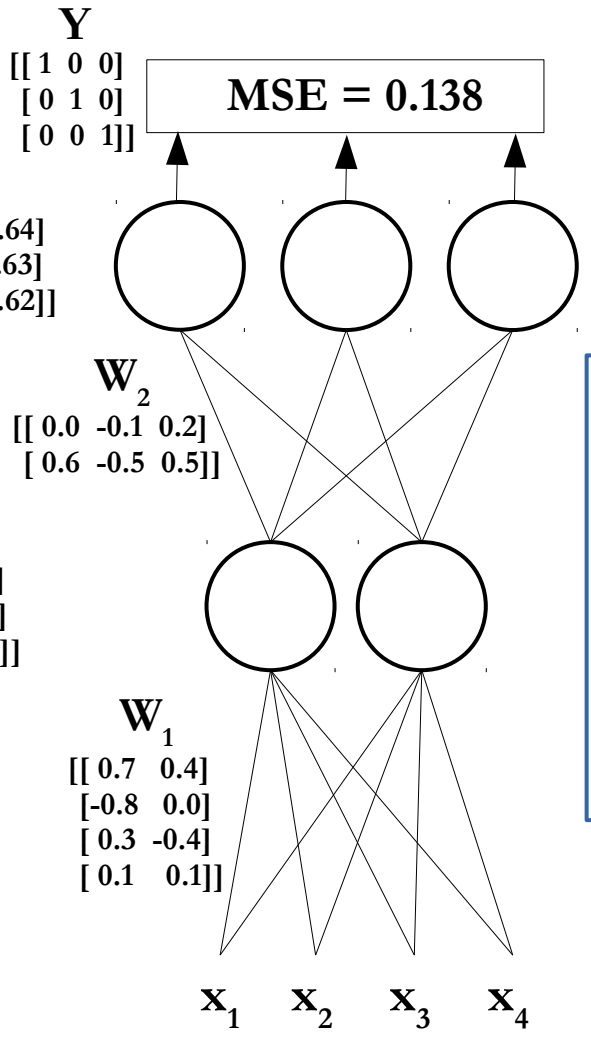
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\ n}^{(e)} - y_n^{(e)}) \cdot o_{out\ n}^{(e)}(1 - o_{out\ n}^{(e)}) \cdot h_{out\ f}^{(e)}$$

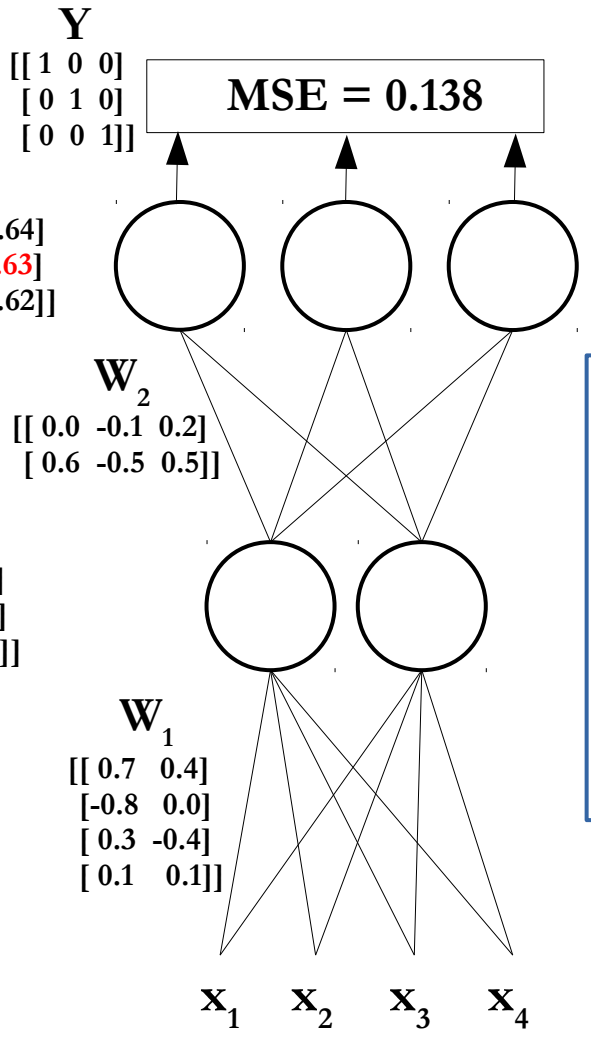
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \cdot \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \cdot \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

$o_{out\ 3}^{(2)}$

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\ n}^{(e)} - y_n^{(e)}) \bullet o_{out\ n}^{(e)}(1 - o_{out\ n}^{(e)}) \bullet h_{out\ f}^{(e)}$$

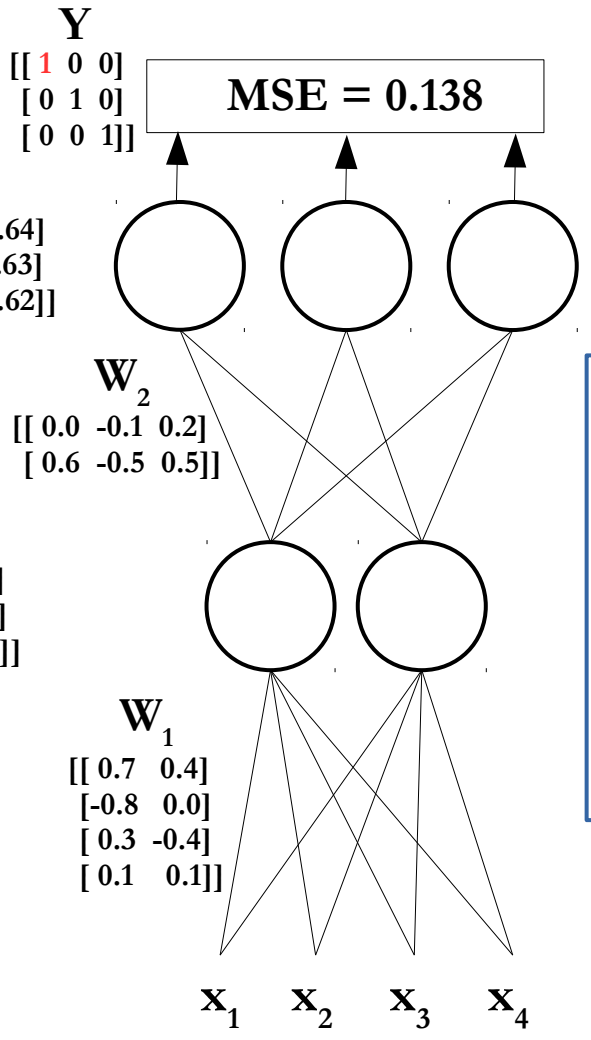
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \cdot o_{out\,n}^{(e)}(1 - o_{out\,n}^{(e)}) \cdot h_{out\,f}^{(e)}$$

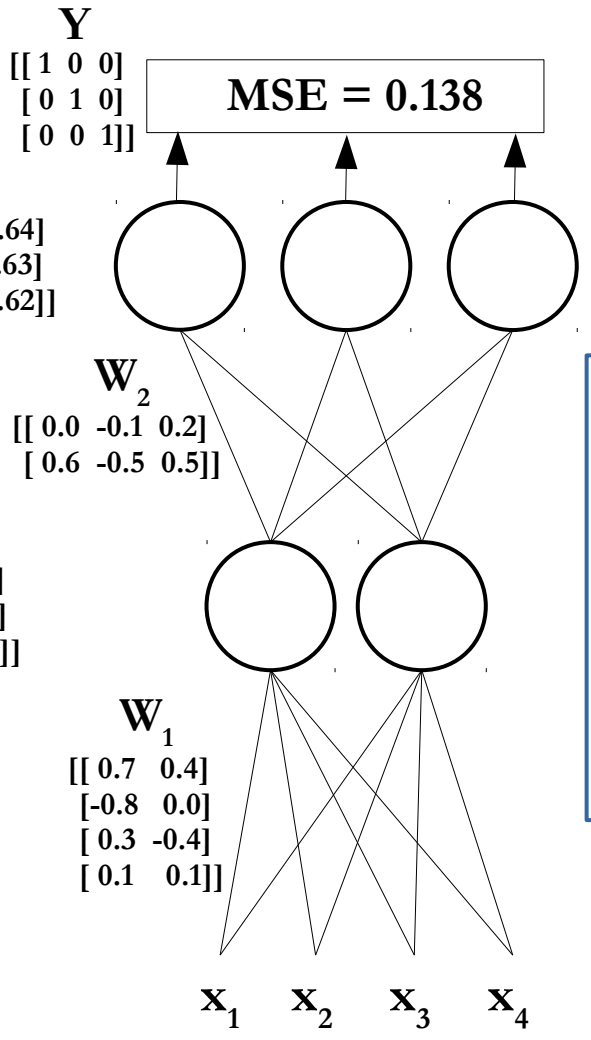
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = \frac{\partial MSE}{\partial O_{out}}$$

$$O_{delta} = O_{error} \cdot \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \cdot \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N} \sum_{e} (o_{out\,n}^{(e)} - y_n^{(e)}) \cdot o_{out\,n}^{(e)} (1 - o_{out\,n}^{(e)}) \cdot h_{out\,f}^{(e)}$$

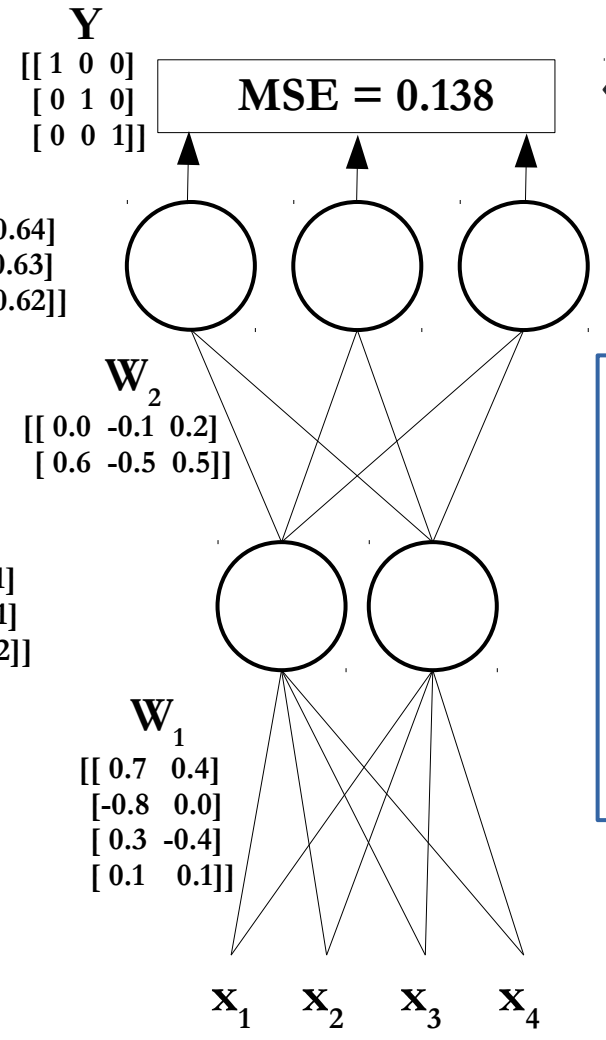
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \cdot \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \cdot \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

O_{error}
[[-0.38 0.38 0.64]
[0.61 -0.61 0.63]
[0.59 0.40 -0.38]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \cdot o_{out\,n}^{(e)}(1 - o_{out\,n}^{(e)}) \cdot h_{out\,f}^{(e)}$$

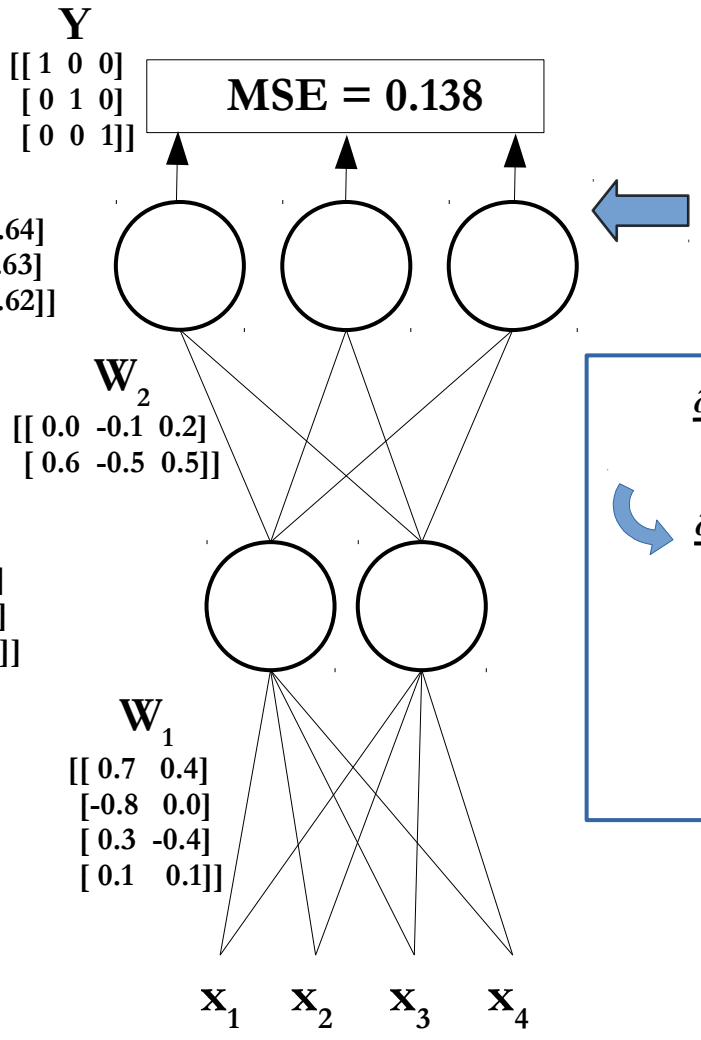
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \cdot \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \cdot \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

O_{error}
[[-0.38 0.38 0.64]
[0.61 -0.61 0.63]
[0.59 0.40 -0.38]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e}(o_{out\,n}^{(e)} - y_n^{(e)}) \bullet o_{out\,n}^{(e)}(1 - o_{out\,n}^{(e)}) \bullet h_{out\,f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N} \sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)} \left(1 - o_{out\ n}^{(e)}\right) \bullet h_{out\ f}^{(e)}$$

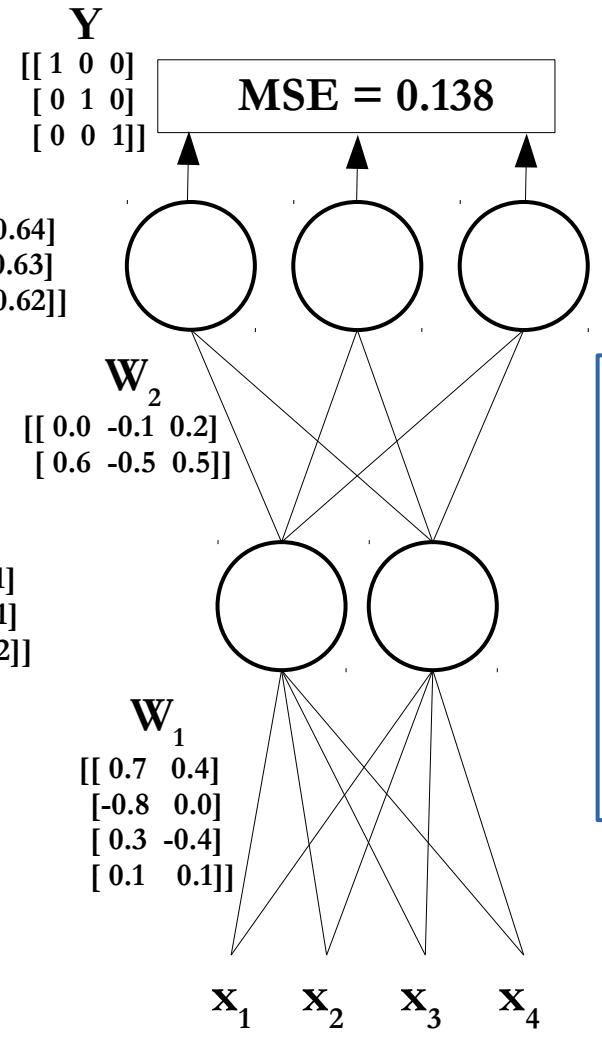
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

O_{error}
[[-0.38 0.38 0.64]
[0.61 -0.61 0.63]
[0.59 0.40 -0.38]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)} (1 - o_{out\ n}^{(e)}) \bullet h_{out\ f}^{(e)}$$

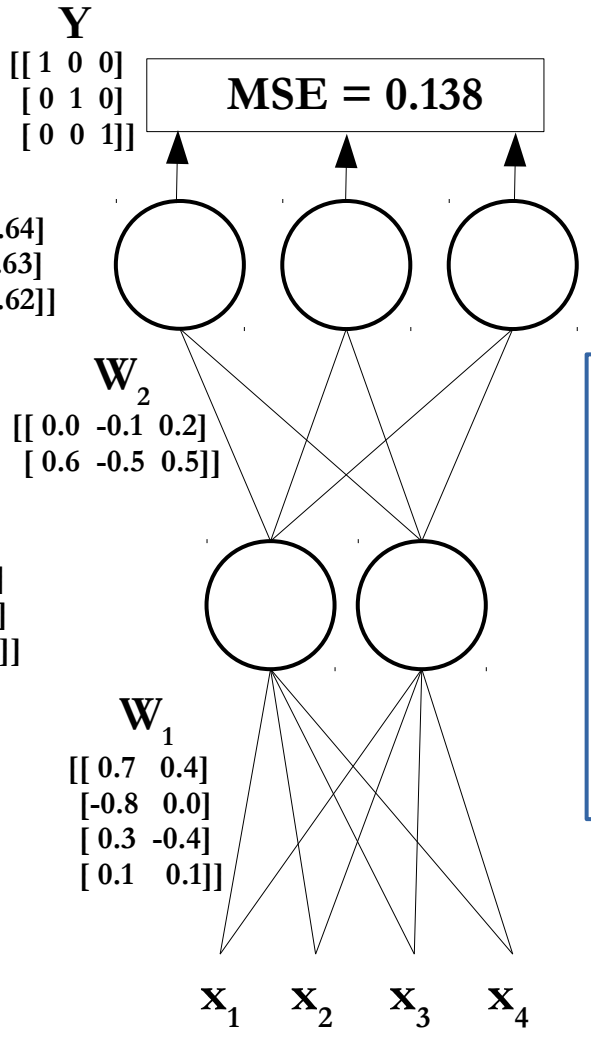
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

$$O_{delta} = O_{error} \bullet \frac{\partial O_{out}}{\partial O_{in}}$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

O_{error}
[[-0.38 0.38 0.64]
[0.61 -0.61 0.63]
[0.59 0.40 -0.38]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)}\left(1 - o_{out\ n}^{(e)}\right) \bullet h_{out\ f}^{(e)}$$

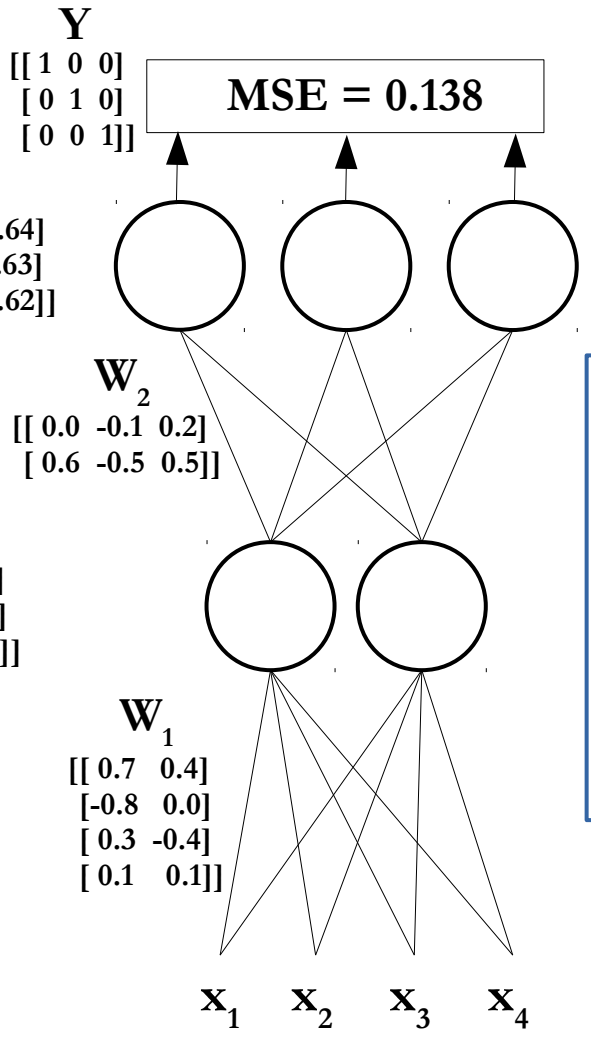
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)} (1 - o_{out\ n}^{(e)}) \bullet h_{out\ f}^{(e)}$$

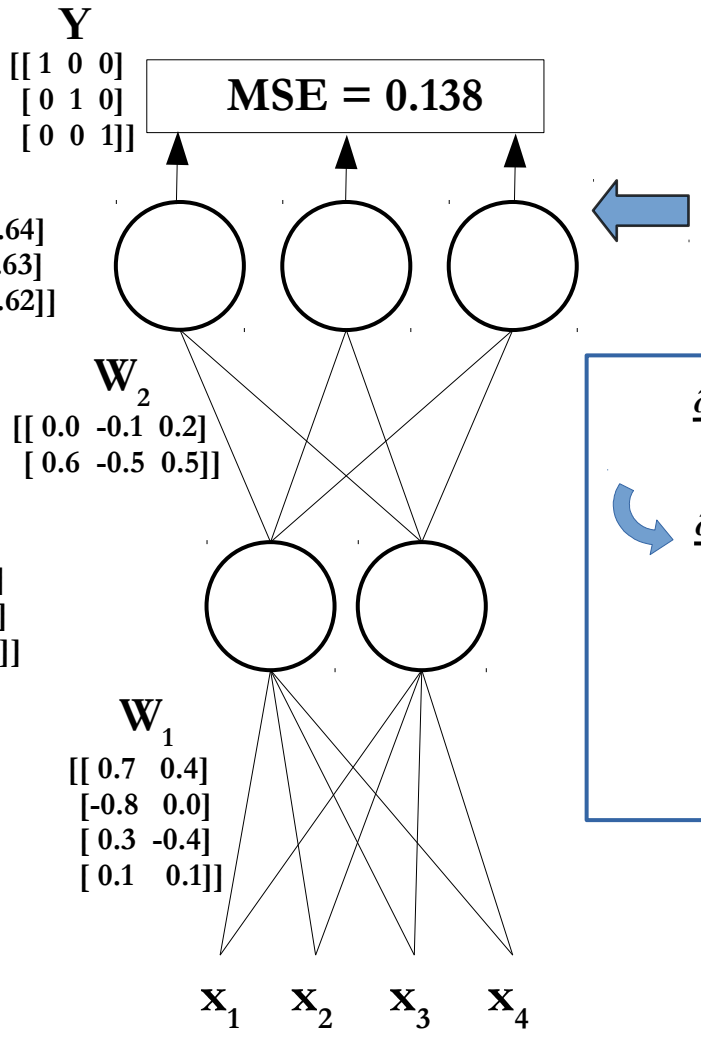
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)} (1 - o_{out\ n}^{(e)}) \bullet h_{out\ f}^{(e)}$$

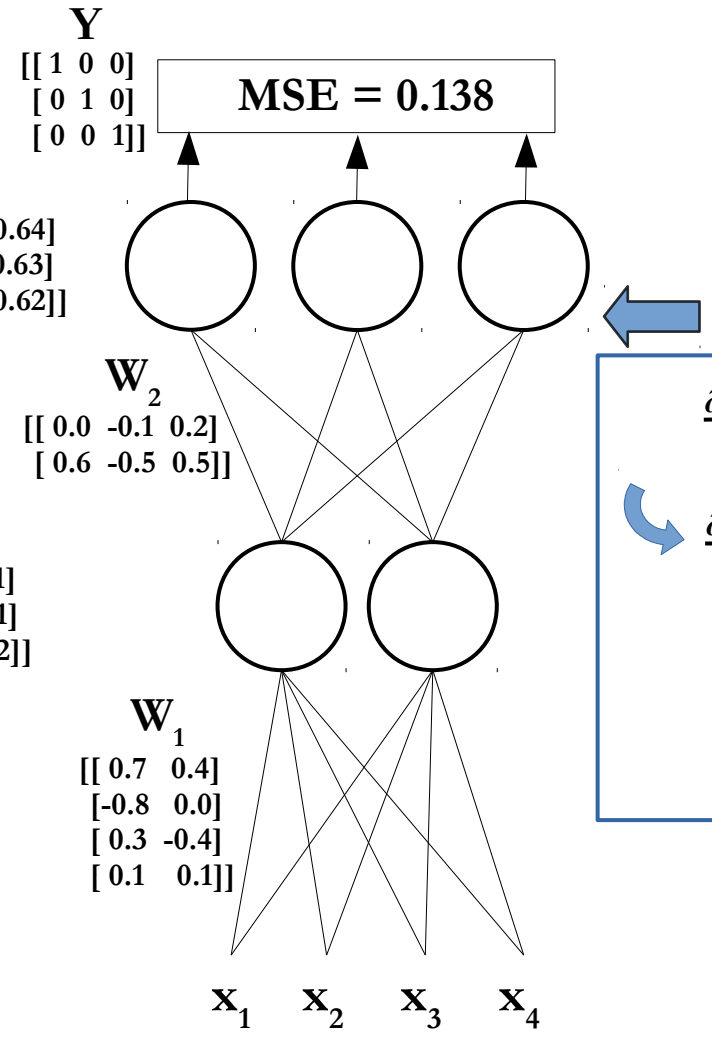
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{error\ n}^{(e)} \bullet o_{out\ n}^{(e)} (1 - o_{out\ n}^{(e)}) \bullet h_{out\ f}^{(e)}$$

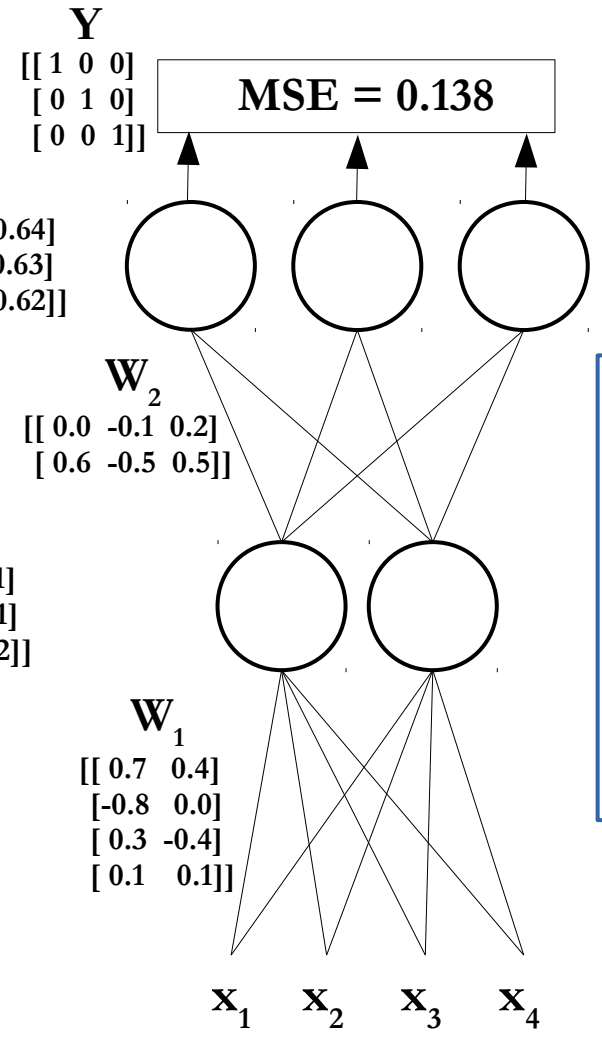
$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

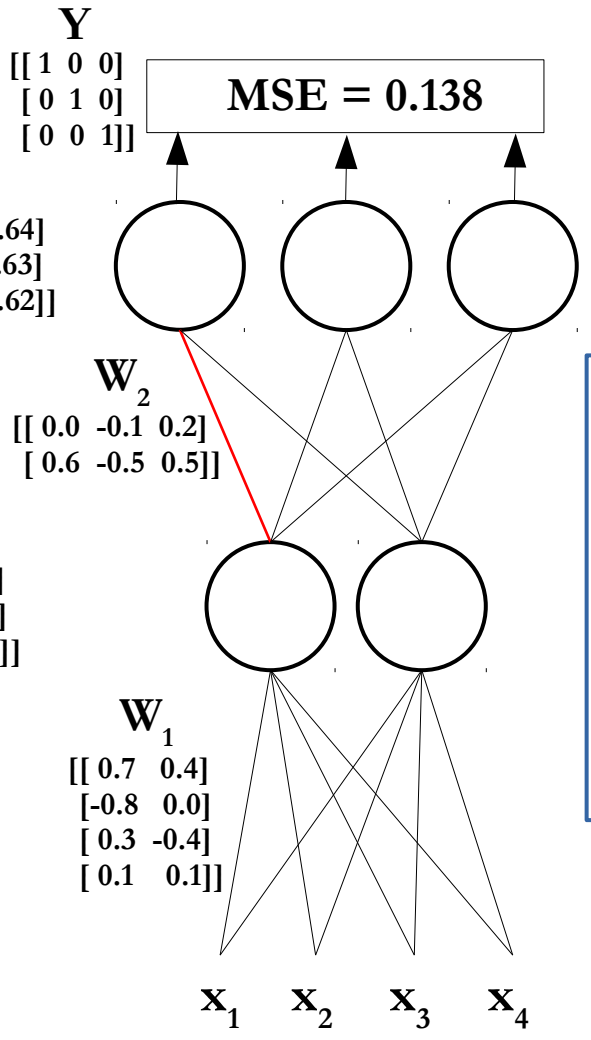
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \qquad \odot \ (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial w_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{\text{delta n}}^{(e)} \bullet h_{\text{out f}}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-11}}} =$$

$$\odot \ (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

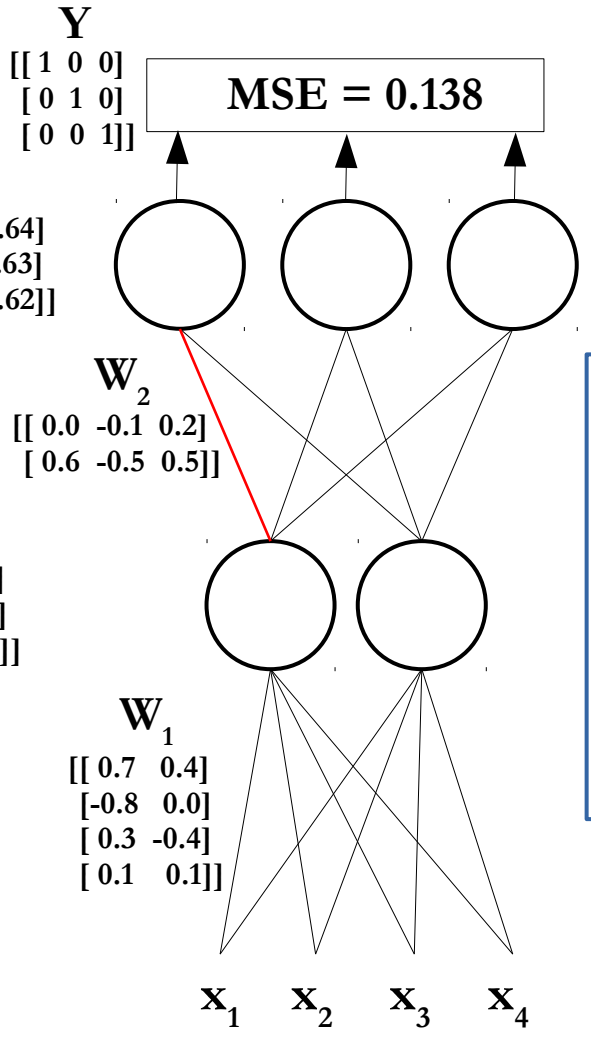
$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{1}{N} \sum_{e} o_{delta\ 1}^{(e)} \bullet h_{out\ 1}^{(e)}$$

$$\odot\ (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

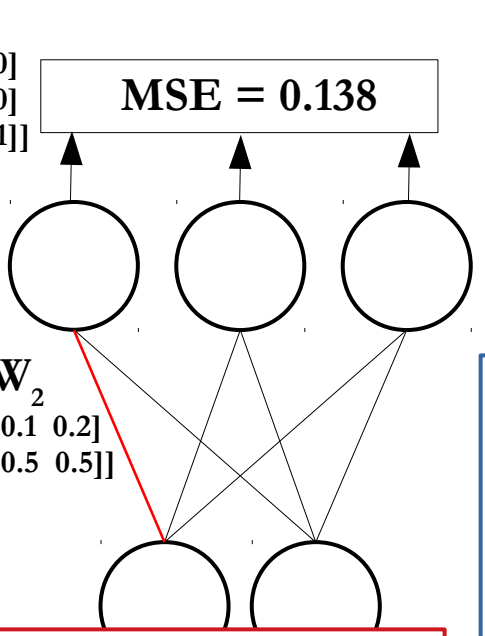
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{1}{N}\sum_{e} o_{delta\ 1}^{(e)} \bullet h_{out\ 1}^{(e)} \odot (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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

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

H_{in}
[[1.47 1.42]
[3.42 0.91]

H_{out}
[[0.81 0.81]
[0.97 0.71]

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]]

$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{\partial MSE}{\partial o_{out\,1}^{(1)}} \frac{do_{out\,1}^{(1)}}{do_{in\,1}^{(1)}} \frac{\partial o_{in\,1}^{(1)}}{\partial w_{2\text{-}11}} + \frac{\partial MSE}{\partial o_{out\,1}^{(2)}} \frac{do_{out\,1}^{(2)}}{do_{in\,1}^{(2)}} \frac{\partial o_{in\,1}^{(2)}}{\partial w_{2\text{-}11}}$$

$$= \frac{1}{N} \sum_{e} (o_{out\,1}^{(e)} - y_1^{(e)}) \cdot o_{out\,1}^{(e)} (1 - o_{out\,1}^{(e)}) \cdot h_{out\,1}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\,n}^{(e)} \cdot h_{out\,f}^{(e)}$$

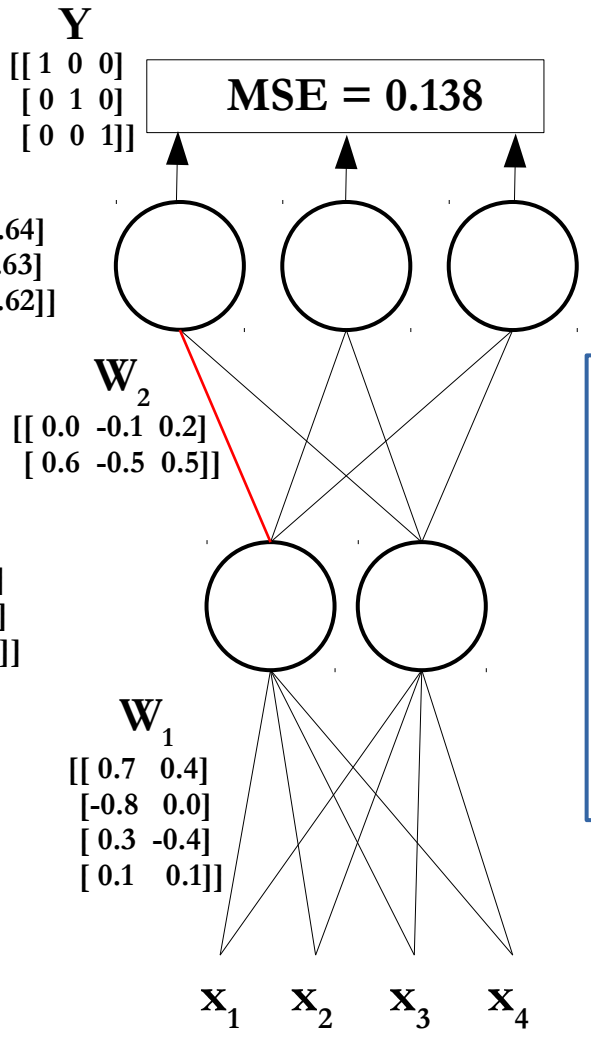
$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{1}{N} \sum_{e} o_{delta\,1}^{(e)} \cdot h_{out\,1}^{(e)}$$

$$\odot (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \cdot \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

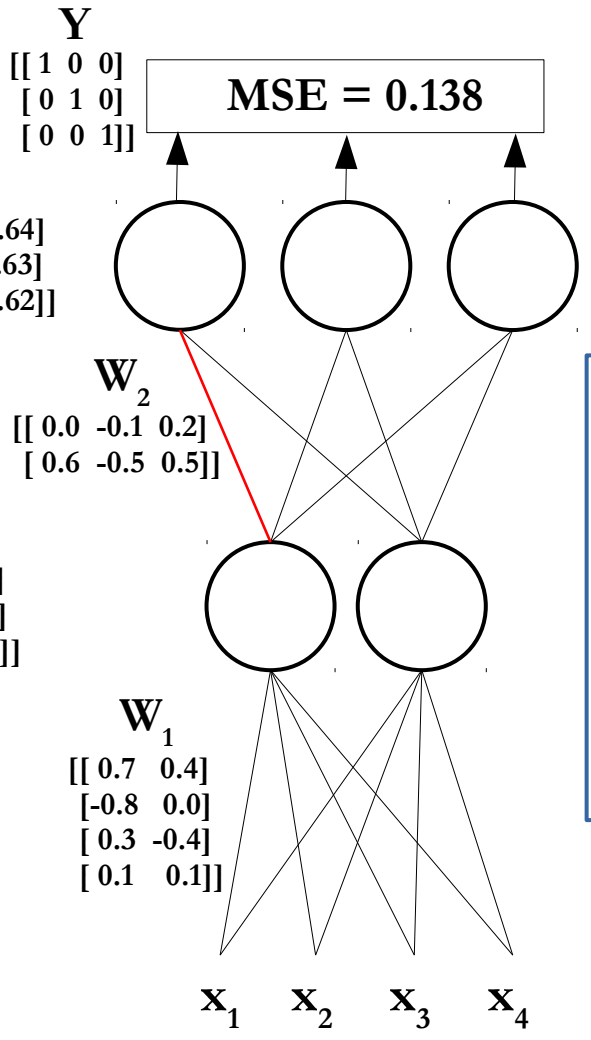
$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{1}{N}\sum_{e} o_{delta\ 1}^{(e)} \bullet h_{out\ 1}^{(e)}$$

$$\odot\ (1 - O_{out})$$

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

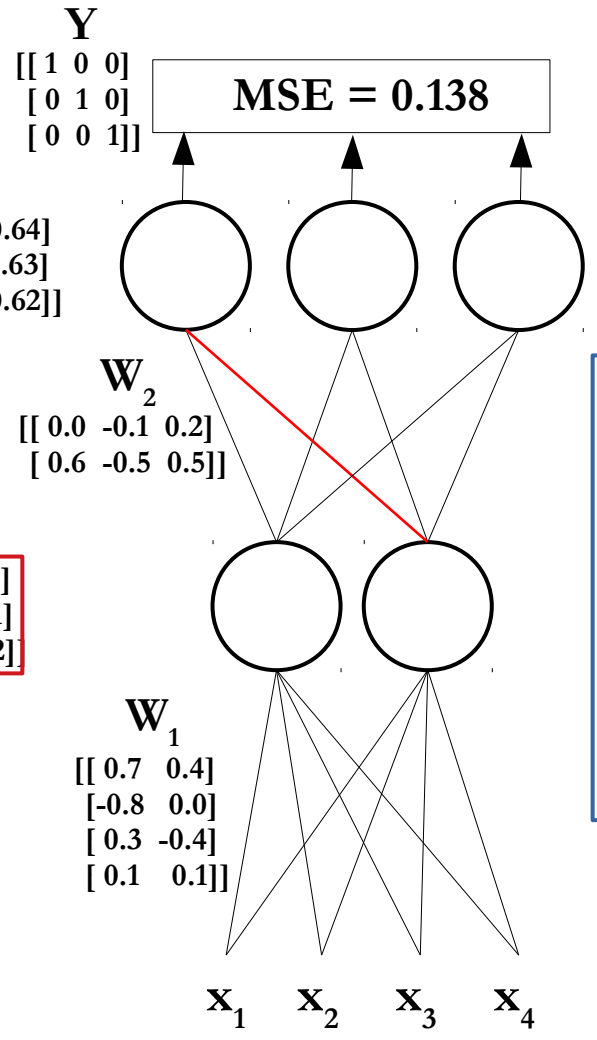
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}11}} = \frac{1}{N}\sum_{e} o_{delta\ 1}^{(e)} \bullet h_{out\ 1}^{(e)} = w_{2\text{-}11\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

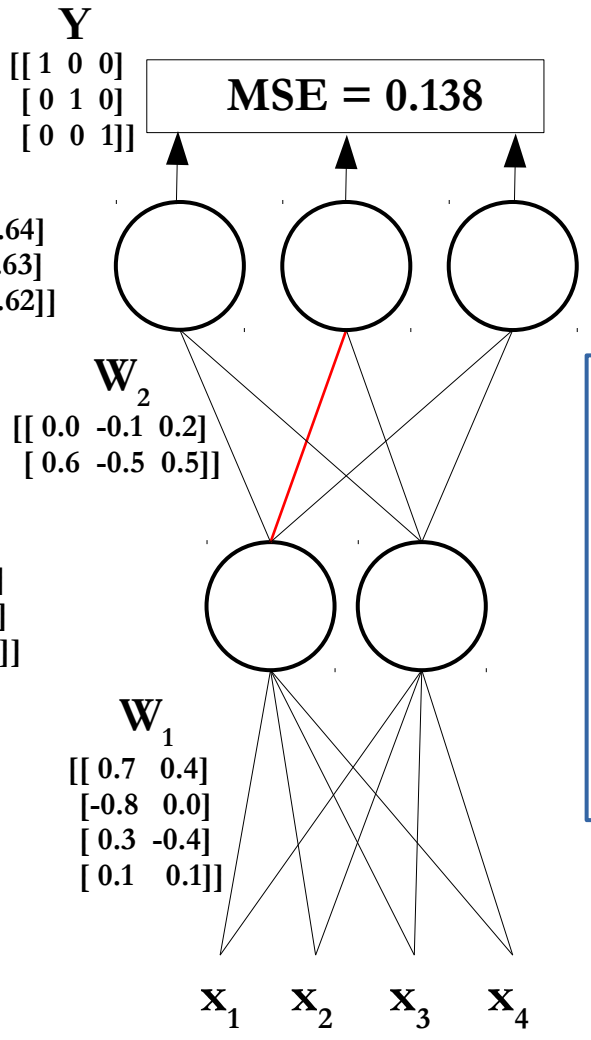
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}21}} = \frac{1}{N}\sum_{e} o_{delta\ 1}^{(e)} \bullet h_{out\ 2}^{(e)} = w_{2\text{-}21\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

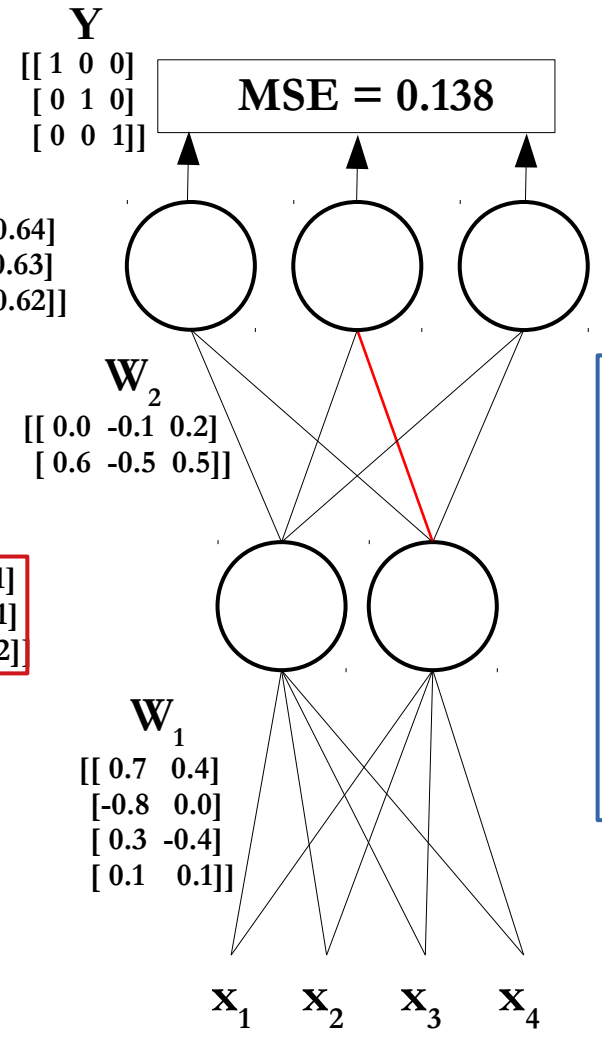
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}12}} = \frac{1}{N}\sum_{e} o_{delta\ 2}^{(e)} \bullet h_{out\ 1}^{(e)} = w_{2\text{-}12\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

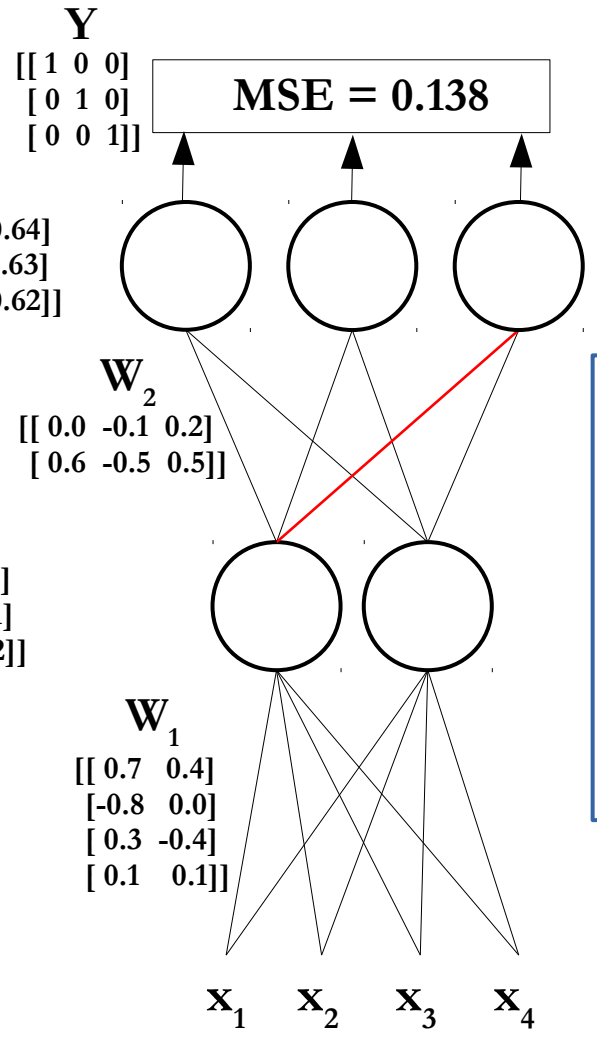
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}22}} = \frac{1}{N}\sum_{e} o_{delta\ 2}^{(e)} \bullet h_{out\ 2}^{(e)} = w_{2\text{-}22\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

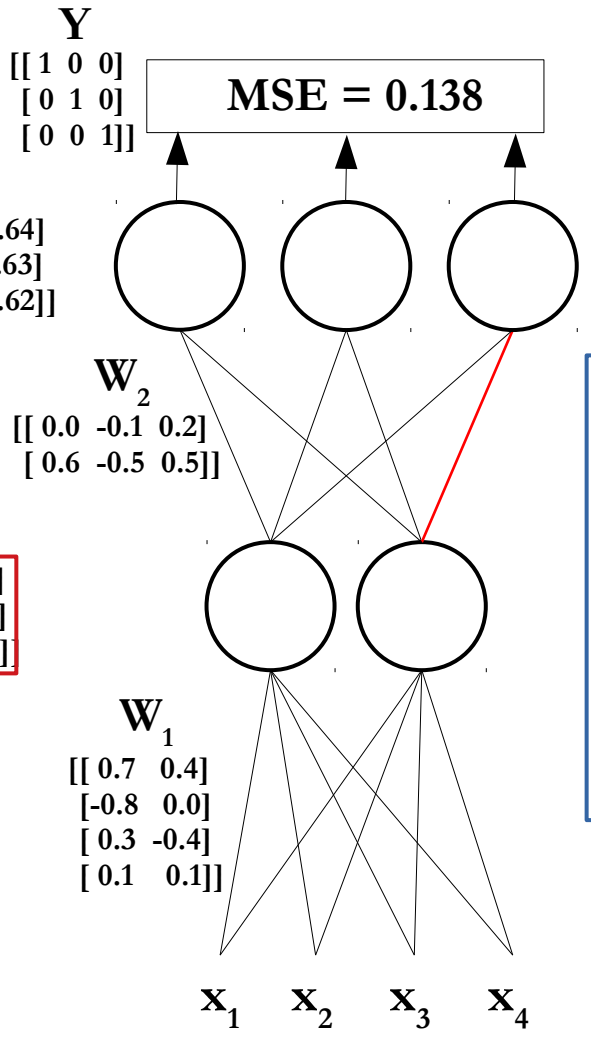
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}13}} = \frac{1}{N}\sum_{e} o_{delta\ 3}^{(e)} \bullet h_{out\ 1}^{(e)} = w_{2\text{-}13\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

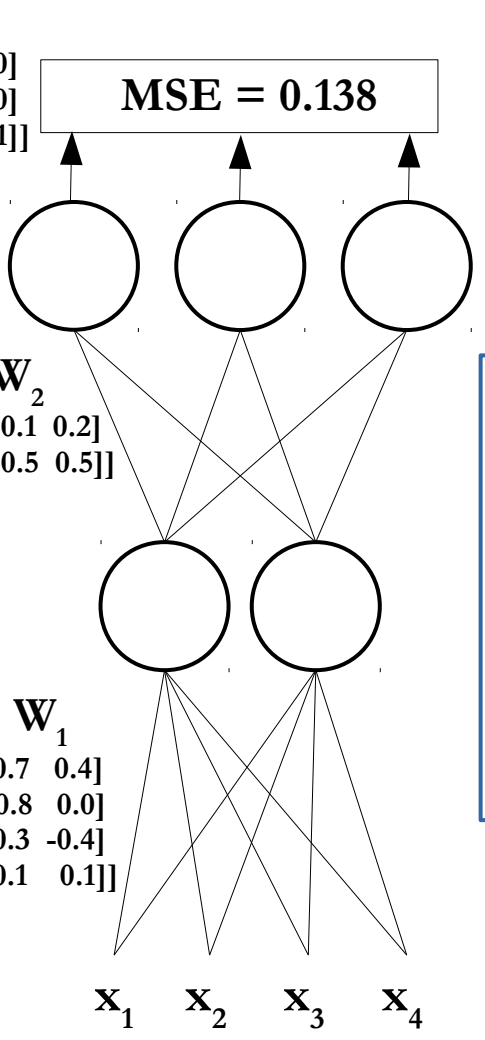
$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}23}} = \frac{1}{N}\sum_{e} o_{delta\ 3}^{(e)} \bullet h_{out\ 2}^{(e)} = w_{2\text{-}23\ update} \odot (1 - O_{out})$$

$$W_{2\text{-}update} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation



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]]

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]]

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]]

O_{delta}^{T}
[[-0.09 0.15 0.14]
[0.09 -0.15 0.10]
[0.15 0.15 -0.09]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta} \bullet \frac{\partial O_{in}}{\partial W_2}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

$$W_{2\text{-update}} = O_{delta}{}^{T} \bullet H_{out}$$

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

3x2 3x3 3x2

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$W_{2\text{-update}} = \begin{bmatrix} w_{2\text{-}11\ update} & w_{2\text{-}21\ update} \\ w_{2\text{-}12\ update} & w_{2\text{-}22\ update} \\ w_{2\text{-}13\ update} & w_{2\text{-}23\ update} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

3x2 3x3 3x2

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$W_{2\text{-update}} = \begin{bmatrix} w_{2\text{-}11\ update} & w_{2\text{-}21\ update} \\ w_{2\text{-}12\ update} & w_{2\text{-}22\ update} \\ w_{2\text{-}13\ update} & w_{2\text{-}23\ update} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

Gradient Descent:

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

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

3x2 3x3 3x2

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_{2\text{-update}} = \begin{bmatrix} w_{2\text{-}11\ update} & w_{2\text{-}12\ update} & w_{2\text{-}13\ update} \\ w_{2\text{-}21\ update} & w_{2\text{-}22\ update} & w_{2\text{-}23\ update} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$\frac{\partial MSE}{\partial w_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

3x2 3x3 3x2

Gradient Descent:

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

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$O_{delta}{}^{T} = \begin{bmatrix} -0.09 & 0.15 & 0.14 \\ 0.09 & -0.15 & 0.10 \\ 0.15 & 0.15 & -0.09 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_{2\text{-update}} = \begin{bmatrix} w_{2\text{-}11\ update} & w_{2\text{-}12\ update} & w_{2\text{-}13\ update} \\ w_{2\text{-}21\ update} & w_{2\text{-}22\ update} & w_{2\text{-}23\ update} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}{}^{(e)} \bullet h_{out\ f}{}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

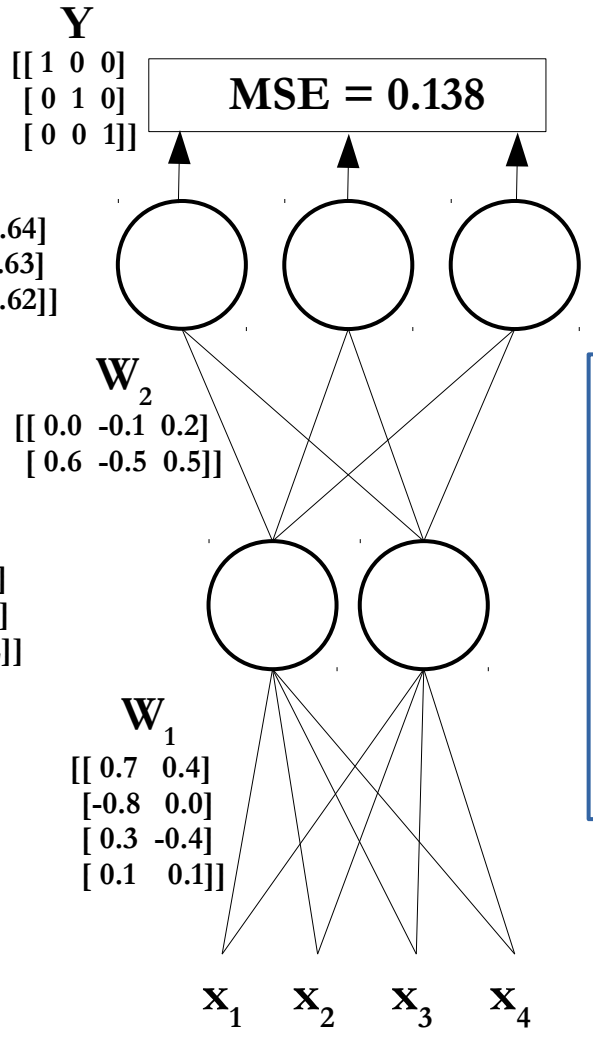
$$O_{delta} = O_{error} \odot O_{out} \odot (1 - O_{out})$$

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

2x3 3x3 3x2

Feedforward

Backpropagation



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]]

$W_{2\text{-update}}$
[[$w_{2\text{-}11\ update}$ $w_{2\text{-}12\ update}$ $w_{2\text{-}13\ update}$]
[$w_{2\text{-}21\ update}$ $w_{2\text{-}22\ update}$ $w_{2\text{-}23\ update}$]]

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]]

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]]

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]]

$$\frac{\partial MSE}{\partial W_{2\text{-fn}}} = \frac{1}{N}\sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

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

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

Feedforward

Backpropagation

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

MSE = 0.138

$$O_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$O_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$H_{out}^{T} = \begin{bmatrix} 0.81 & 0.97 & 0.97 \\ 0.81 & 0.71 & 0.62 \end{bmatrix}$$

$$O_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$O_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$W_{2\text{-update}} = \begin{bmatrix} w_{2\text{-}11\ update} & w_{2\text{-}12\ update} & w_{2\text{-}13\ update} \\ w_{2\text{-}21\ update} & w_{2\text{-}22\ update} & w_{2\text{-}23\ update} \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$H_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$H_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$W_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$X = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

x_1 x_2 x_3 x_4

$$\frac{\partial MSE}{\partial W_{2\text{-}fn}} = \frac{1}{N} \sum_{e} o_{delta\ n}^{(e)} \bullet h_{out\ f}^{(e)}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad O_{error} = O_{out} - Y$$

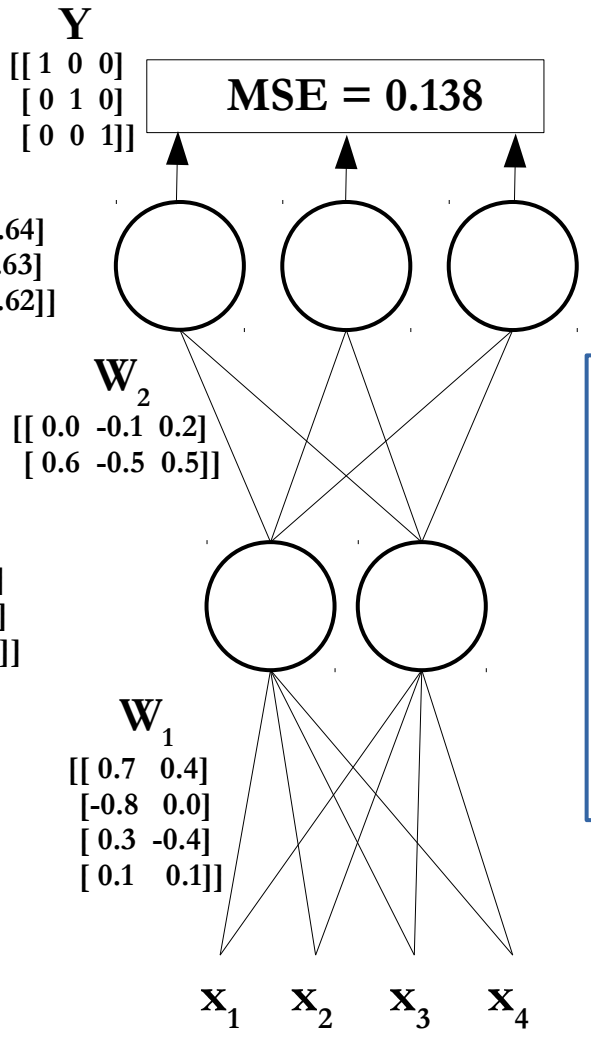
$$O_{delta} = O_{error} \odot O_{out} \odot (1 - O_{out})$$

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

2x3 2x3 3x3

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

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

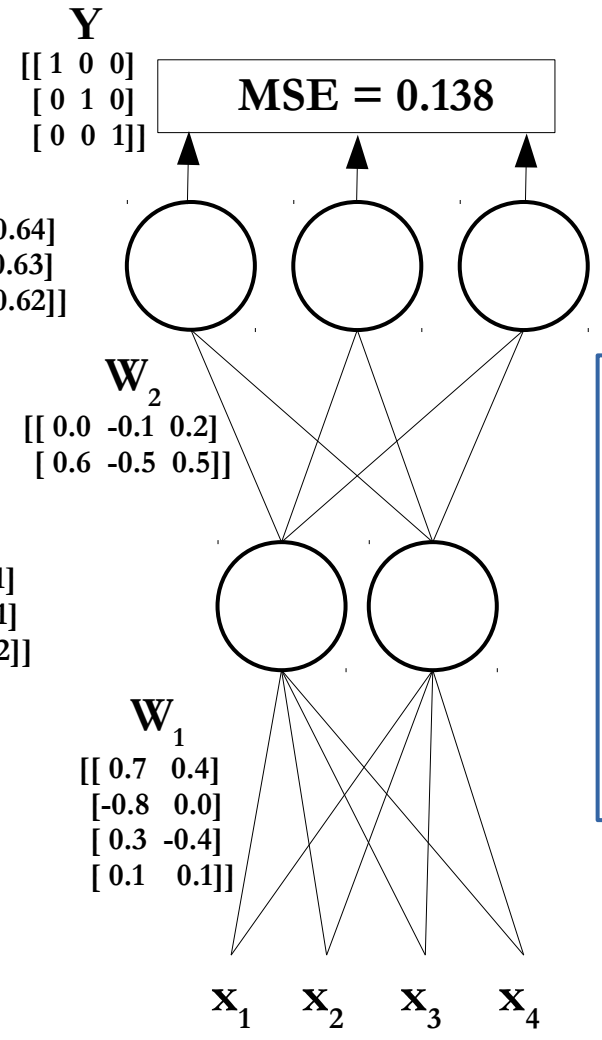
$$O_{error} = O_{out} - Y$$

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

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

Feedforward

Backpropagation



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]]

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]]

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]]

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]]

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

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

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

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

Feedforward

Backpropagation

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

MSE = 0.138

$$\mathbf{O}_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$\mathbf{O}_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$\mathbf{H}_{out}^{T} = \begin{bmatrix} 0.81 & 0.97 & 0.97 \\ 0.81 & 0.71 & 0.62 \end{bmatrix}$$

$$\mathbf{O}_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$\mathbf{O}_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$\mathbf{W}_{2\text{-update}} = \begin{bmatrix} 0.023 & 0.003 & 0.020 \\ 0.013 & 0.003 & 0.019 \end{bmatrix}$$

$$\mathbf{W}_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$\frac{\partial MSE}{\partial W_2}(W_2): \quad \mathbf{O}_{error} = \mathbf{O}_{out} - \mathbf{Y}$$

$$\mathbf{O}_{delta} = \mathbf{O}_{error} \odot \mathbf{O}_{out} \odot (1 - \mathbf{O}_{out})$$

$$\mathbf{W}_{2\text{-update}} = \frac{1}{N} (\mathbf{H}_{out}^{T} \bullet \mathbf{O}_{delta})$$

$$\mathbf{H}_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

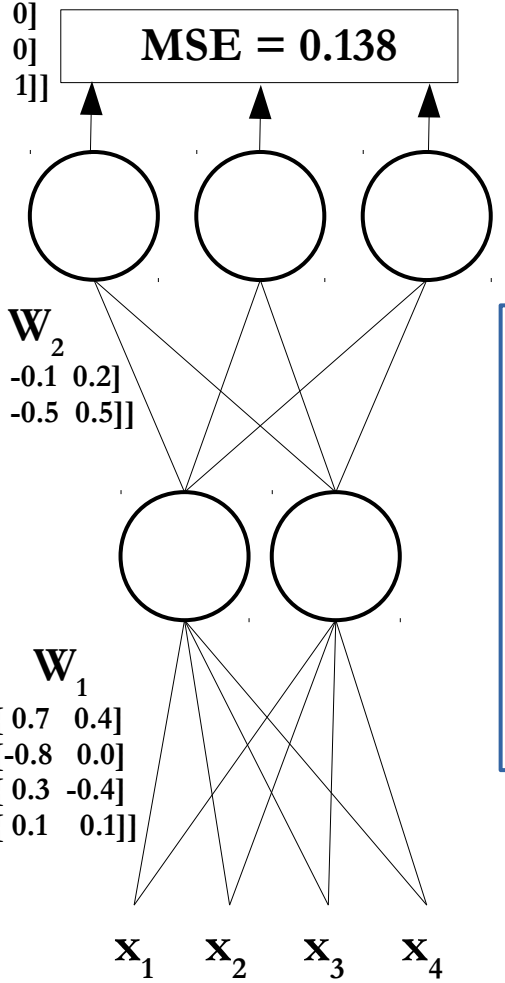
$$\mathbf{H}_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$\mathbf{W}_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

Gradient Descent:

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

$$\mathbf{X} = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$



Feedforward

Backpropagation

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

MSE = 0.138

$$\mathbf{O}_{in} = \begin{bmatrix} 0.49 & -0.49 & 0.57 \\ 0.43 & -0.45 & 0.55 \\ 0.37 & -0.41 & 0.50 \end{bmatrix}$$

$$\mathbf{O}_{out} = \begin{bmatrix} 0.62 & 0.38 & 0.64 \\ 0.61 & 0.39 & 0.63 \\ 0.59 & 0.40 & 0.62 \end{bmatrix}$$

$$\mathbf{H}_{out}^{T} = \begin{bmatrix} 0.81 & 0.97 & 0.97 \\ 0.81 & 0.71 & 0.62 \end{bmatrix}$$

$$\mathbf{O}_{delta} = \begin{bmatrix} -0.09 & 0.09 & 0.15 \\ 0.15 & -0.15 & 0.15 \\ 0.14 & 0.10 & -0.09 \end{bmatrix}$$

$$\mathbf{O}_{error} = \begin{bmatrix} -0.38 & 0.38 & 0.64 \\ 0.61 & -0.61 & 0.63 \\ 0.59 & 0.40 & -0.38 \end{bmatrix}$$

$$\mathbf{W}_{2\text{-update}} = \begin{bmatrix} 0.023 & 0.003 & 0.020 \\ 0.013 & 0.003 & 0.019 \end{bmatrix}$$

$$\mathbf{W}_2 = \begin{bmatrix} 0.0 & -0.1 & 0.2 \\ 0.6 & -0.5 & 0.5 \end{bmatrix}$$

$$\mathbf{H}_{in} = \begin{bmatrix} 1.47 & 1.42 \\ 3.42 & 0.91 \\ 3.62 & 0.47 \end{bmatrix}$$

$$\mathbf{H}_{out} = \begin{bmatrix} 0.81 & 0.81 \\ 0.97 & 0.71 \\ 0.97 & 0.62 \end{bmatrix}$$

$$\mathbf{W}_1 = \begin{bmatrix} 0.7 & 0.4 \\ -0.8 & 0.0 \\ 0.3 & -0.4 \\ 0.1 & 0.1 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 4.9 & 3.0 & 1.4 & 0.2 \\ 6.4 & 3.2 & 4.5 & 1.5 \\ 5.8 & 2.7 & 5.1 & 1.9 \end{bmatrix}$$

$\mathbf{x}_1$ $\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_4$

Gradient Descent:

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

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{W}_2}(\mathbf{W}_2): \quad \mathbf{O}_{error} = \mathbf{O}_{out} - \mathbf{Y}$$

$$\mathbf{O}_{delta} = \mathbf{O}_{error} \odot \mathbf{O}_{out} \odot (1 - \mathbf{O}_{out})$$

$$\mathbf{W}_{2\text{-update}} = \frac{1}{N} (\mathbf{H}_{out}^{T} \bullet \mathbf{O}_{delta})$$

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{W}_1}(\mathbf{W}_1):$$