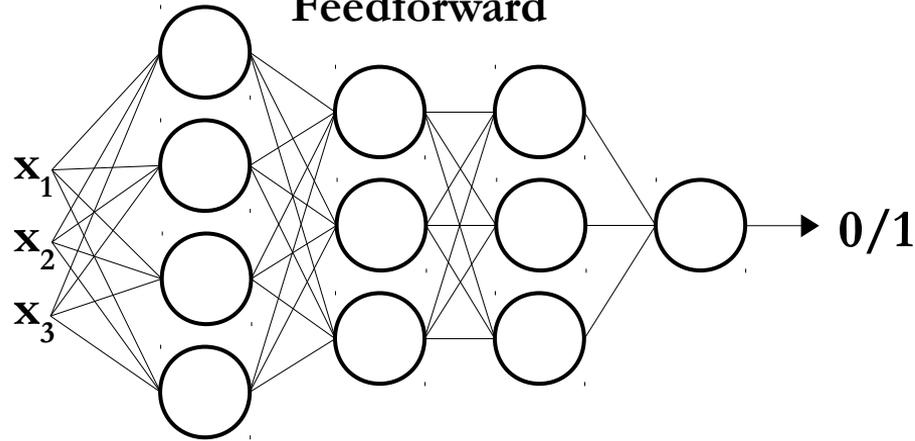


# Deep Learning

Feedforward



How does the algorithm make a decision?

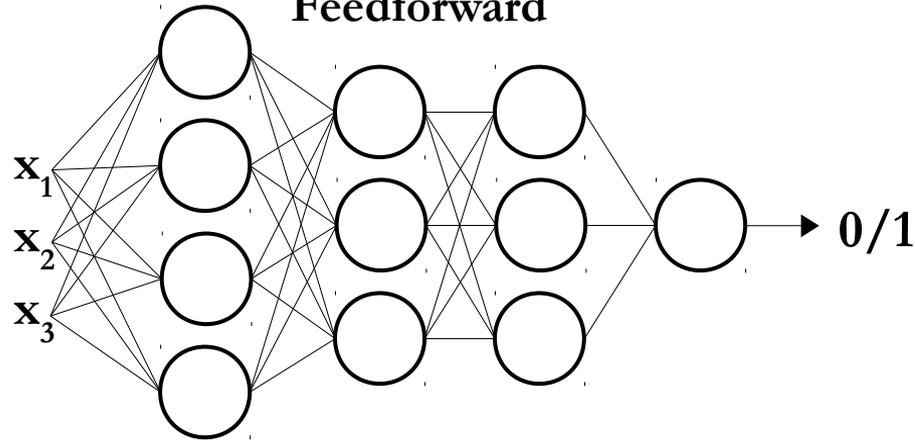


How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



How does the algorithm make a decision?



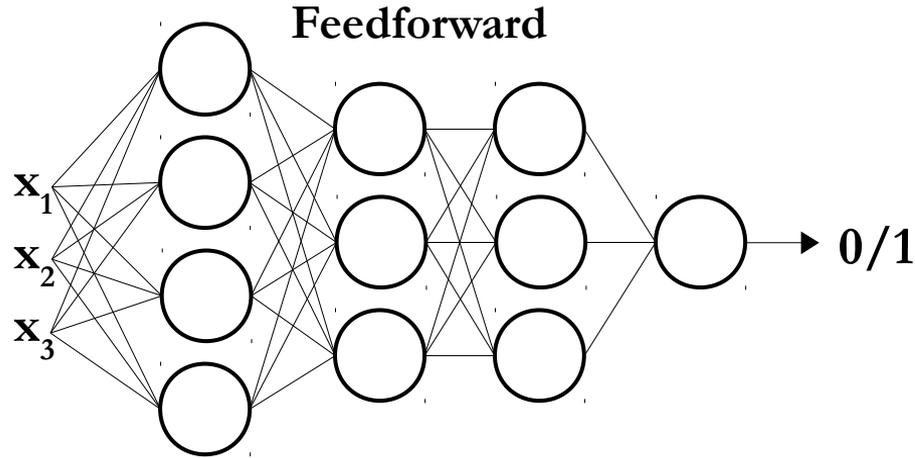
## Cost function/Loss function:

How do you determine the right parameters for the algorithm?



# Deep Learning

$$\underline{\text{Output} - \text{Label} = \text{Error}}$$



How does the algorithm make a decision?



Cost function/Loss function:

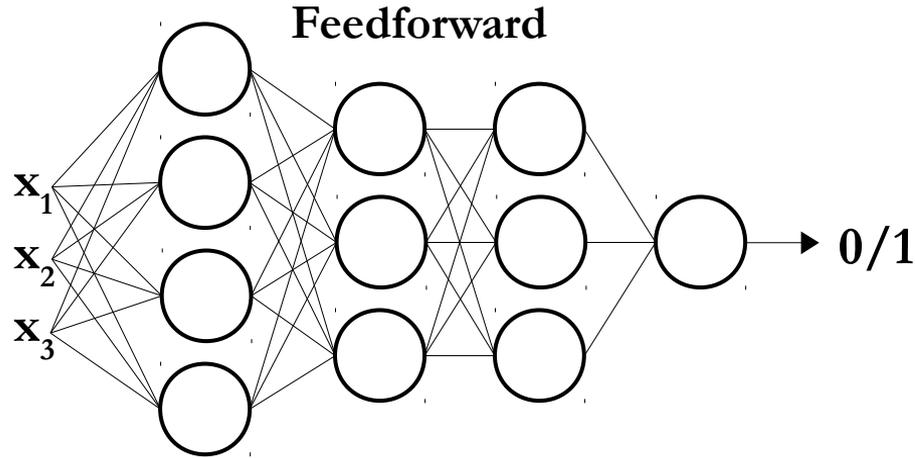
How do you determine the right parameters for the algorithm?



# Deep Learning

$$\text{Output} - \text{Label} = \text{Error}$$

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



## Cost function/Loss function:

How does the algorithm make a decision?



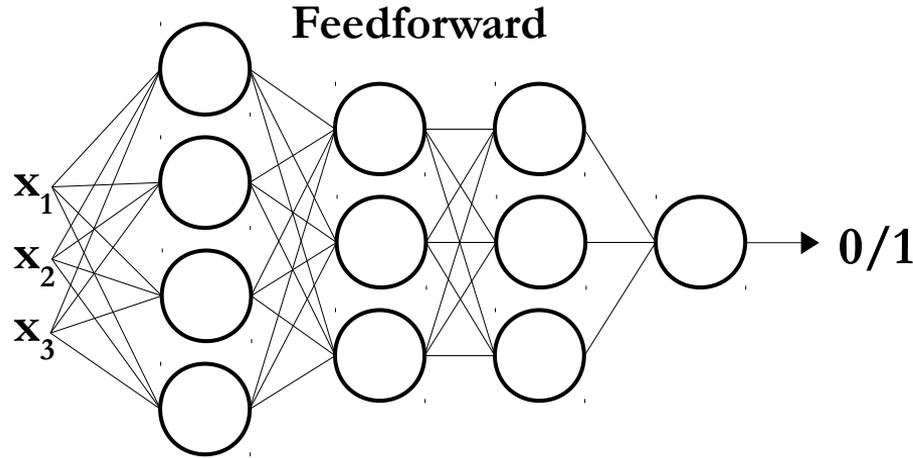
How do you determine the right parameters for the algorithm?



# Deep Learning

$$\underline{\text{Output}} - \underline{\text{Label}} = \underline{\text{Error}}$$

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$



How does the algorithm make a decision?



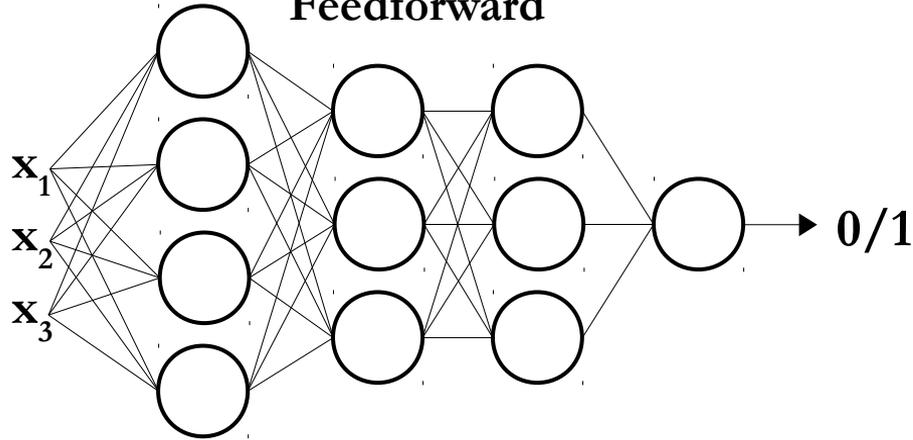
## Cost function/Loss function:

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



Output - Label = Error

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

0

How does the algorithm make a decision?



## Cost function/Loss function:

Sum of Errors

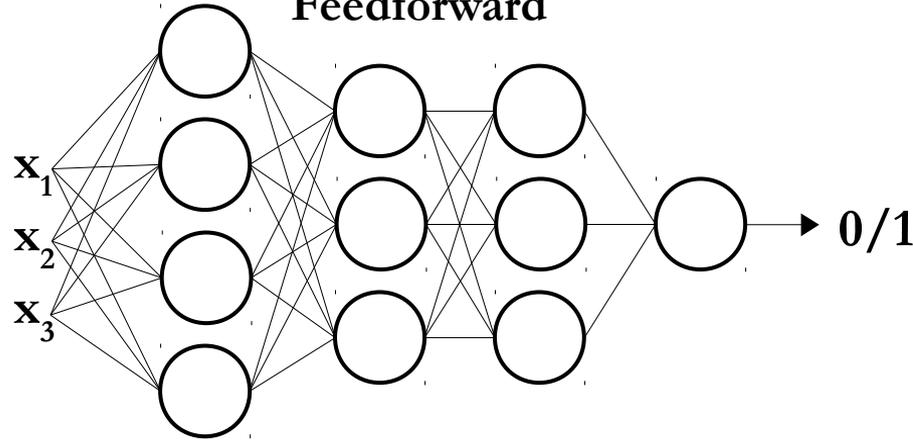
How do you determine the right parameters for the algorithm?



# Deep Learning

$$\frac{(\text{Output} - \text{Label})^2}{2} = \text{Squared Error}$$

Feedforward



$$\left( \begin{bmatrix} [1] \\ [0] \\ [1] \\ [0] \end{bmatrix} - \begin{bmatrix} [1] \\ [0] \\ [0] \\ [1] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [0] \\ [0] \\ [1] \\ [1] \end{bmatrix}}{2}$$

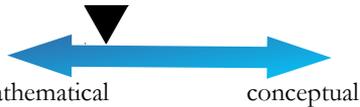
How does the algorithm make a decision?



## Cost function/Loss function:

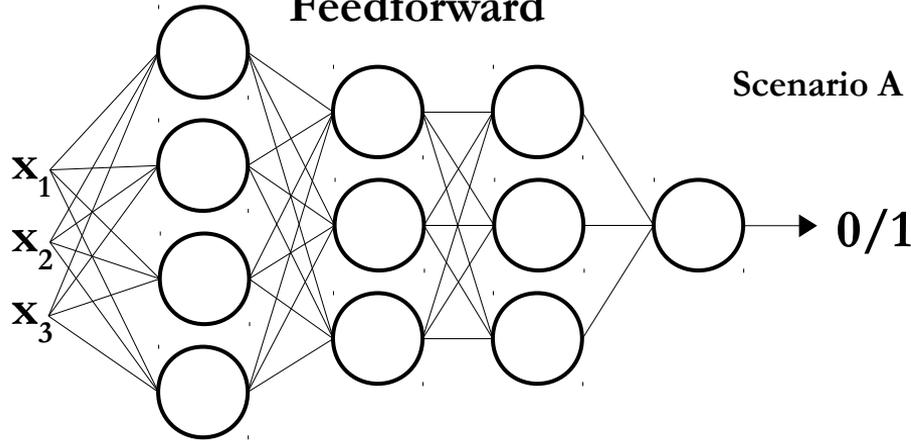
Sum of Squared Errors

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$\frac{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1] \\ [0] \\ [1] \\ [0] \end{bmatrix} - \begin{bmatrix} [1] \\ [0] \\ [0] \\ [1] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [0] \\ [0] \\ [1] \\ [1] \end{bmatrix}}{\underline{2}}$$

$$\frac{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1] \\ [1] \\ [1] \\ [0] \\ [1] \\ [0] \\ [0] \\ [0] \\ [1] \\ [0] \end{bmatrix} - \begin{bmatrix} [1] \\ [1] \\ [1] \\ [0] \\ [1] \\ [0] \\ [1] \\ [0] \\ [0] \\ [1] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [0] \\ [0] \\ [0] \\ [0] \\ [0] \\ [0] \\ [1] \\ [1] \\ [1] \\ [1] \end{bmatrix}}{\underline{3}}$$

## Cost function/Loss function:

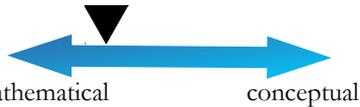
Sum of Squared Errors

Scenario B

How does the algorithm make a decision?

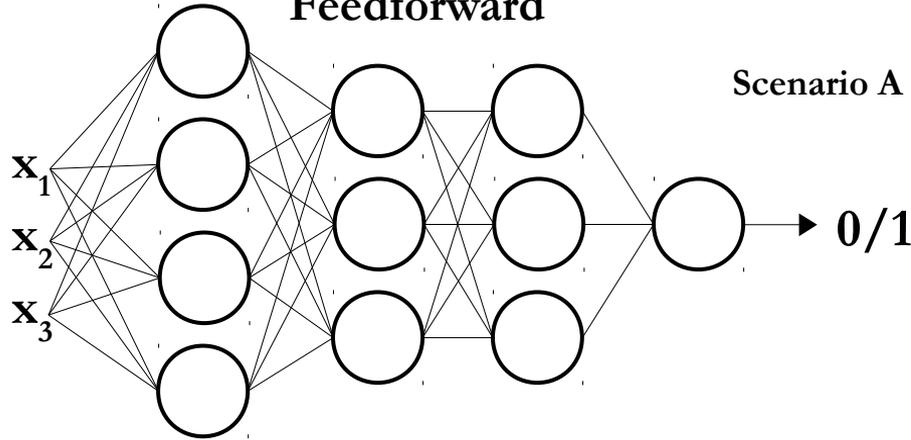


How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



## Cost function/Loss function:

Mean Squared Error

Scenario B

$$\frac{(\text{Output} - \text{Label})^2}{2} = \text{Squared Error}$$

$$\left( \begin{matrix} [1] \\ [0] \\ [1] \\ [0] \end{matrix} - \begin{matrix} [1] \\ [0] \\ [0] \\ [1] \end{matrix} \right)^2 = \frac{\begin{matrix} [0] \\ [0] \\ [1] \\ [1] \end{matrix}}{2} = \underline{0.5}$$

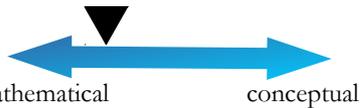
$$\frac{(\text{Output} - \text{Label})^2}{3} = \text{Squared Error}$$

$$\left( \begin{matrix} [1] \\ [1] \\ [1] \\ [0] \\ [1] \\ [0] \\ [0] \\ [0] \\ [1] \\ [0] \end{matrix} - \begin{matrix} [1] \\ [1] \\ [1] \\ [0] \\ [1] \\ [0] \\ [0] \\ [1] \\ [0] \\ [1] \end{matrix} \right)^2 = \frac{\begin{matrix} [0] \\ [0] \\ [0] \\ [0] \\ [0] \\ [0] \\ [1] \\ [1] \\ [1] \\ [1] \end{matrix}}{3} = \underline{0.3}$$

How does the algorithm make a decision?



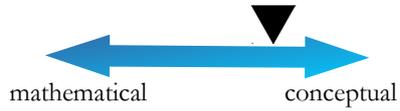
How do you determine the right parameters for the algorithm?



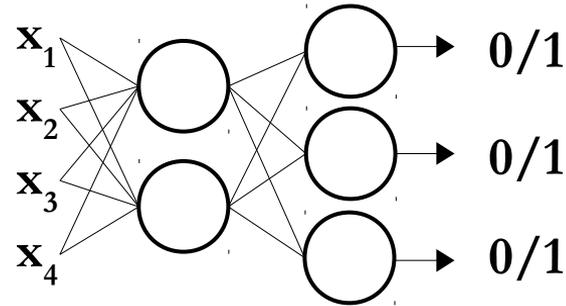
# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

How does the algorithm make a decision?



Feedforward



Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?



# Deep Learning

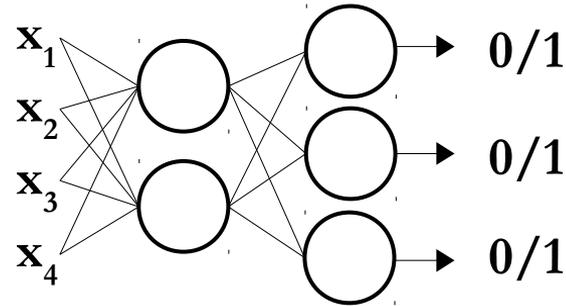
$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

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

How does the algorithm make a decision?



Feedforward



Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?

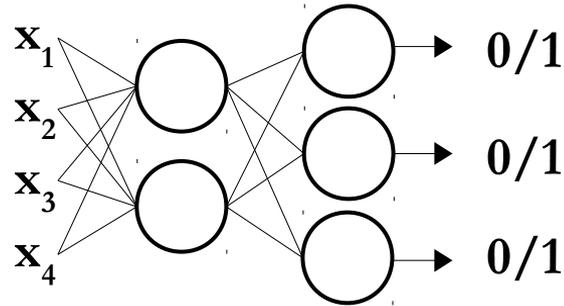


# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}$$

## Feedforward



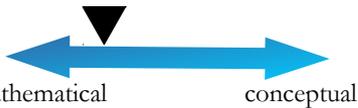
How does the algorithm make a decision?



## Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?

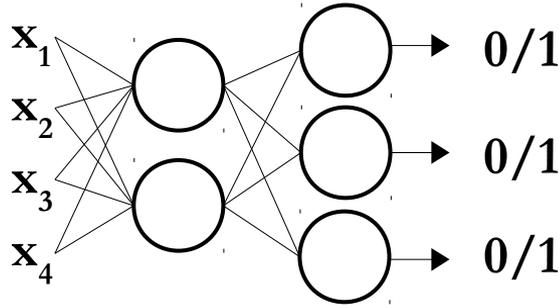


# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]}$$

## Feedforward



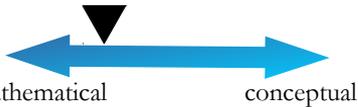
How does the algorithm make a decision?



## Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?

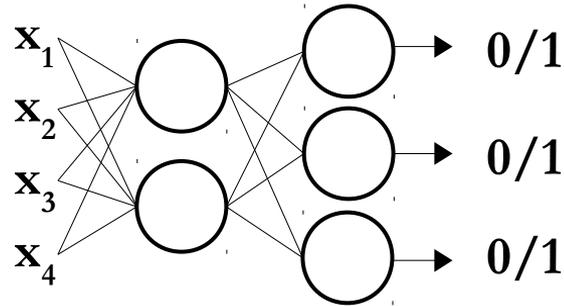


# Deep Learning

$$\frac{(\text{Output} - \text{Label})^2}{2} = \text{Squared Error}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \rightarrow 6$$

## Feedforward



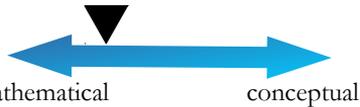
How does the algorithm make a decision?



## Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?

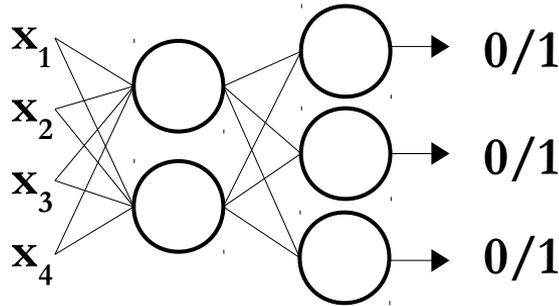


# Deep Learning

$$\frac{(\text{Output} - \text{Label})^2}{\text{Label}} = \text{Squared Error}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

## Feedforward



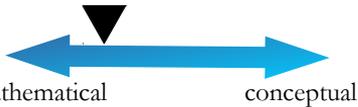
How does the algorithm make a decision?



## Cost function/Loss function:

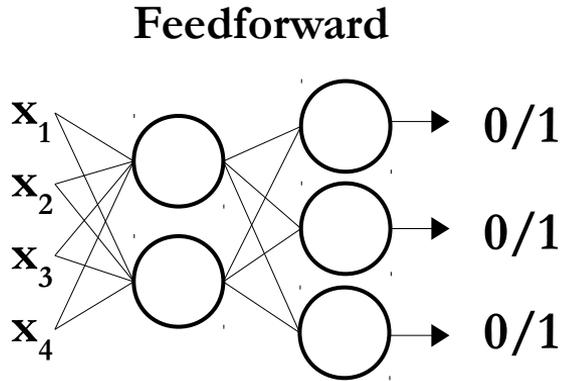
Mean Squared Error

How do you determine the right parameters for the algorithm?



# Deep Learning

How does the algorithm make a decision?



## Cost function/Loss function:

Mean Squared Error

How do you determine the right parameters for the algorithm?



$$\frac{(\text{Output} - \text{Label})^2}{\text{Label}} = \text{Squared Error}$$

Scenario C

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

$$\frac{(\text{Output} - \text{Label})^2}{\text{Label}} = \text{Squared Error}$$

Scenario A

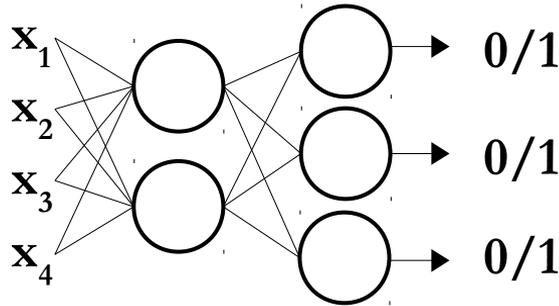
$$\left( \begin{bmatrix} [1] \\ [0] \\ [1] \\ [0] \end{bmatrix} - \begin{bmatrix} [1] \\ [0] \\ [0] \\ [1] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [0] \\ [0] \\ [1] \\ [1] \end{bmatrix}}{2} : 4 = \underline{0.5}$$

# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

## Feedforward



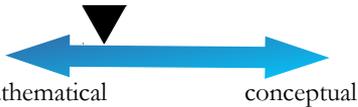
How does the algorithm make a decision?



## Cost function/Loss function:

$$\text{Mean Squared Error} = \sum (O - Y)^2$$

How do you determine the right parameters for the algorithm?



# Deep Learning

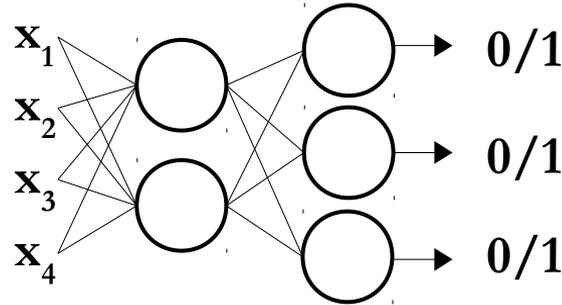
$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

How does the algorithm make a decision?



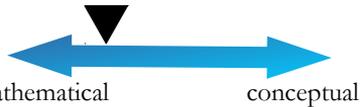
Feedforward



Cost function/Loss function:

$$\text{Mean Squared Error} = (\text{O} - \text{Y})^2$$

How do you determine the right parameters for the algorithm?



# Deep Learning

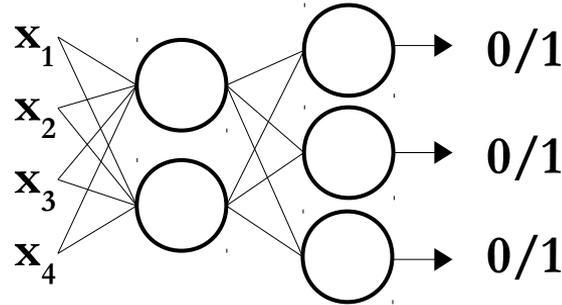
$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

How does the algorithm make a decision?



Feedforward



Cost function/Loss function:

$$\text{Mean Squared Error} = \sum_c (O_c - Y_c)^2$$

How do you determine the right parameters for the algorithm?

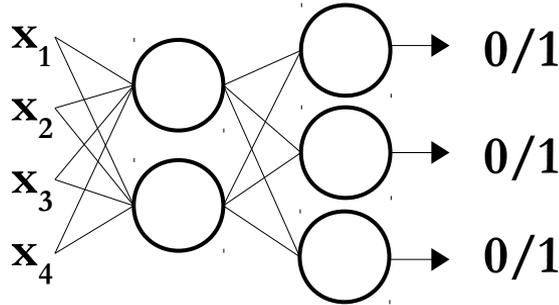


# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \frac{\begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix}}{[2 \ 2 \ 2]} \Rightarrow 6 : 12 = \underline{0.5}$$

## Feedforward



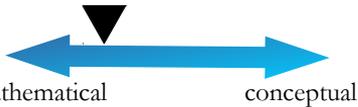
How does the algorithm make a decision?



## Cost function/Loss function:

$$\text{Mean Squared Error} = \sum_n \sum_c (O_{c,n} - Y_{c,n})^2$$

How do you determine the right parameters for the algorithm?



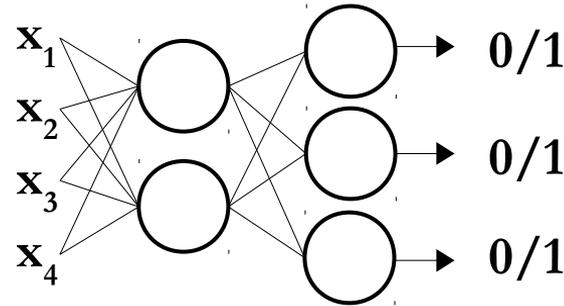
# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix} \quad \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

↓  
6 : 12 = 0.5

## Feedforward



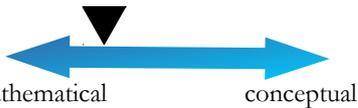
How does the algorithm make a decision?



## Cost function/Loss function:

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

How do you determine the right parameters for the algorithm?



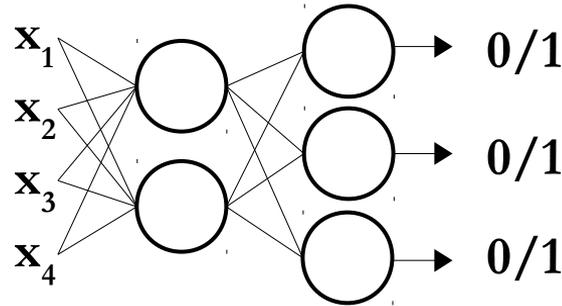
# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

↓  
6 : 12 = 0.5

## Feedforward



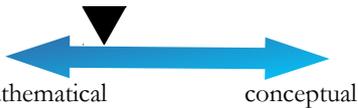
How does the algorithm make a decision?



## Cost function/Loss function:

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

How do you determine the right parameters for the algorithm?



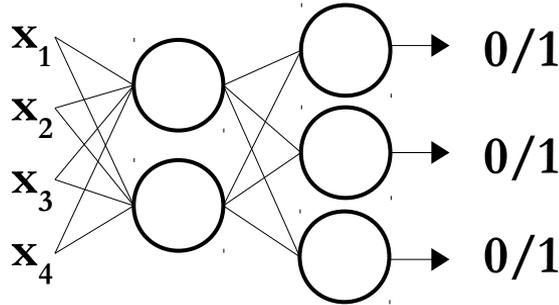
# Deep Learning

$$\underline{(\text{Output} - \text{Label})^2 = \text{Squared Error}}$$

$$\left( \begin{bmatrix} [1 & 0 & 0] \\ [0 & 1 & 0] \\ [0 & 1 & 0] \\ [0 & 0 & 1] \end{bmatrix} - \begin{bmatrix} [0 & 0 & 1] \\ [0 & 1 & 0] \\ [1 & 0 & 0] \\ [0 & 1 & 0] \end{bmatrix} \right)^2 = \begin{bmatrix} [1 & 0 & 1] \\ [0 & 0 & 0] \\ [1 & 1 & 0] \\ [0 & 1 & 1] \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

↓  
6 : 12 = 0.5

## Feedforward



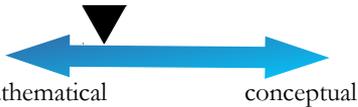
How does the algorithm make a decision?



## Cost function/Loss function:

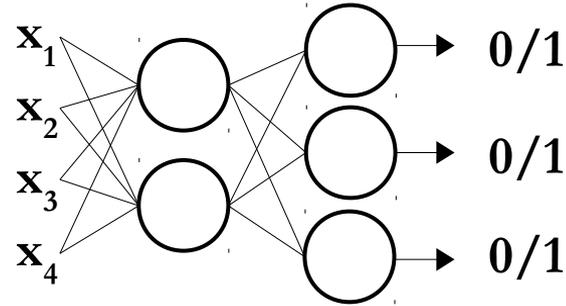
$$\text{Mean Squared Error} = \frac{1}{N} \sum_e \sum_n (O_{e,n} - Y_{e,n})^2$$

How do you determine the right parameters for the algorithm?

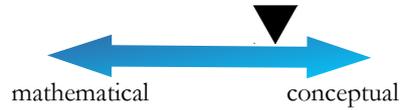


# Deep Learning

## Feedforward



How does the algorithm make a decision?



## Cost function/Loss function:

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

How do you determine the right parameters for the algorithm?

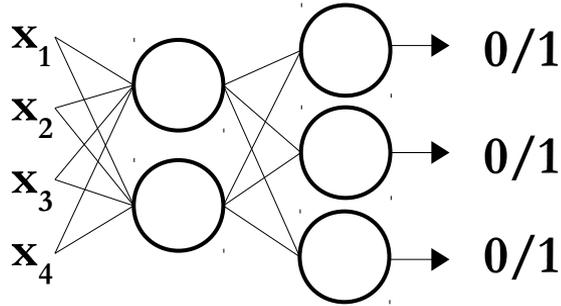


# Deep Learning

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

Cost function/Loss function:

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

How do you determine the right parameters for the algorithm?

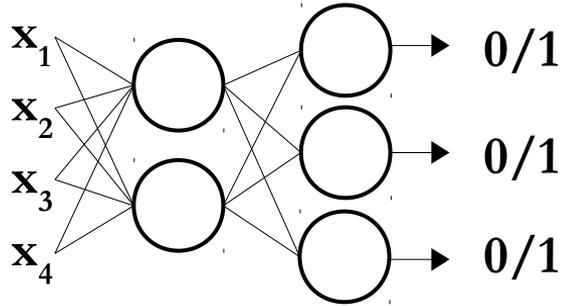


# Deep Learning

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

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

Cost function/Loss function:

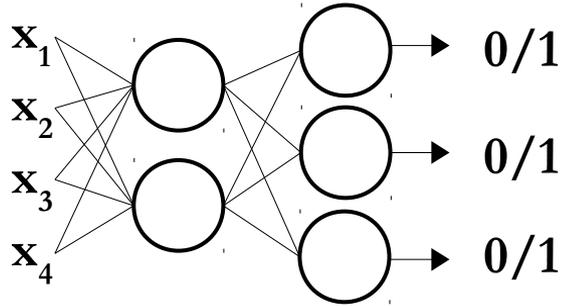
$$\text{Mean Squared Error} = \frac{1}{N} \sum_e \sum_n (O_{e,n} - Y_{e,n})^2$$

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$H_{in} = XW_1$$

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

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

How does the algorithm make a decision?



## Cost function/Loss function:

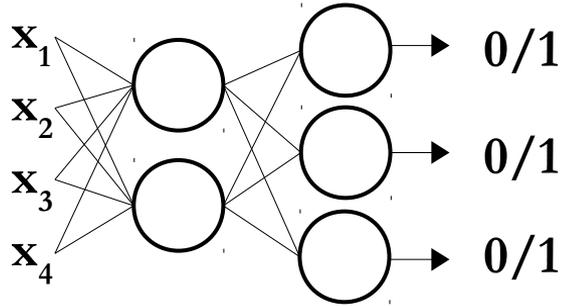
$$\text{Mean Squared Error} = \frac{1}{N} \sum_e \sum_n (O_{e,n} - Y_{e,n})^2$$

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



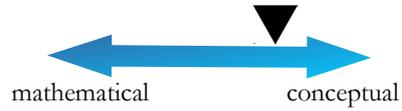
$$H_{in} = XW_1$$

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

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

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

How does the algorithm make a decision?



## Cost function/Loss function:

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

How do you determine the right parameters for the algorithm?

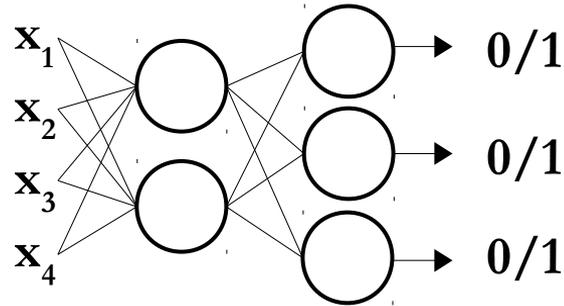


# Deep Learning

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

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

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

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

Cost function/Loss function:

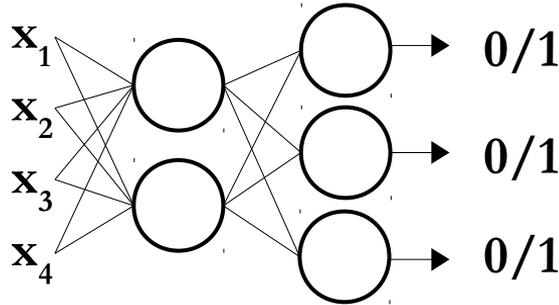
$$\text{Mean Squared Error} = \frac{1}{N} \sum_e \sum_n (O_{out\ e,n} - Y_{e,n})^2$$

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$H_{in} = XW_1$$

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

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

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



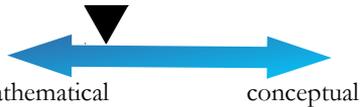
## Cost function/Loss function:

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

How does the algorithm make a decision?

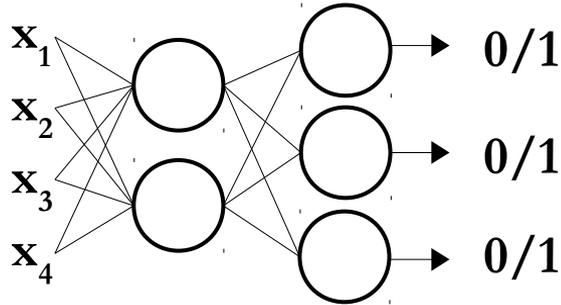


How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$H_{in} = XW_1$$

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

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

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



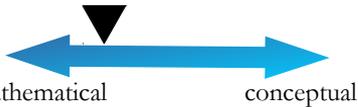
## Cost function/Loss function:

Mean Squared Error =  $\frac{1}{N} \sum_e \sum_n (O_{out\ e,n} - Y_{e,n})^2$   $\rightarrow$  MSE =  $XW_1$

How does the algorithm make a decision?



How do you determine the right parameters for the algorithm?

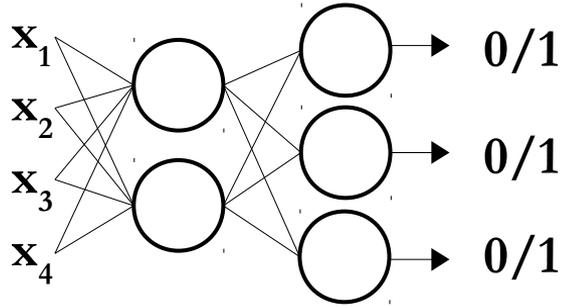


# Deep Learning

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

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

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

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



Cost function/Loss function:

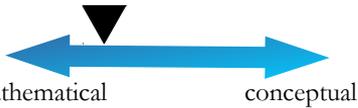
$$\text{Mean Squared Error} = \frac{1}{N} \sum_e \sum_n (O_{out\ e,n} - Y_{e,n})^2$$



MSE =

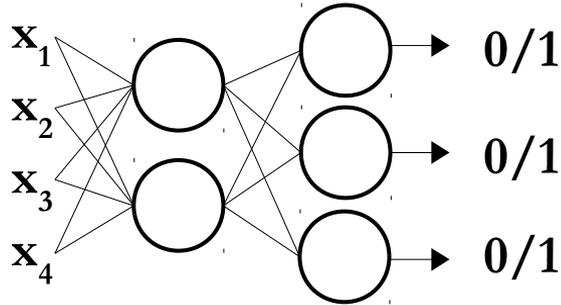
step(XW<sub>1</sub>)

How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$H_{in} = XW_1$$

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

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

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



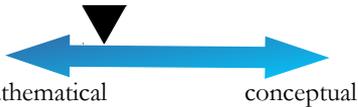
## Cost function/Loss function:

Mean Squared Error =  $\frac{1}{N} \sum_e \sum_n (O_{out\ e,n} - Y_{e,n})^2$   $\rightarrow$  MSE =  $\text{step}(XW_1)W_2$

How does the algorithm make a decision?

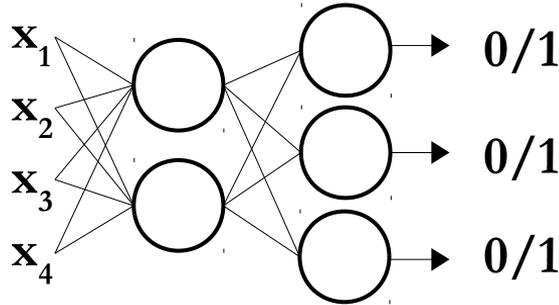


How do you determine the right parameters for the algorithm?



# Deep Learning

## Feedforward



$$H_{in} = XW_1$$

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

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

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



## Cost function/Loss function:

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



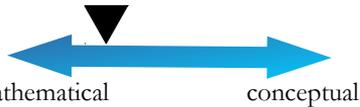
MSE =

$$\text{step}(\text{step}(XW_1)W_2)$$

How does the algorithm make a decision?



How do you determine the right parameters for the algorithm?

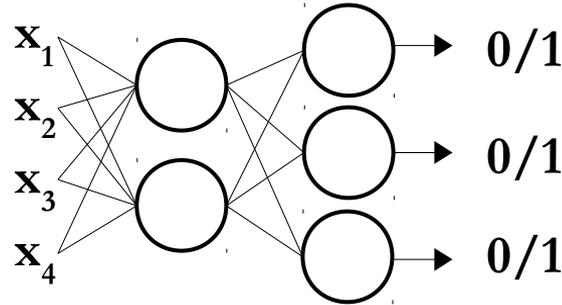


# Deep Learning

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

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

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

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



Cost function/Loss function:

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

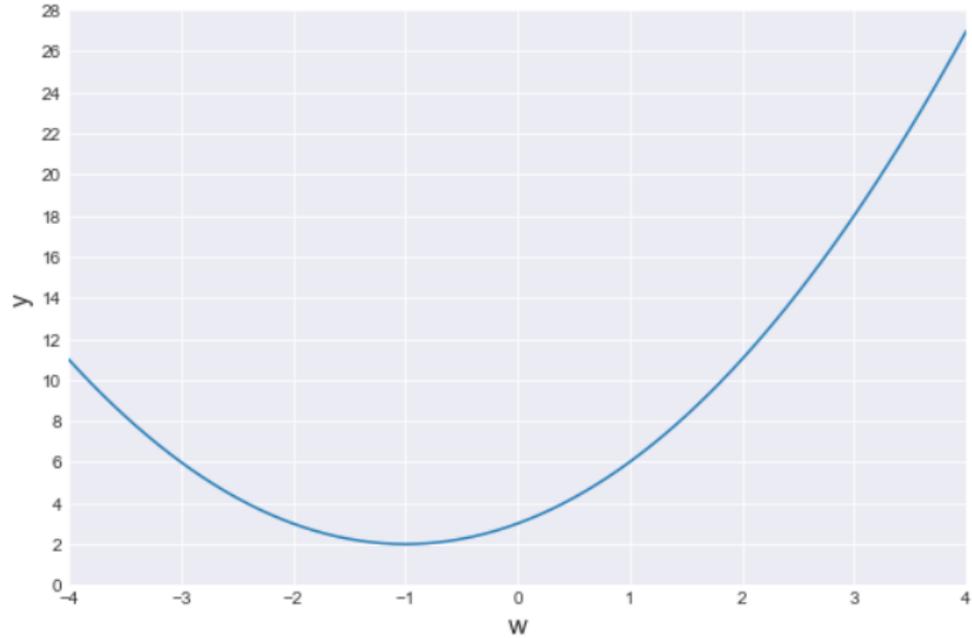


$$\text{MSE} = \frac{1}{N} \sum_e \sum_n (\text{step}(\text{step}(XW_1)W_2)_{e,n} - Y_{e,n})^2$$

How do you determine the right parameters for the algorithm?

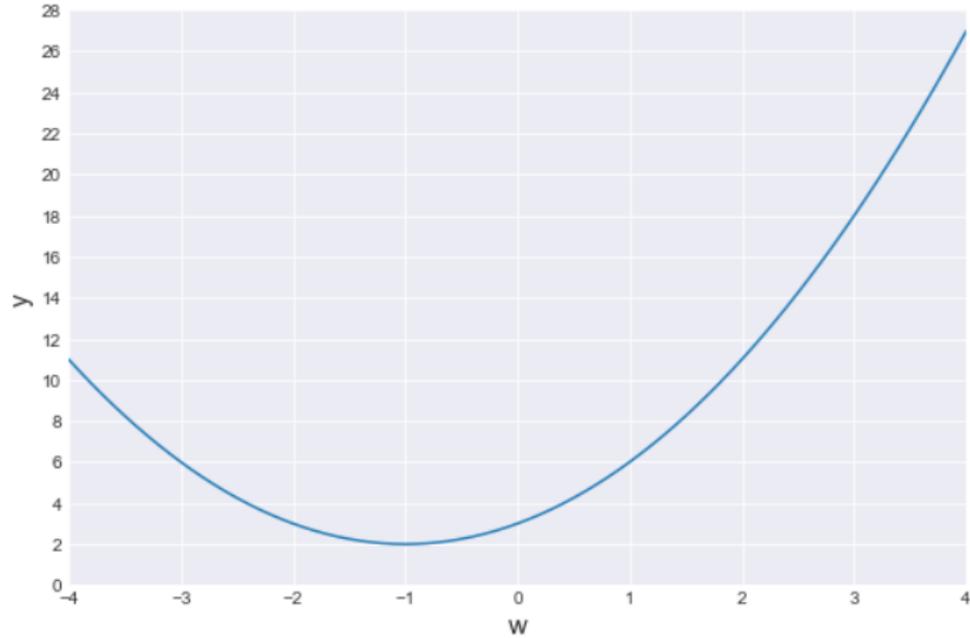


# Derivative



$$y = f(w) = w^2 + 2w + 3$$

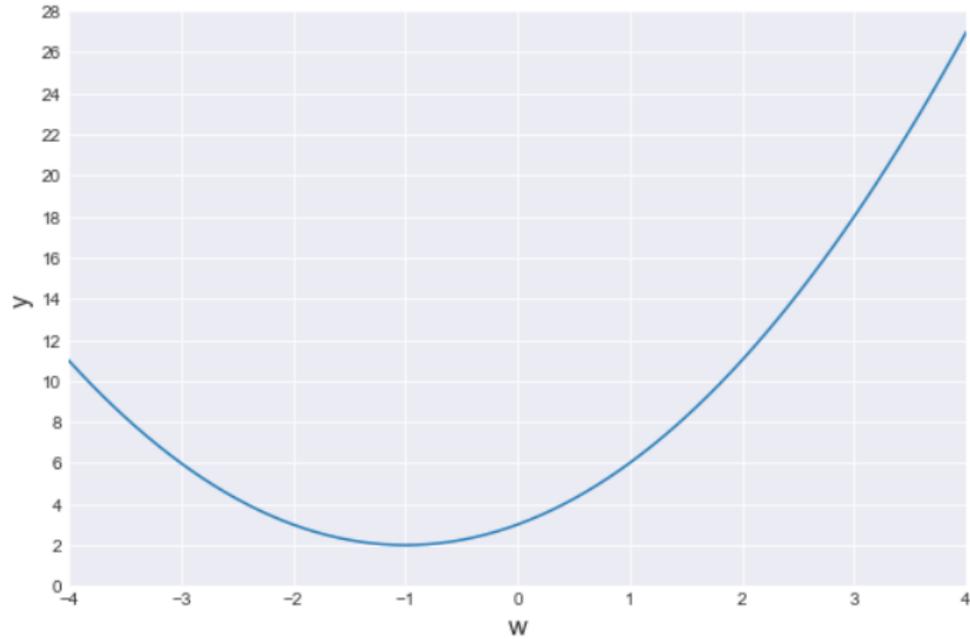
# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) =$$

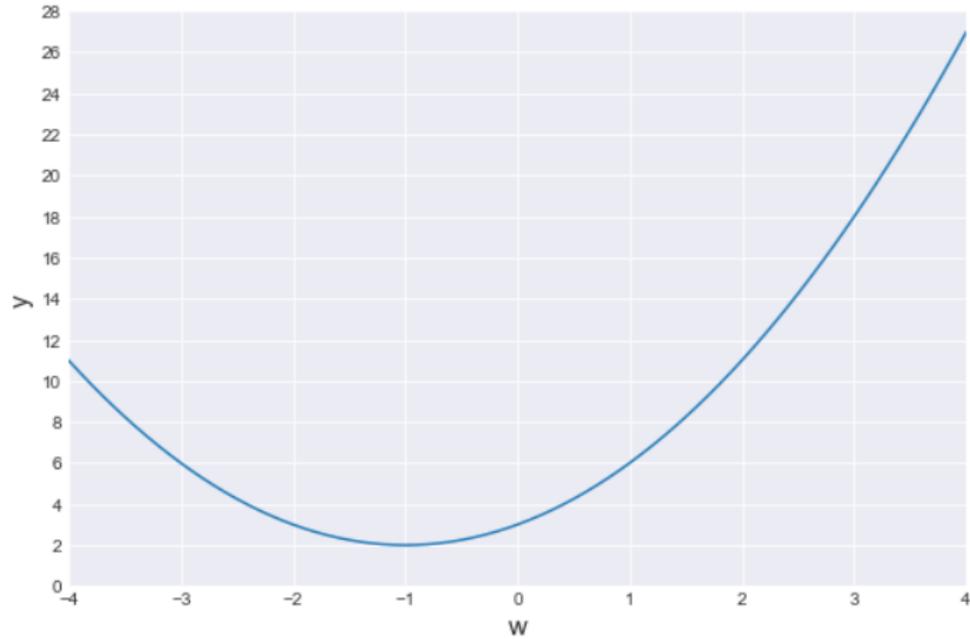
# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w$$

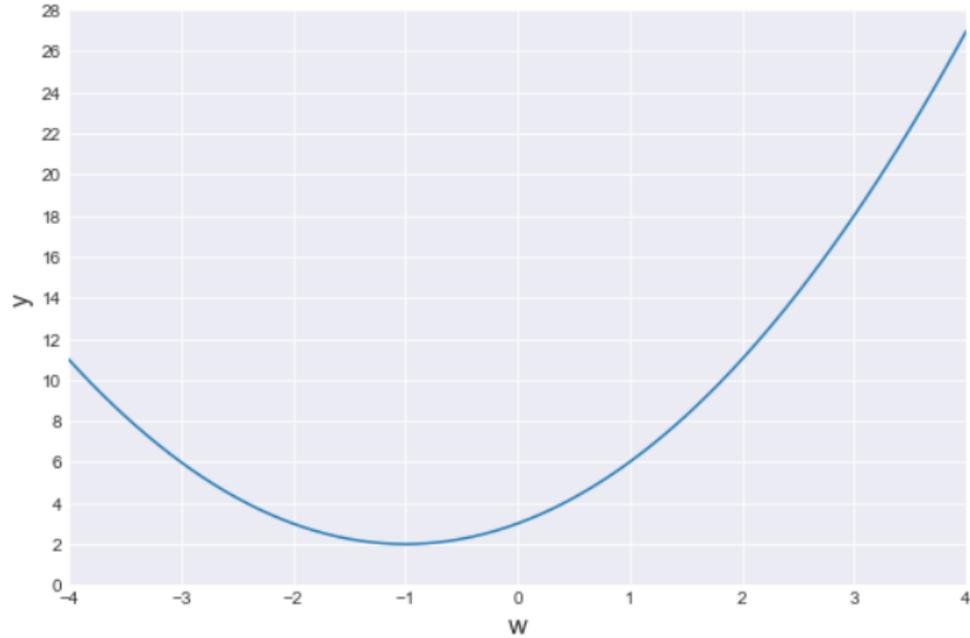
# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2w^0$$

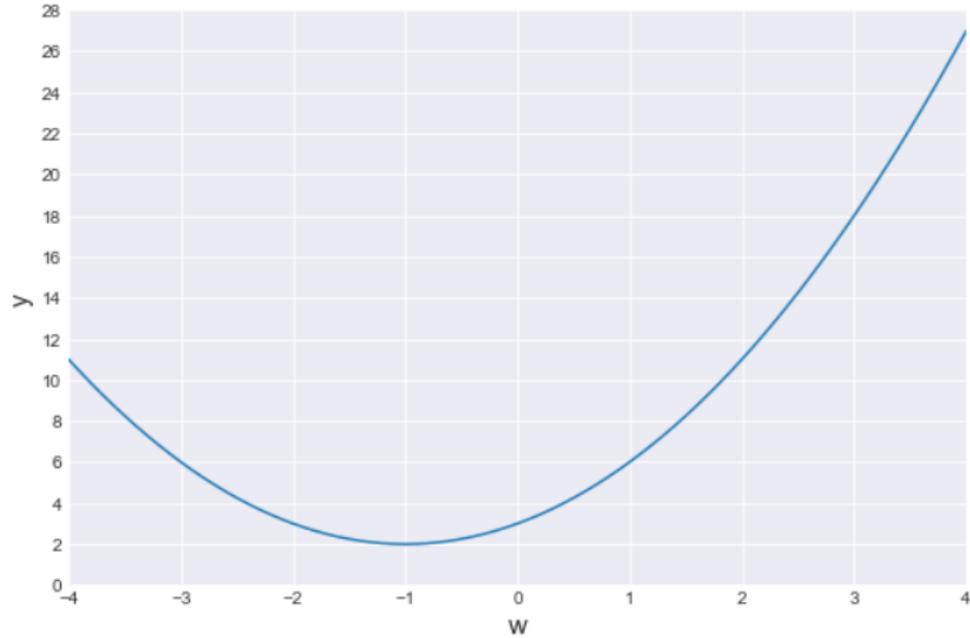
# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

# Derivative

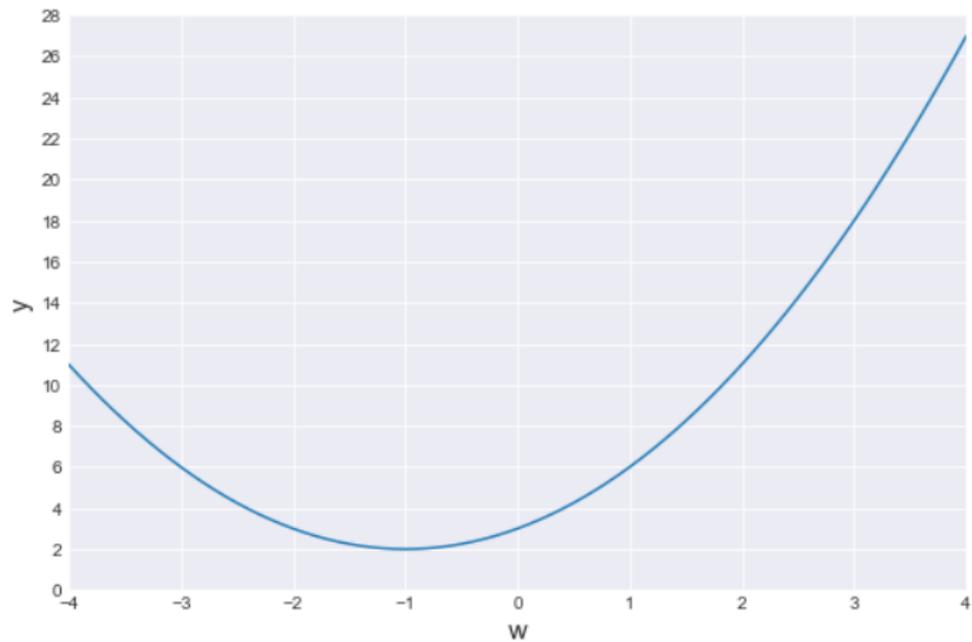


w = 1:

$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

# Derivative



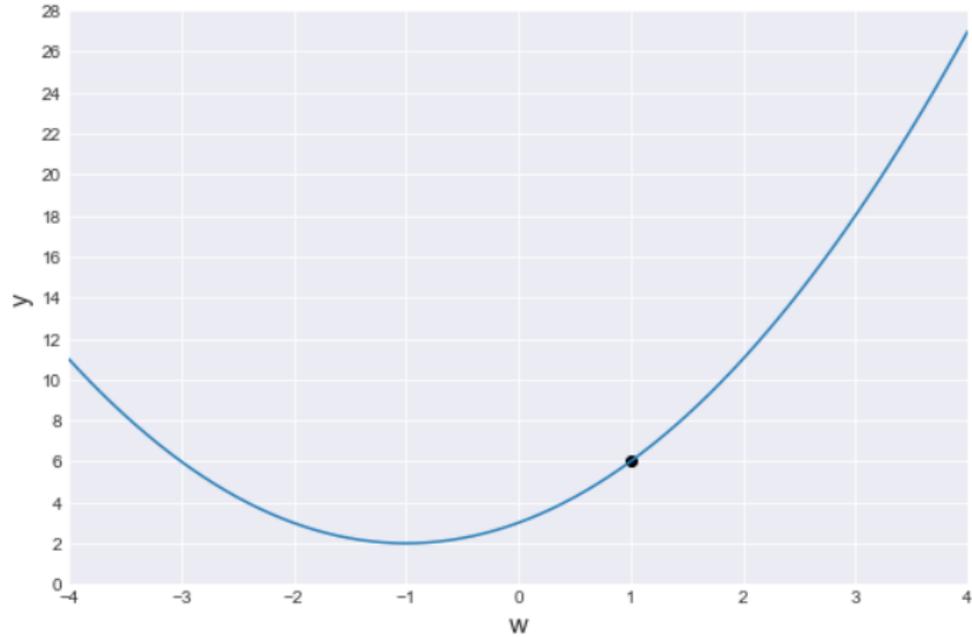
$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$w = 1$ :

$$y(1) = 6$$

# Derivative



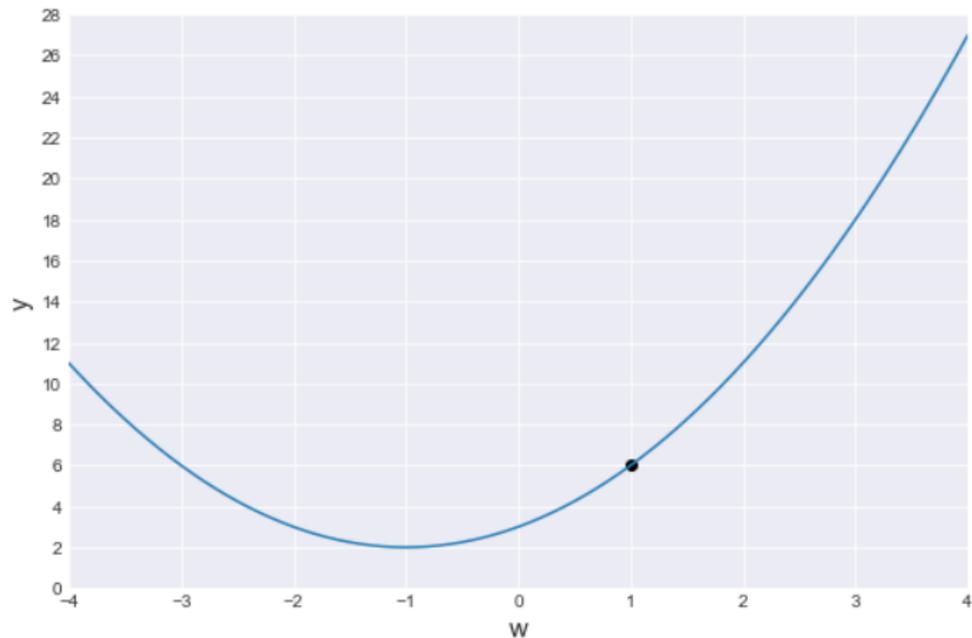
$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = 1:}$$

$$y(1) = 6$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

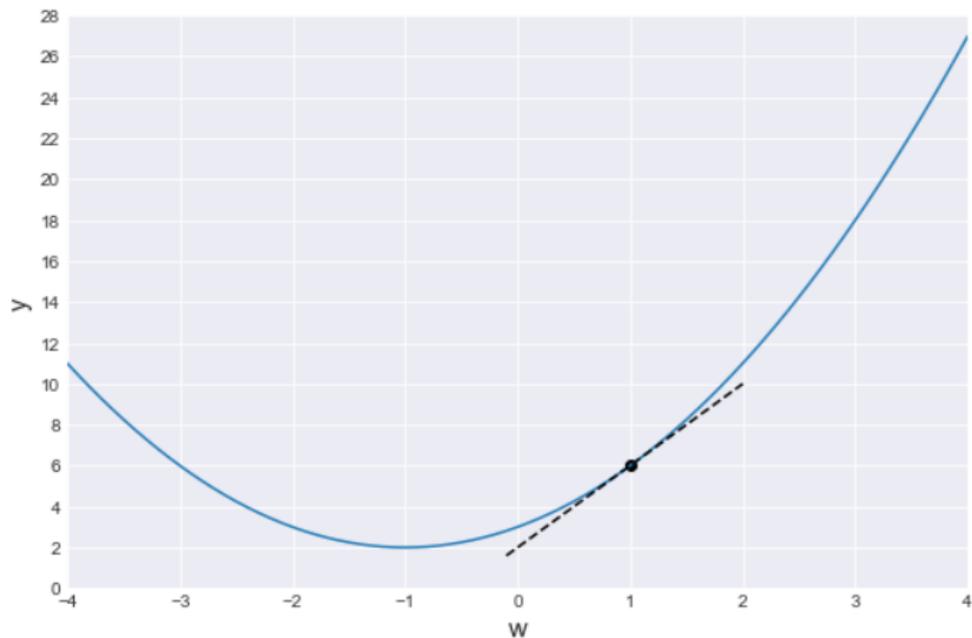
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = 1:}$$

$$y(1) = 6$$

$$\frac{dy}{dw}(1) = 4$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

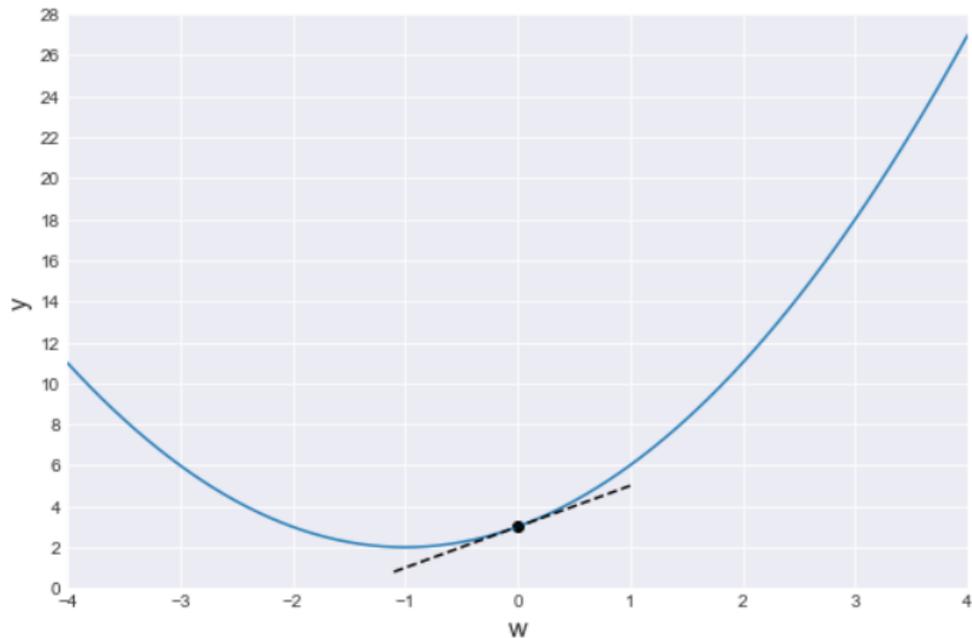
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = 1:}$$

$$y(1) = 6$$

$$\frac{dy}{dw}(1) = 4$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

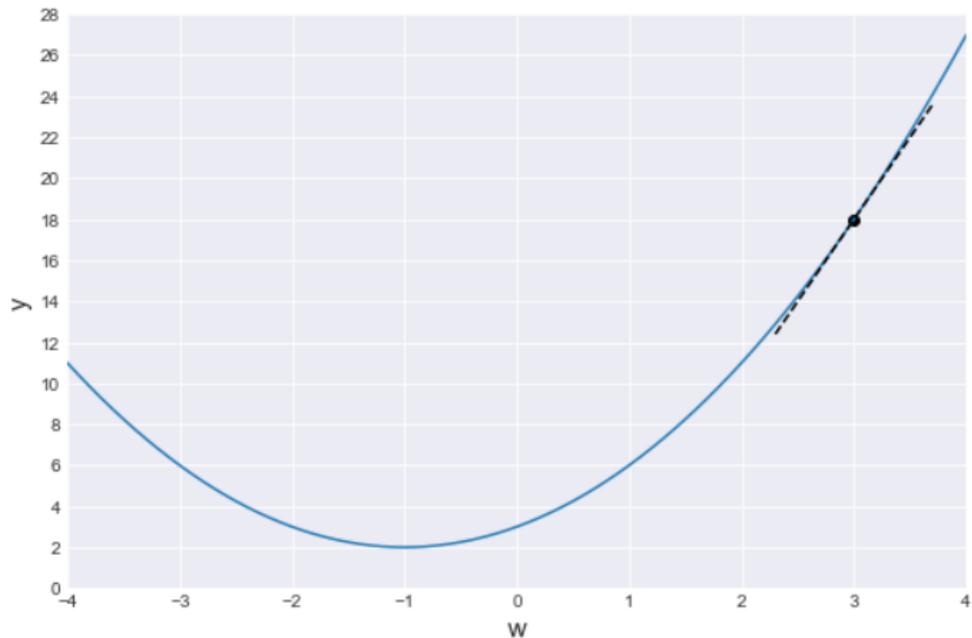
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = 0:}$$

$$y(0) = 3$$

$$\frac{dy}{dw}(0) = 2$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

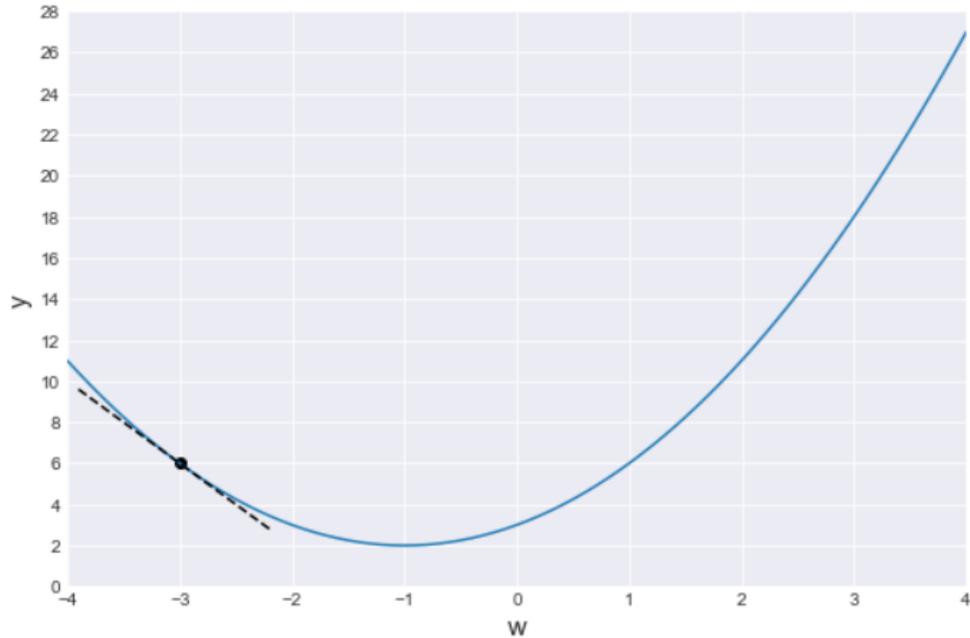
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = 3:}$$

$$y(3) = 18$$

$$\frac{dy}{dw}(3) = 8$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

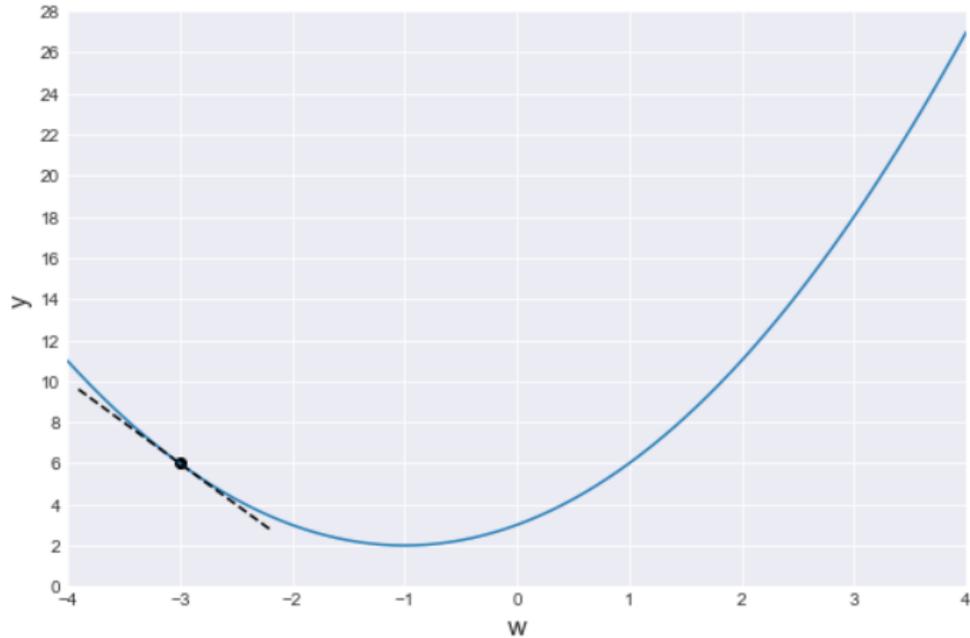
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

$$\underline{w = -3:}$$

$$y(3) = 6$$

$$\frac{dy}{dw}(3) = -4$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

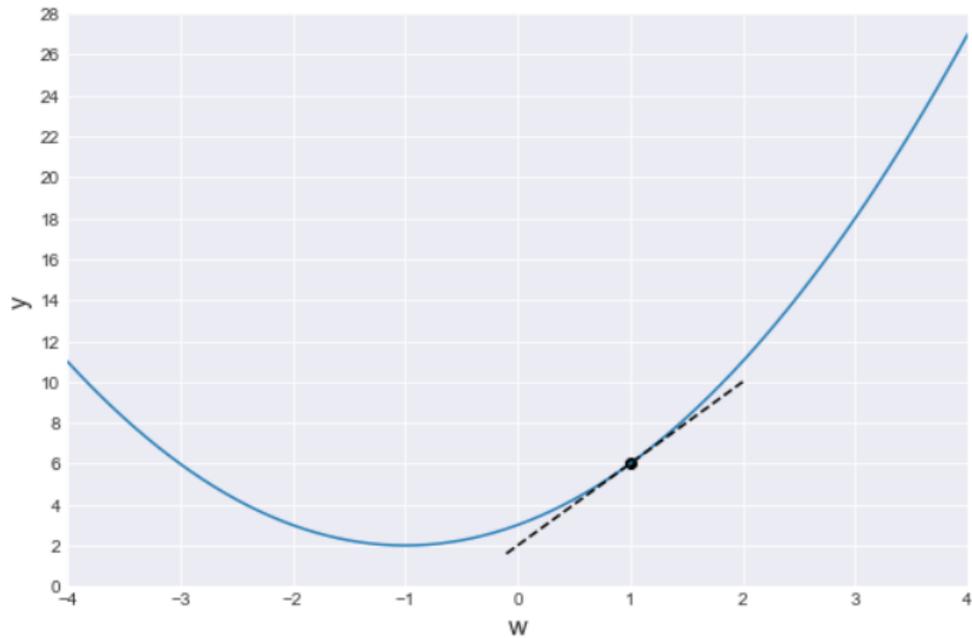
$$\underline{w = -3:}$$

$$y(3) = 6$$

$$\frac{dy}{dw}(3) = -4$$

direction

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

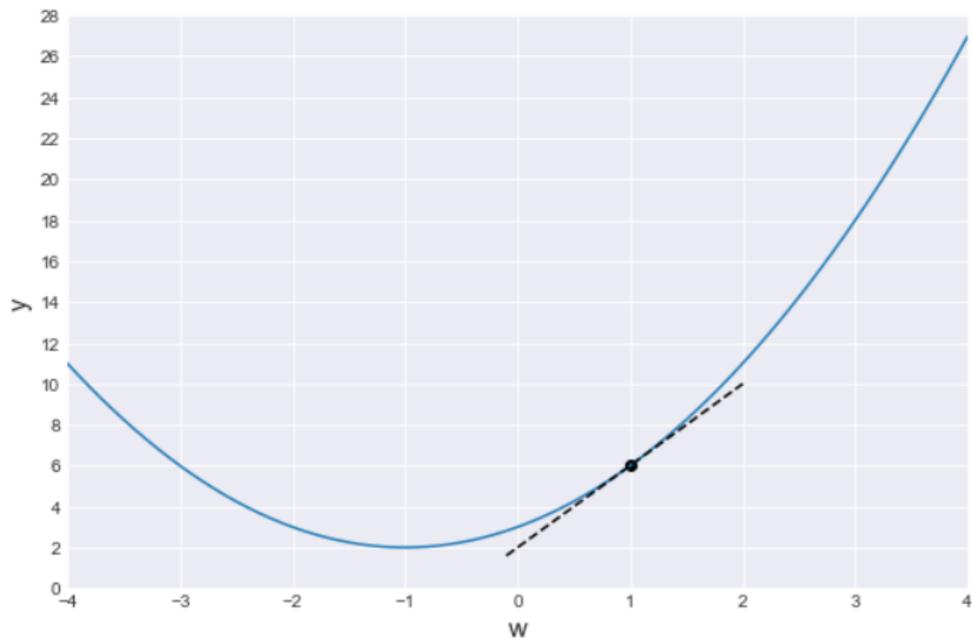
$$\underline{w = 1:}$$

$$y(1) = 6$$

$$\frac{dy}{dw} (1) = + 4$$

direction

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

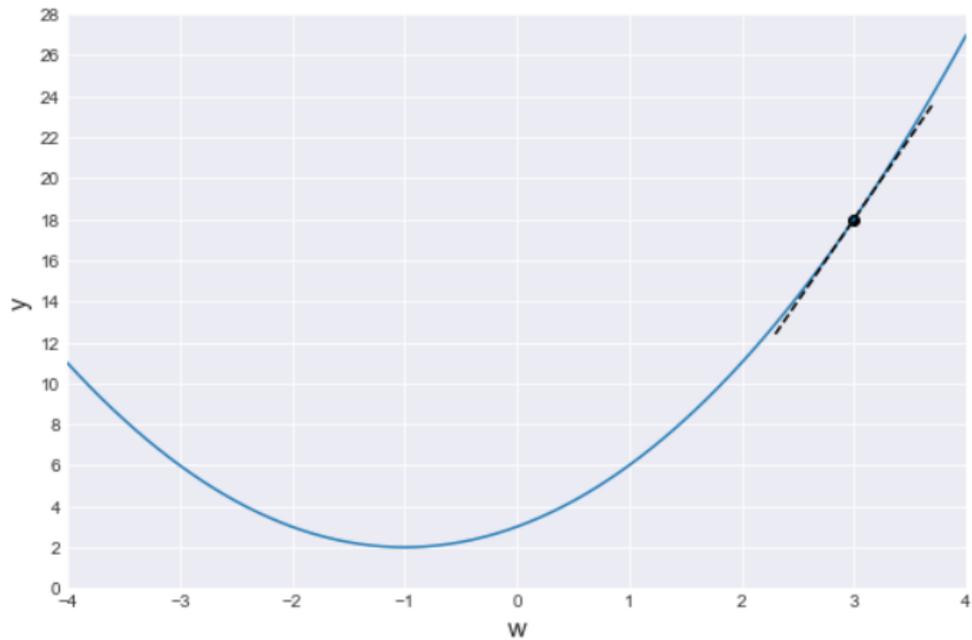
$$\underline{w = 1:}$$

$$y(1) = 6$$

$$\frac{dy}{dw}(1) = +4$$

direction      amount

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

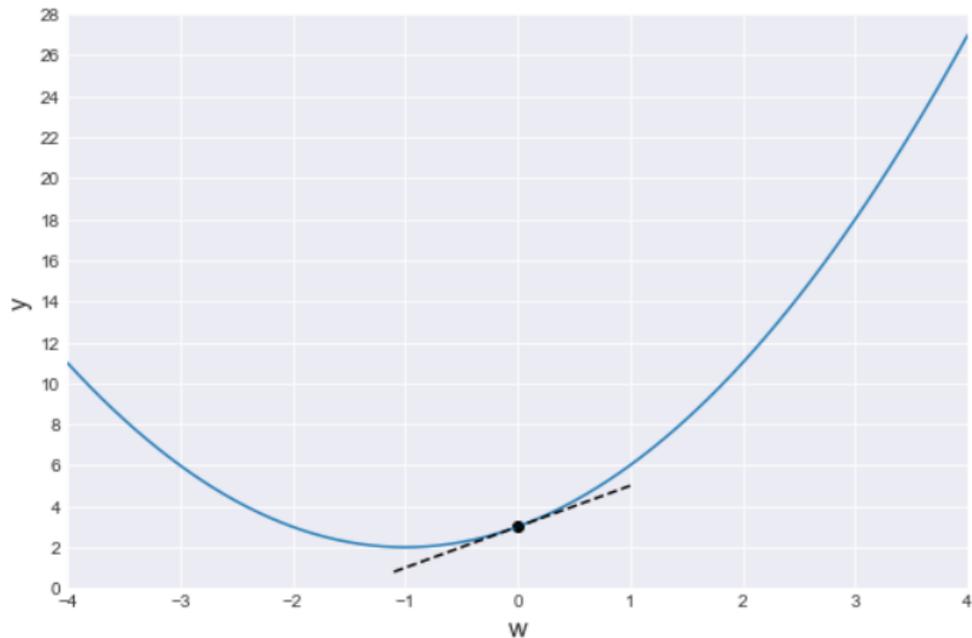
$w = 3:$

$$y(3) = 18$$

$$\frac{dy}{dw}(3) = +8$$

direction      amount

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

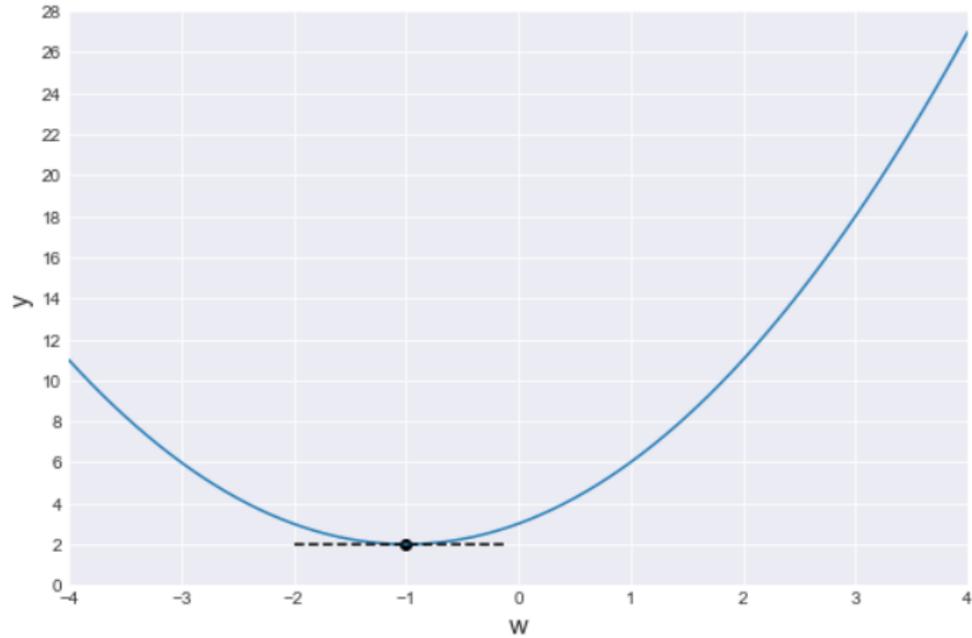
$$\underline{w = 0:}$$

$$y(0) = 3$$

$$\frac{dy}{dw}(0) = +2$$

direction      amount

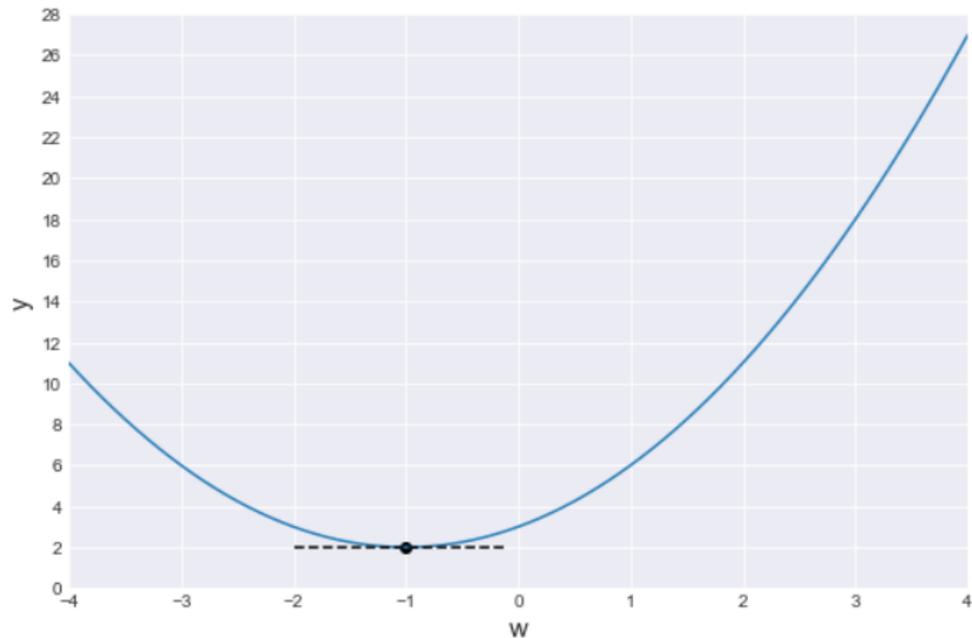
# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

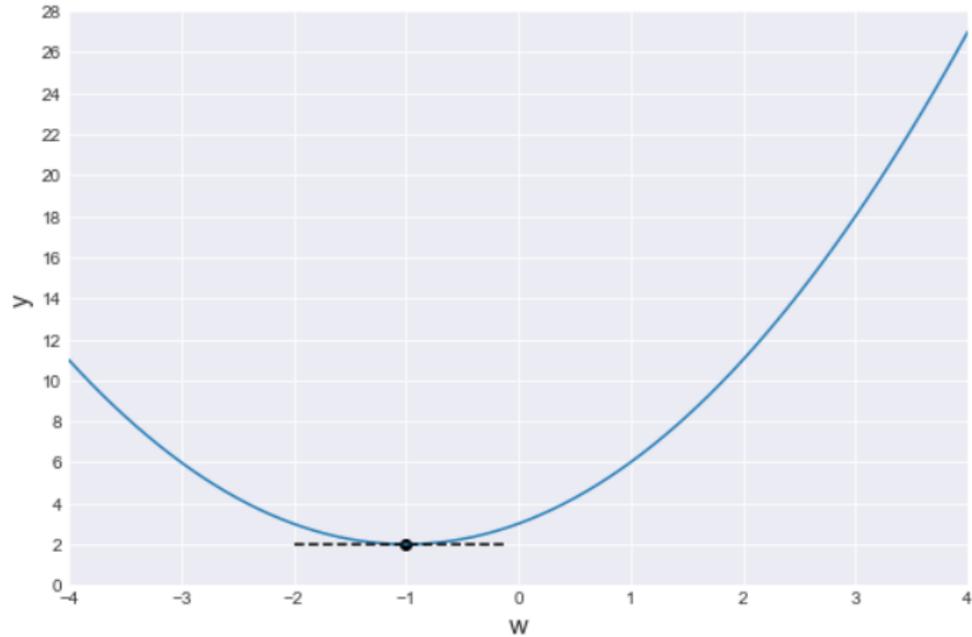
$$\frac{dy}{dw} = f'(w) = 2w + 2$$

Minima/  
Maxima



$$2w + 2 = 0$$

# Derivative



$$y = f(w) = w^2 + 2w + 3$$

$$\frac{dy}{dw} = f'(w) = 2w + 2$$

Minima/  
Maxima

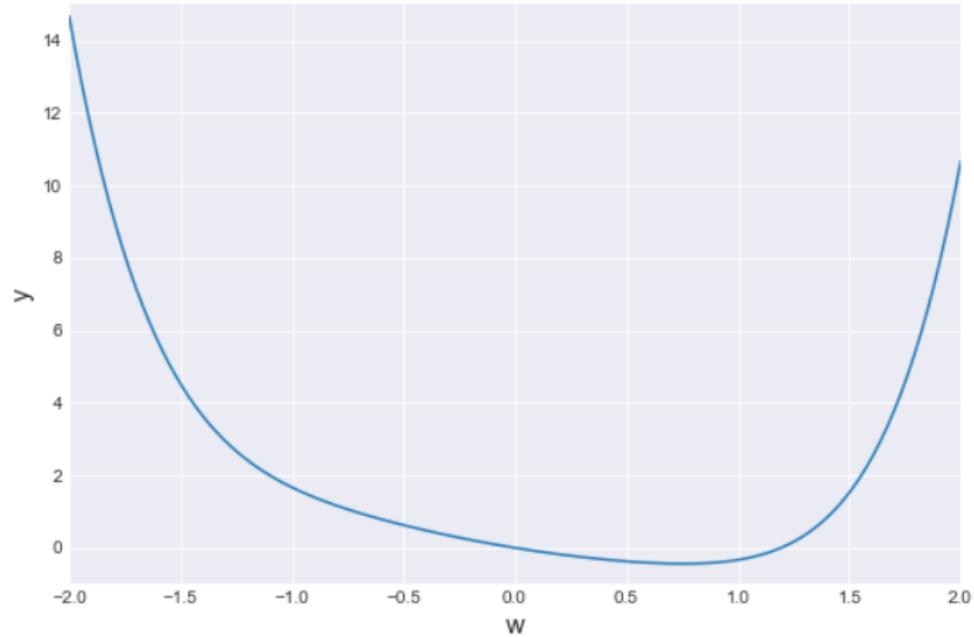


$$2w + 2 = 0$$

$$2w = -2$$

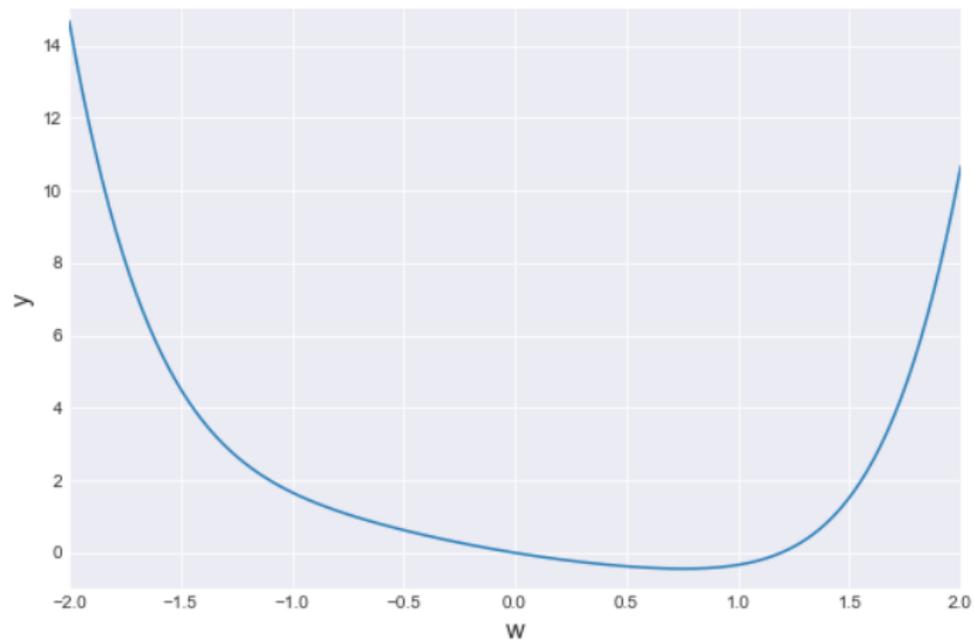
$$\underline{\underline{w = -1}}$$

# Derivative



$$y = f(w) = \frac{1}{6}w^6 + \frac{1}{2}w^2 - w$$

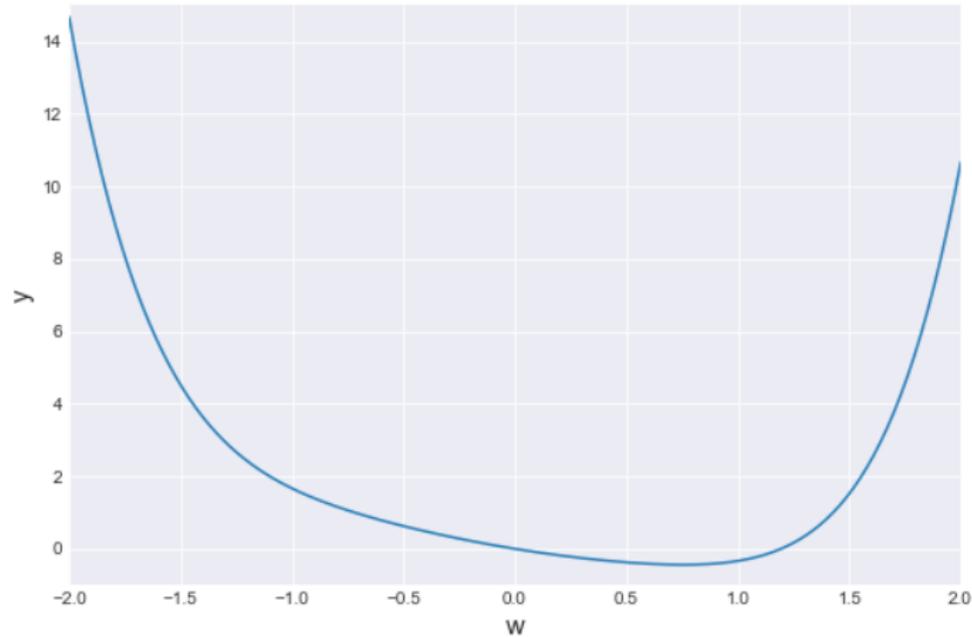
# Derivative



$$y = f(w) = \frac{1}{6}w^6 + \frac{1}{2}w^2 - w$$

$$\frac{dy}{dw} = f'(w) = w^5 + w - 1$$

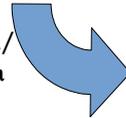
# Derivative



$$y = f(w) = \frac{1}{6} w^6 + \frac{1}{2} w^2 - w$$

$$\frac{dy}{dw} = f'(w) = w^5 + w - 1$$

Minima/  
Maxima



$$w^5 + w - 1 = 0$$

$$w^5 + w = 1$$

$$w(w^4 + 1) = 1$$

?