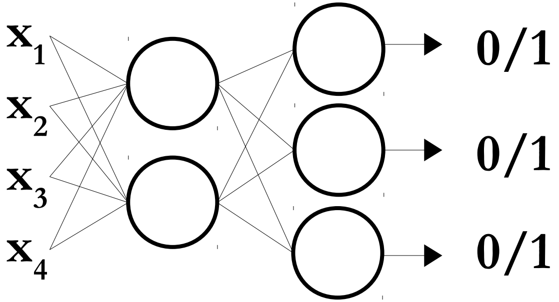
Deep Learning

How does the algorithm make a decision?

mathematical conceptual

Feedforward



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

mathematical conceptual

Cost function/Loss function:

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

$$\textbf{MSE} = \frac{1}{N}\sum_{e}\sum_{n}(step(step(XW_1)W_2)_{e,n} - Y_{e,n})^2$$

Gradient Descent:

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

$$\frac{\partial\, \textbf{MSE}}{\partial\, \textbf{W}_1}(\textbf{W}_1) =$$

$$\frac{\partial\, \textbf{MSE}}{\partial\, \textbf{W}_2}(\textbf{W}_2) =$$

N: e * n

Chain Rule

$y = 3x - 2$

$z = y^2$

Chain Rule

$y = 3x - 2$

$z = y^2$

$$\frac{dz}{dx} =$$

Chain Rule

$y = 3x - 2$

$z = y^2$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

x = 5v - 2

y = 3x - 2

$z = y^2$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

$x = 5v - 2$

$y = 3x - 2$

$z = y^2$

$$\frac{dz}{dv} = \frac{dz}{dy} \cdot \frac{dy}{dx} \cdot \frac{dx}{dv}$$

Chain Rule

$y = 3x - 2$

$z = y^2$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

$y = 3x - 2 \qquad \frac{dy}{dx} = 3$

$z = y^2 \qquad \frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

x = 2:

$y = 3x - 2$ $\frac{dy}{dx} = 3$

$z = y^2$ $\frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

x = 2:

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$ $\qquad y(2) = 4$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$ $\qquad z(4) = 16$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Chain Rule

x = 2:

$y = 3x - 2 \qquad \frac{dy}{dx} = 3 \qquad y(2) = 4$

$z = y^2 \qquad \frac{dz}{dy} = 2y \qquad z(4) = 16$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} \qquad \frac{dz}{dx}(2) =$$

Chain Rule

x = 2:

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$ $\qquad y(2) = 4$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$ $\qquad z(4) = 16$ $\qquad \frac{dz}{dy}(4) =$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$\frac{dz}{dx}(2) =$

Chain Rule

x = 2:

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$ $\qquad y(2) = 4$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$ $\qquad z(4) = 16$ $\qquad \frac{dz}{dy}(4) = 8$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$$\frac{dz}{dx}(2) =$$

Chain Rule

$x = 2$:

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$ $\qquad y(2) = 4$ $\qquad \frac{dy}{dx}(2) =$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$ $\qquad z(4) = 16$ $\qquad \frac{dz}{dy}(4) = 8$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$$\frac{dz}{dx}(2) =$$

Chain Rule

x = 2:

$y = 3x - 2$ $\frac{dy}{dx} = 3$ $y(2) = 4$ $\frac{dy}{dx}(2) = 3$

$z = y^2$ $\frac{dz}{dy} = 2y$ $z(4) = 16$ $\frac{dz}{dy}(4) = 8$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$$\frac{dz}{dx}(2) =$$

Chain Rule

x = 2:

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$ $\qquad y(2) = 4$ $\qquad \frac{dy}{dx}(2) = 3$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$ $\qquad z(4) = 16$ $\qquad \frac{dz}{dy}(4) = 8$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$$\frac{dz}{dx}(2) = 8 * 3$$

$$= 24$$

Chain Rule

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

x = 2:

$z = (3x - 2)^2$

$$\frac{dz}{dx}(2) = 8 * 3$$

$$= 24$$

Chain Rule

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

x = 2:

$z = 9x^2 - 12x + 4$

$$\frac{dz}{dx}(2) = 8 * 3$$

$$= 24$$

Chain Rule

$y = 3x - 2$ $\frac{dy}{dx} = 3$

$z = y^2$ $\frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

$z = 9x^2 - 12x + 4$ $\frac{dz}{dx} = 18x - 12$

x = 2:

$\frac{dz}{dx}(2) = 8 * 3$

$= 24$

Chain Rule

$y = 3x - 2$ $\qquad \frac{dy}{dx} = 3$

$z = y^2$ $\qquad \frac{dz}{dy} = 2y$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

x = 2:

$z = 9x^2 - 12x + 4$ $\qquad \frac{dz}{dx} = 18x - 12$

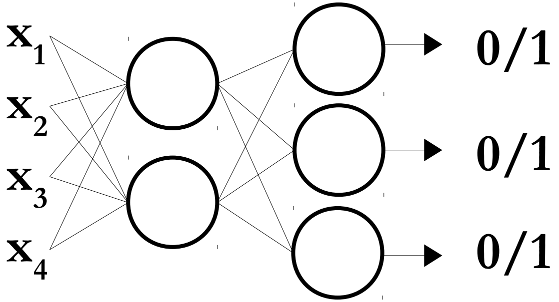
$$\frac{dz}{dx}(2) = 18 * 2 - 12$$
$$= 24$$

$$\frac{dz}{dx}(2) = 8 * 3$$
$$= 24$$

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

$$\mathbf{MSE} = \frac{1}{N}\sum_e\sum_n (step(step(XW_1)W_2)_{e,n} - Y_{e,n})^2$$

Gradient Descent:

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

$$\frac{\partial\, \mathbf{MSE}}{\partial\, \mathbf{W}_1}(\mathbf{W}_1) =$$

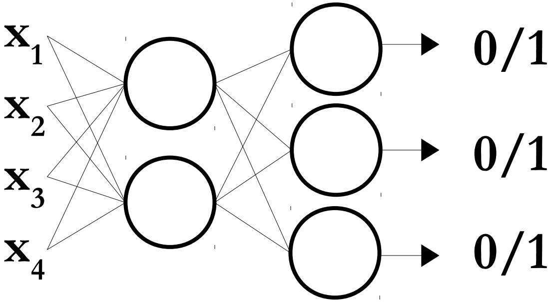
$$\frac{\partial\, \mathbf{MSE}}{\partial\, \mathbf{W}_2}(\mathbf{W}_2) =$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$H_{in} = XW_1$

$H_{out} = step(H_{in})$

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

$O_{out} = step(O_{in})$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

Gradient Descent:

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

$$\frac{\partial \text{ MSE}}{\partial W_1}(W_1) =$$

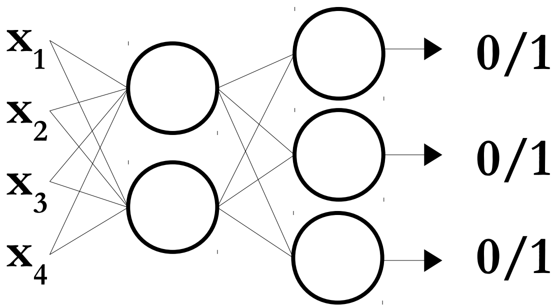
$$\frac{\partial \text{ MSE}}{\partial W_2}(W_2) =$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

Cost function/Loss function:

How do you determine the right parameters for the algorithm?

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

Gradient Descent:

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

$$\frac{\partial\, MSE}{\partial\, W_1}(W_1) =$$

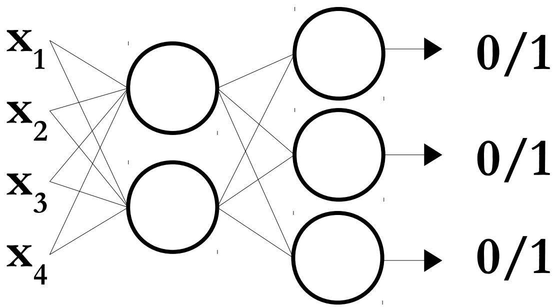
$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}}$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

Cost function/Loss function:

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

Gradient Descent:

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

How do you determine the right parameters for the algorithm?

$$\frac{\partial\, MSE}{\partial\, W_1}(W_1) =$$

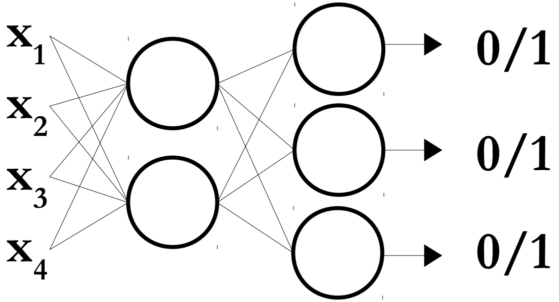
$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}}\frac{\partial\, O_{out}}{\partial\, O_{in}}$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

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

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

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

Cost function/Loss function:

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

Gradient Descent:

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

How do you determine the right parameters for the algorithm?

$$\frac{\partial\, MSE}{\partial\, W_1}(W_1) =$$

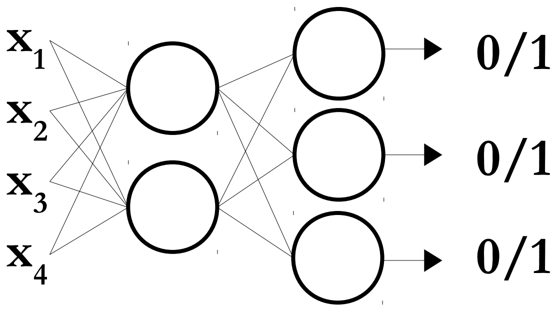
$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, W_2}$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

Gradient Descent:

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

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

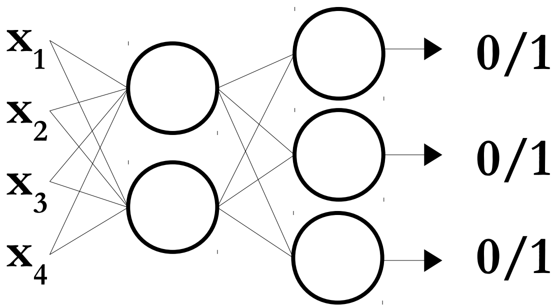
$$\frac{\partial MSE}{\partial W_2}(W_2) = \frac{\partial MSE}{\partial O_{out}}\frac{\partial O_{out}}{\partial O_{in}}\frac{\partial O_{in}}{\partial W_2}$$

N: e * n

Deep Learning

How does the algorithm make a decision?

Feedforward



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

Gradient Descent:

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

Chain Rule:

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

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

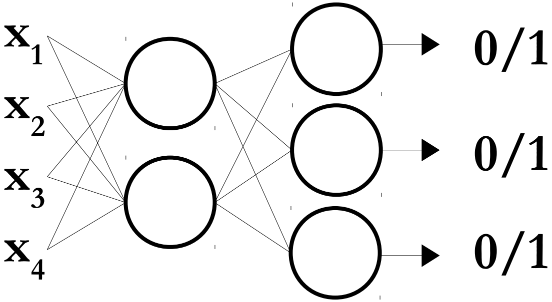
$$\frac{\partial MSE}{\partial W_2}(W_2) = \frac{\partial MSE}{\partial O_{out}} \frac{\partial O_{out}}{\partial O_{in}} \frac{\partial O_{in}}{\partial W_2}$$

N: e * n

Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

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

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

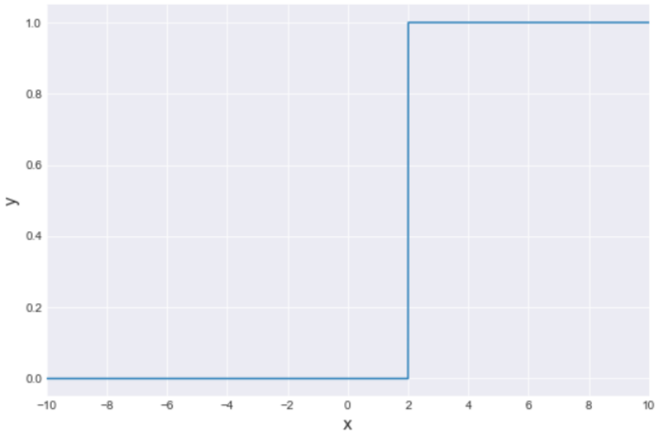
$$\frac{\partial MSE}{\partial W_2}(W_2) = \frac{\partial MSE}{\partial O_{out}} \frac{\partial O_{out}}{\partial O_{in}} \frac{\partial O_{in}}{\partial W_2}$$

N: e * n

Activation Function

Step Function

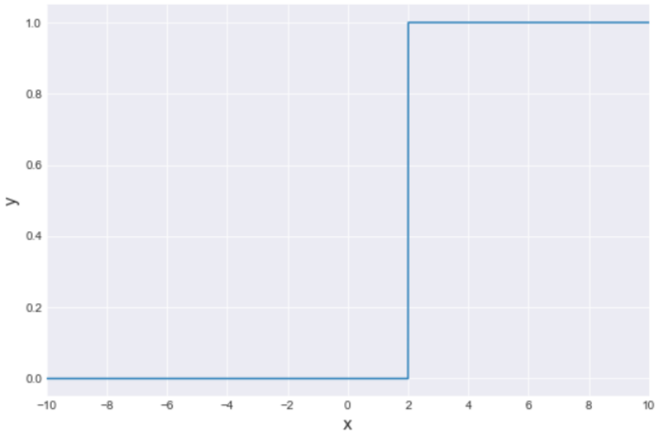
$$y = \begin{cases} 0 & \text{if } x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$



Activation Function

Step Function

$$y = \begin{cases} 0 & \text{if } x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

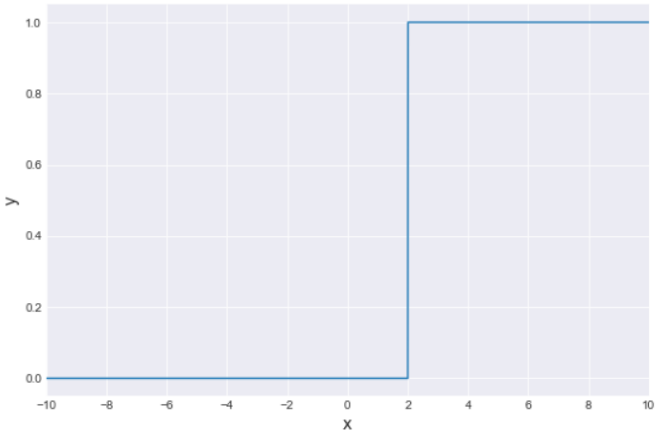
$$\frac{\partial\, MSE}{\partial\, W_1}(W_1) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, H_{out}} \frac{\partial\, H_{out}}{\partial\, H_{in}} \frac{\partial\, H_{in}}{\partial\, W_1}$$

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

Activation Function

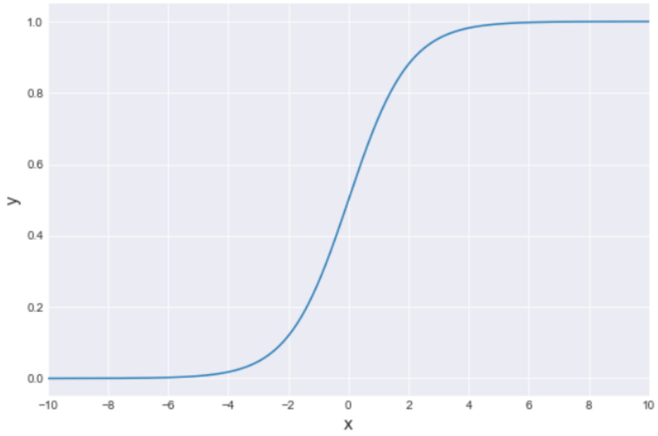
Step Function

$$y = \begin{cases} 0 & \text{if } x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$



Sigmoid Function

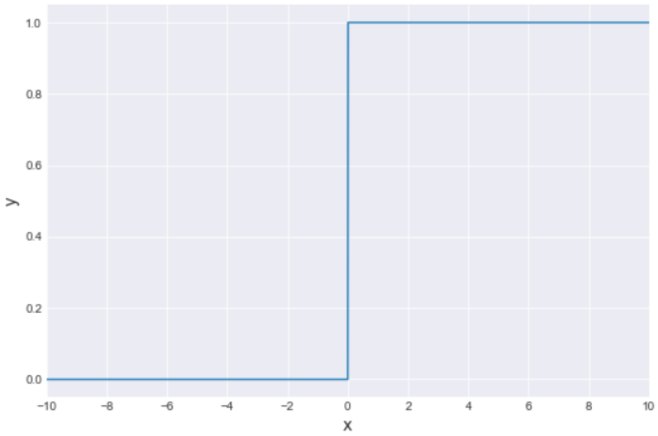
$$y = \frac{1}{1 + e^{-x}}$$



Activation Function

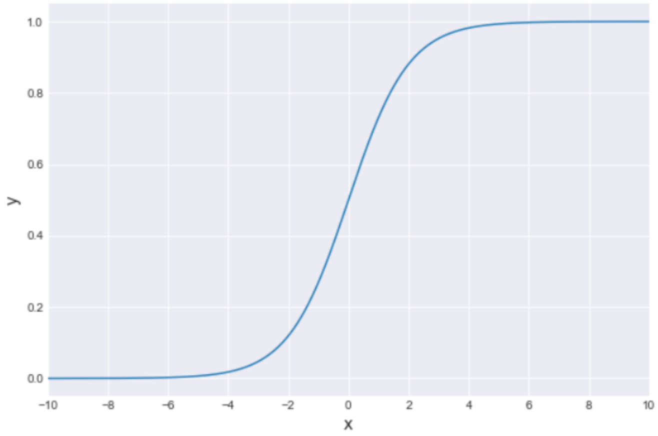
Step Function

$$y = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$



Sigmoid Function

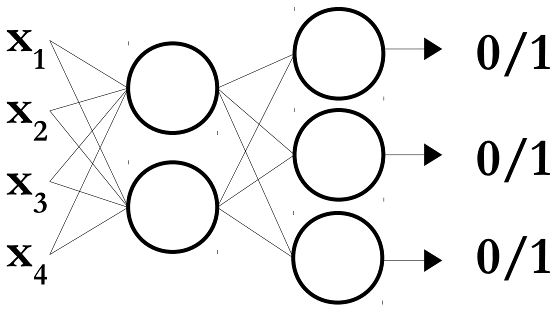
$$y = \frac{1}{1 + e^{-x}}$$



Deep Learning

Feedforward

How does the algorithm make a decision?



$$H_{in} = XW_1$$

$$H_{out} = step(H_{in})$$

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

$$O_{out} = step(O_{in})$$

How do you determine the right parameters for the algorithm?

Cost function/Loss function:

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

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

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

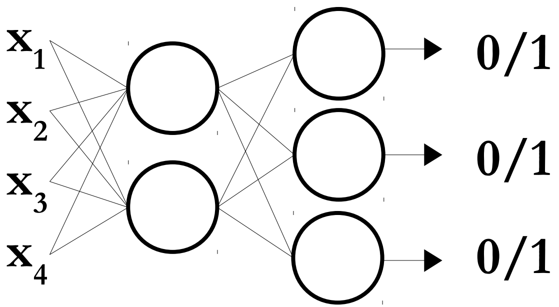
$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, W_2}$$

N: e * n

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:

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

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

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

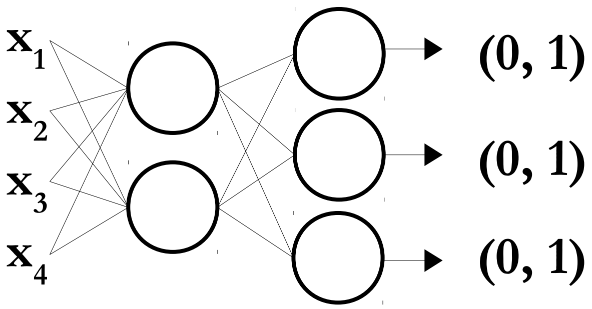
$$\frac{\partial\, MSE}{\partial\, W_2}(W_2) = \frac{\partial\, MSE}{\partial\, O_{out}} \frac{\partial\, O_{out}}{\partial\, O_{in}} \frac{\partial\, O_{in}}{\partial\, W_2}$$

N: e * n

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:

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

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