Neural Network and Deep Learning

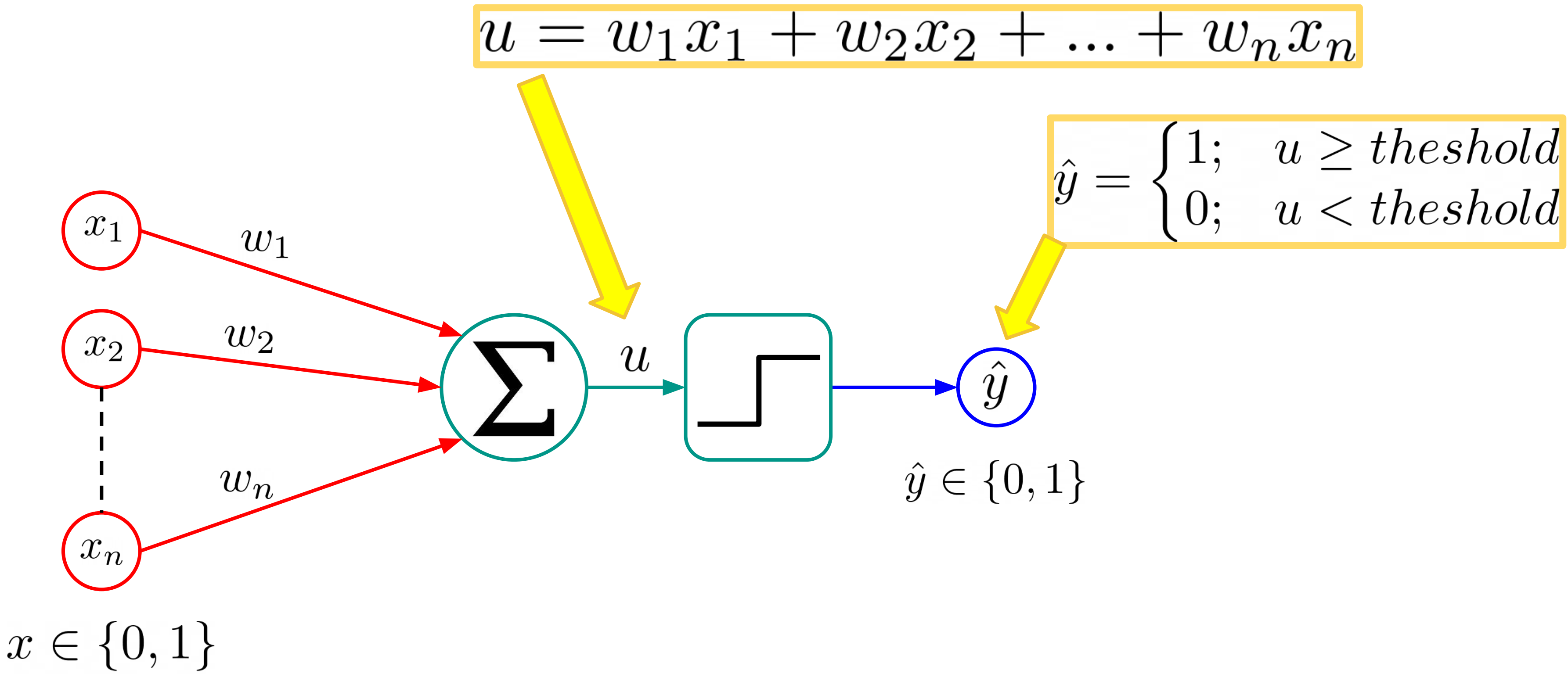
Delta Learning Rule

YANIKA KONGSOROT

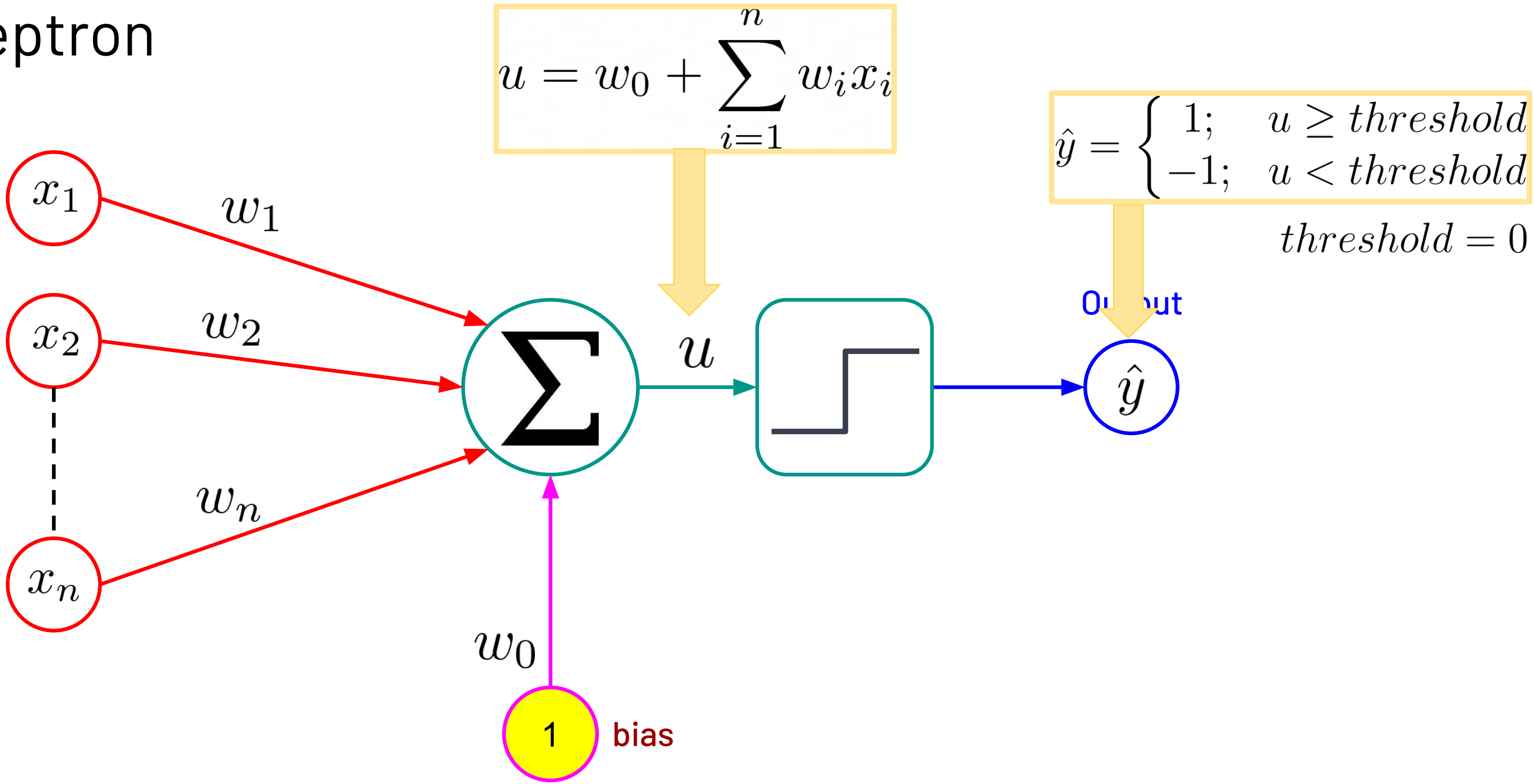
Recap

- McCulloch-Pitts Neuron
- Perceptron
- Adaline

McCulloch-Pitts Neuron

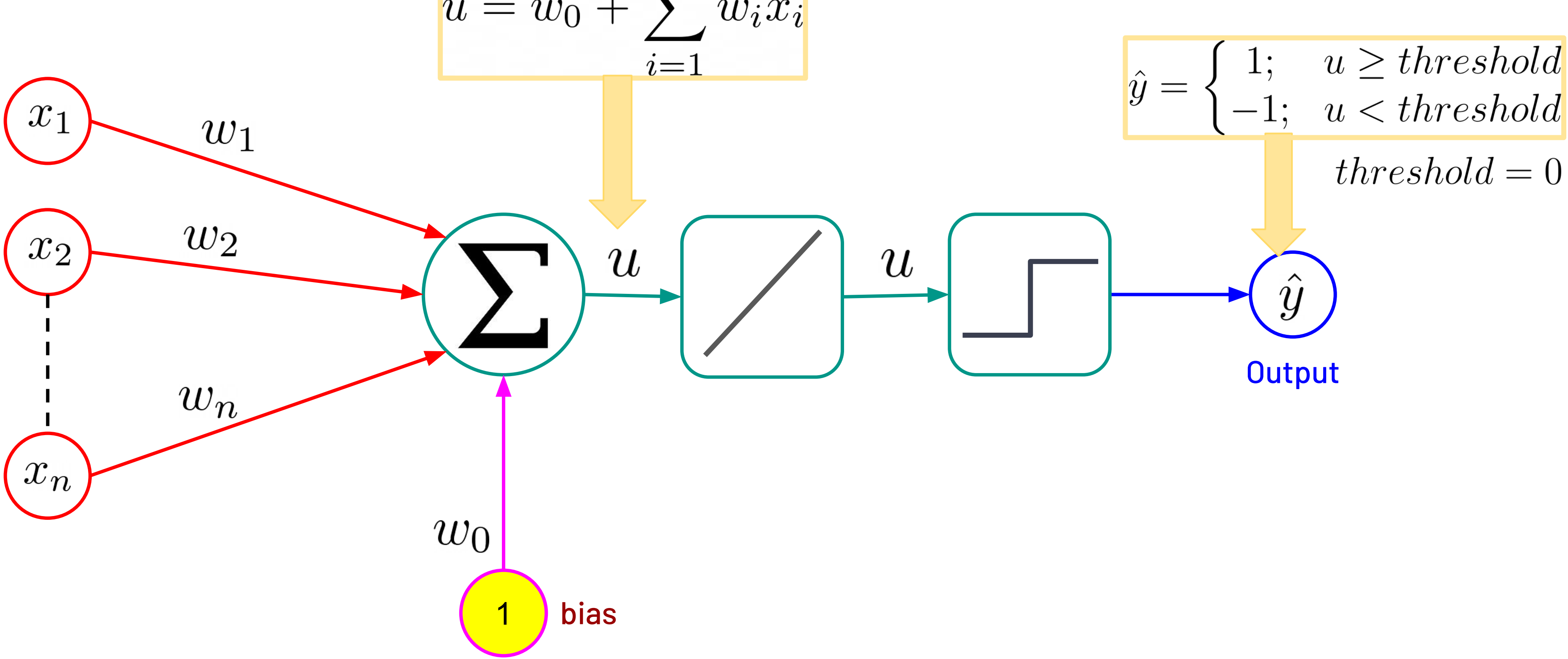


Perceptron



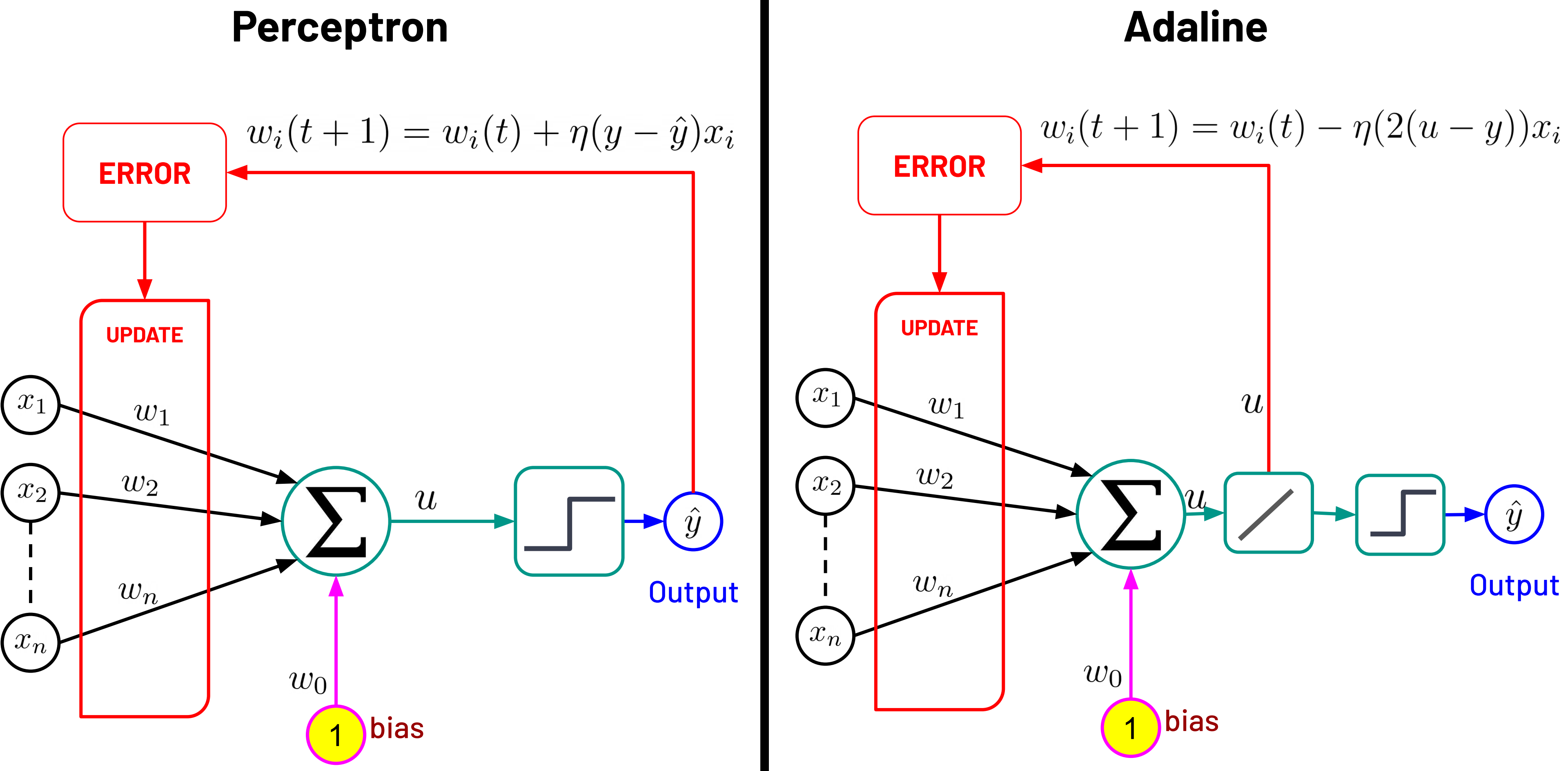
$u = w_0 + \sum_{i=1}^{n} w_i x_i$
$\hat{y} = \begin{cases} 1; & u \geq threshold \\ -1; & u < threshold \end{cases}$
$threshold = 0$
x_1
x_2
x_n
w_1
w_2
w_n
Σ
u
Output
$\hat{y}$
w_0
1
bias

Adaline



$u = w_0 + \sum_{i=1}^{n} w_i x_i$
$\hat{y} = \begin{cases} 1; & u \geq threshold \\ -1; & u < threshold \end{cases}$
$threshold = 0$
x_1
x_2
x_n
w_1
w_2
w_n
w_0
Σ
u
u
$\hat{y}$
Output
1
bias

Perceptron Learning
VS
Adaline Learning


Perceptron
$w_i(t+1) = w_i(t) + \eta(y - \hat{y})x_i$
ERROR
UPDATE
x_1
x_2
x_n
w_1
w_2
w_n
Σ
u
$\hat{y}$
Output
w_0
1
bias
Adaline
$w_i(t+1) = w_i(t) - \eta(2(u - y))x_i$
ERROR
UPDATE
u
x_1
x_2
x_n
w_1
w_2
w_n
Σ
u
$\hat{y}$
Output
w_0
1
bias

Behind the Adaline Learning

Delta learning rule
Gradient Descent

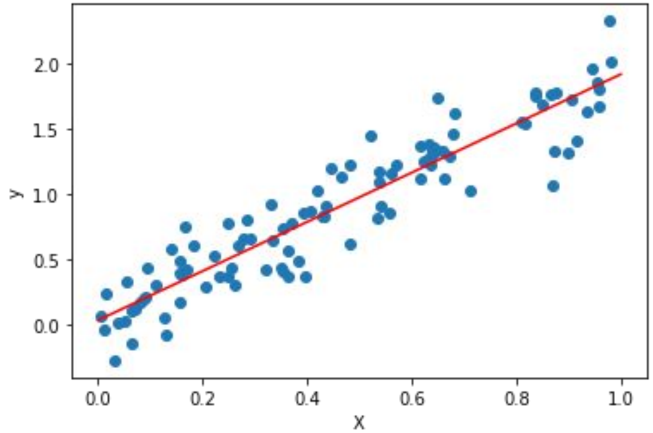
Outline of today

- Function Approximation
- Gradient-descent
- Delta Learning Rule
- Derivative of Error
- Gradient descent with Linear activation Function
- Gradient descent with non-Linear activation Function

Function Approximation

Input data

Output



Target

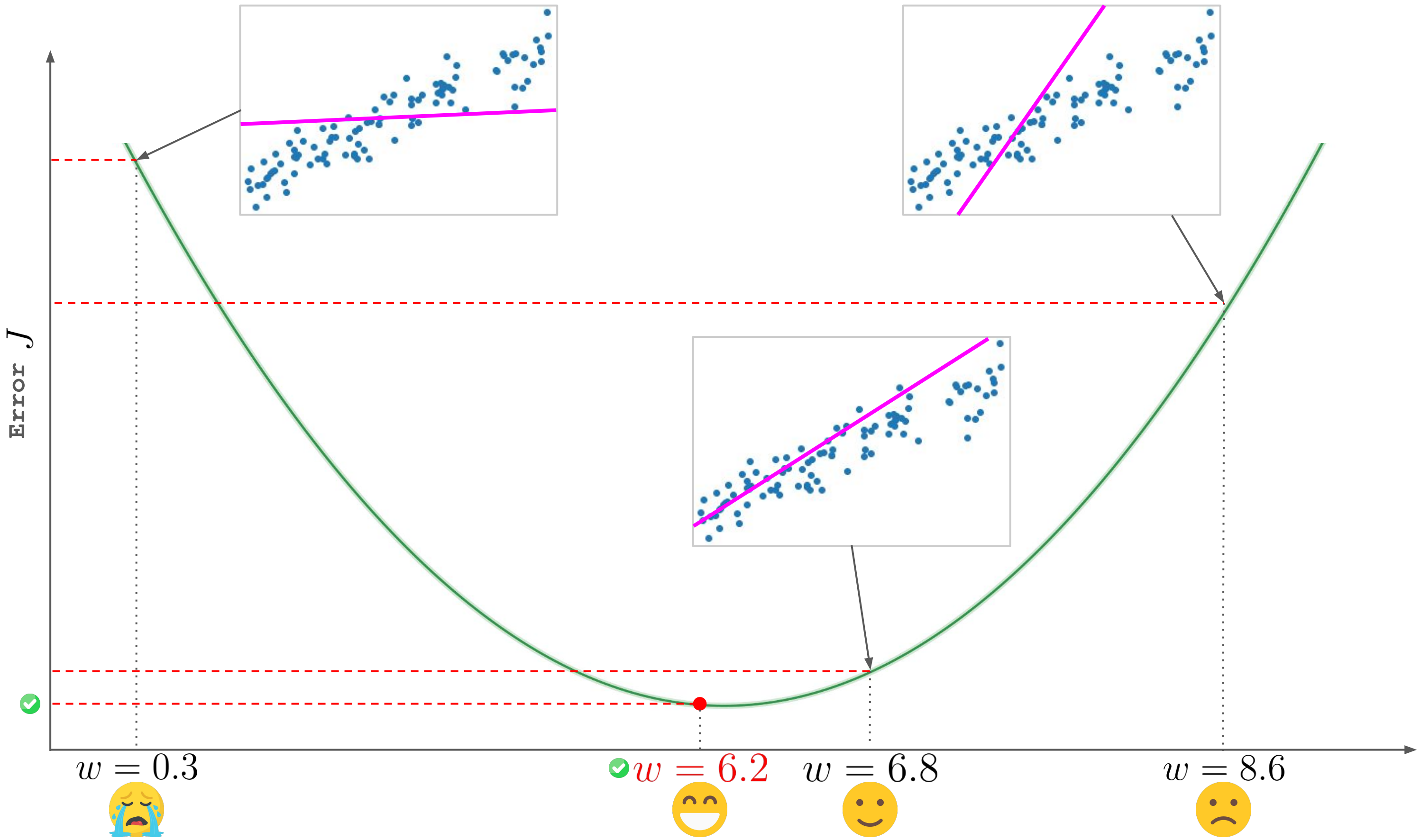
Actual value of each data point

Function Approximation

Function Approximation

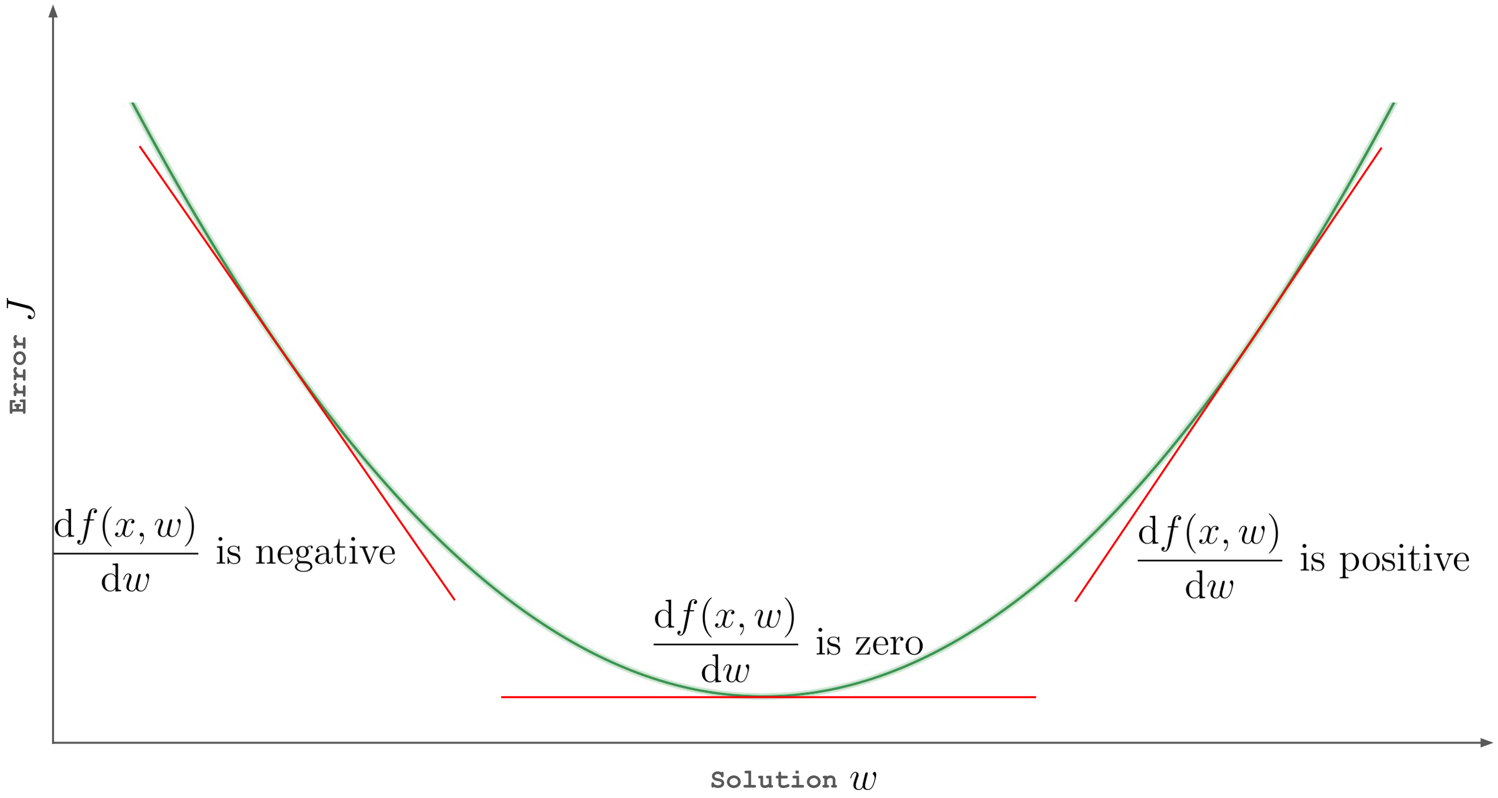
- The function to approximate the desire (target) value y can be modeled as
  - target function $y = \hat{y} = f(x, w)$
- This **parameters** of this function includes **data** x and **weight** w
- The objective of this function is **to compute the weight w that can make the minimum error of the function $f(w, x)$**, then the objective function can be represented as
  - $\boxed{J = \|y - f(x, w)\|}$
  - So, the optimum value of w must produce the minimum value for J

Gradient-descent

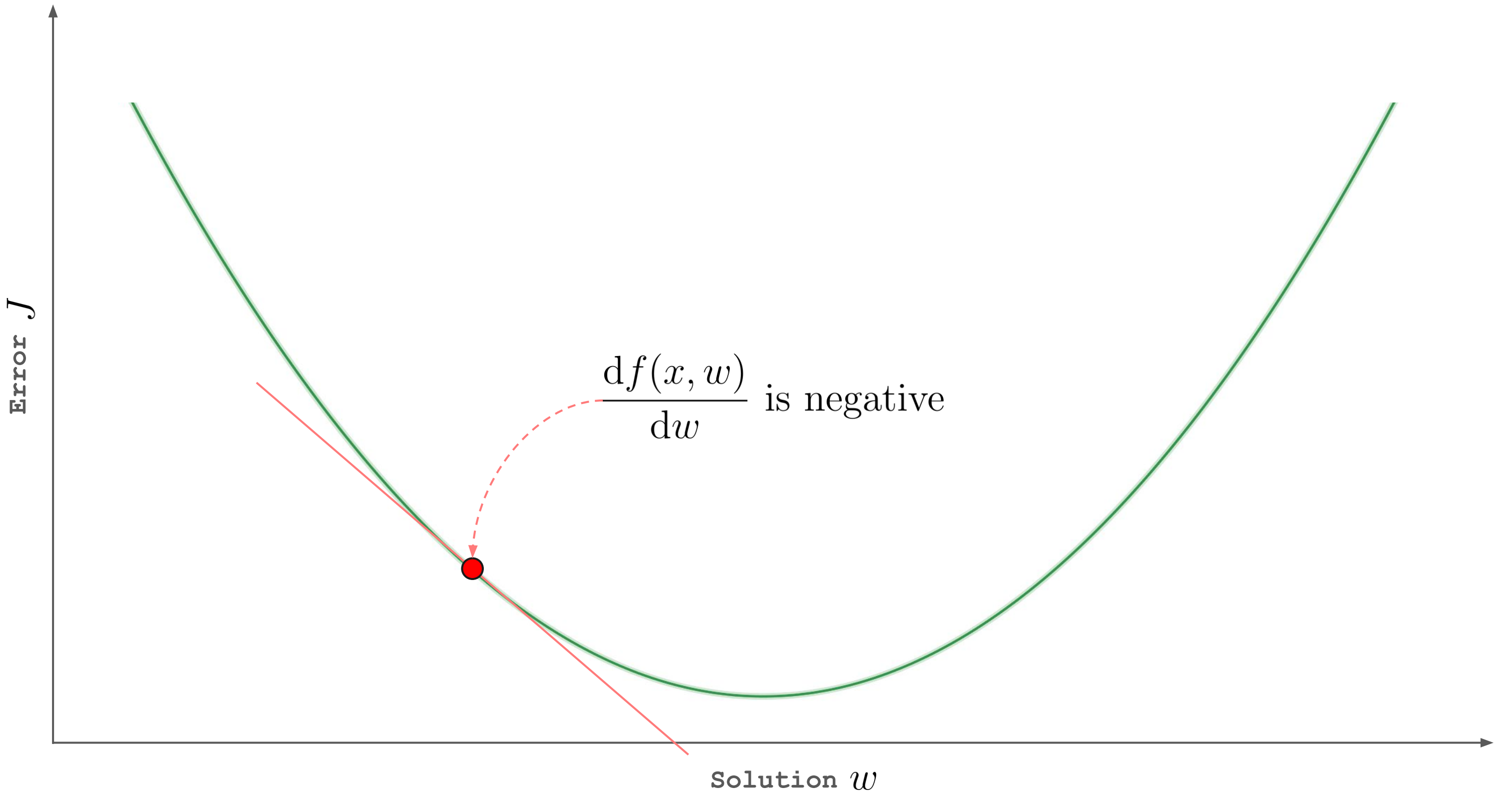


Gradient-descent

Gradient-descent



Gradient-descent



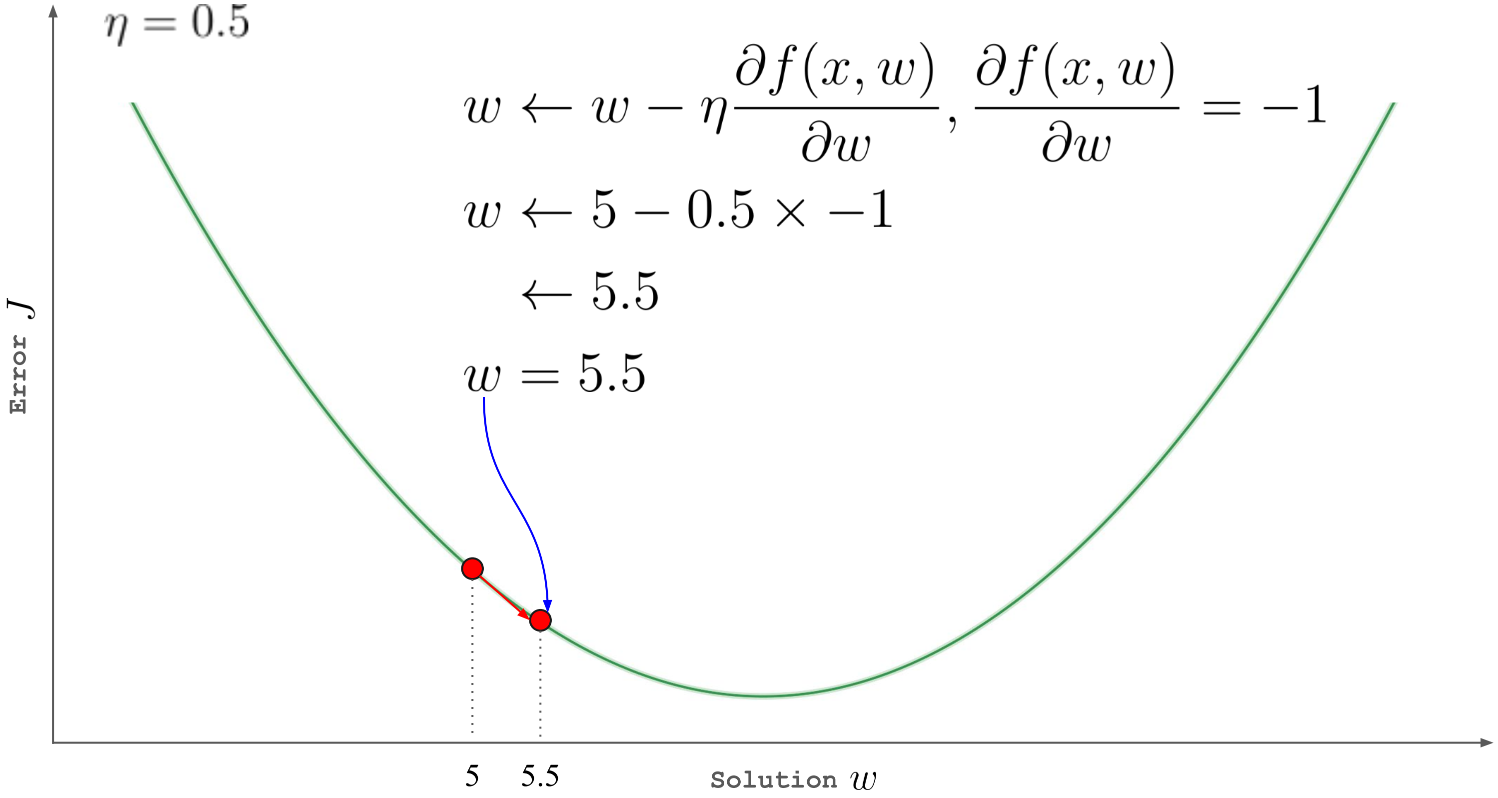
Gradient-descent

$$w \leftarrow w - \eta \frac{\partial f(x,w)}{\partial w}, \quad \frac{\partial f(x,w)}{\partial w} = -1$$

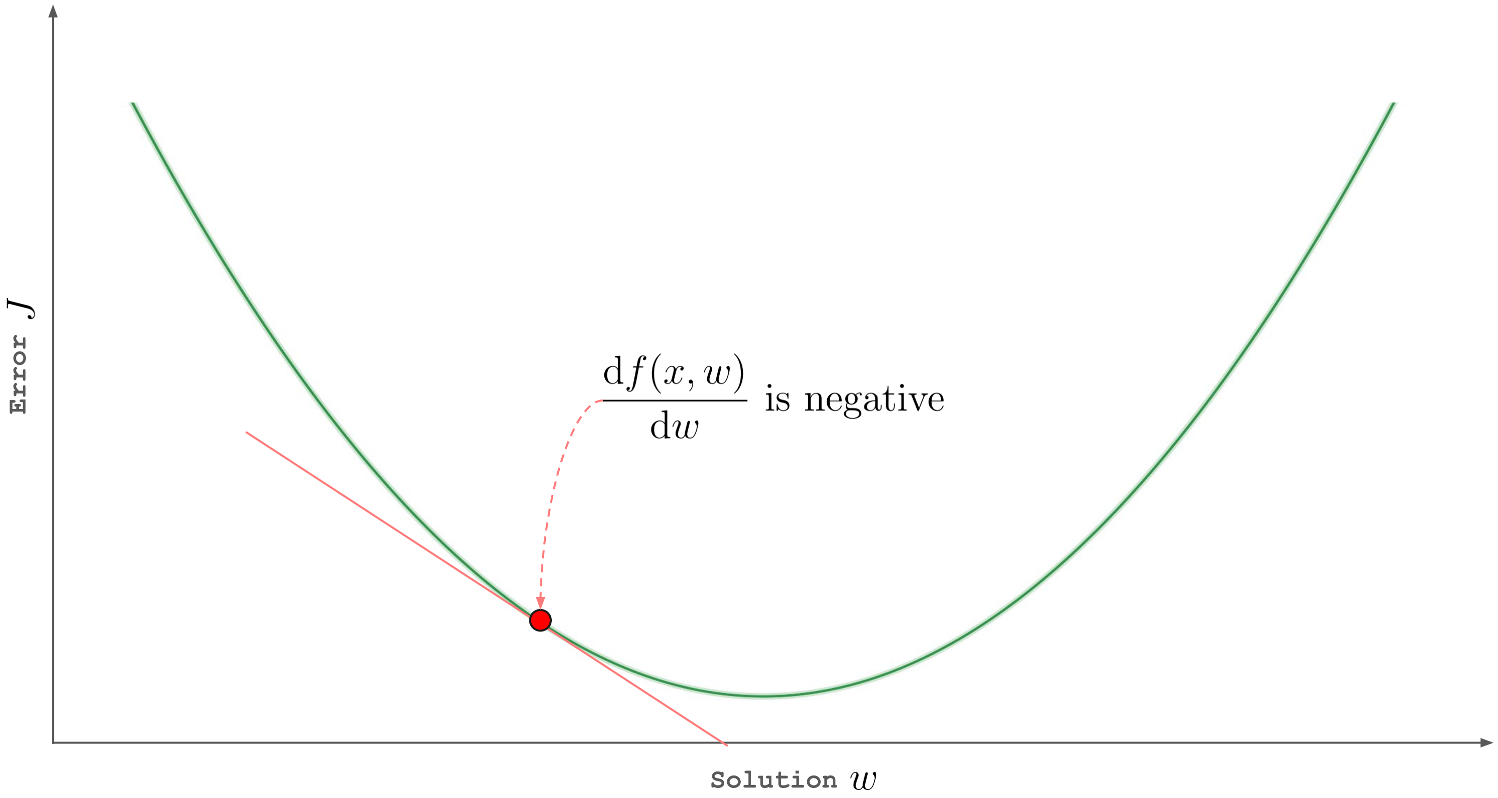
$$w \leftarrow 5 - 0.5 \times -1$$

$$\leftarrow 5.5$$

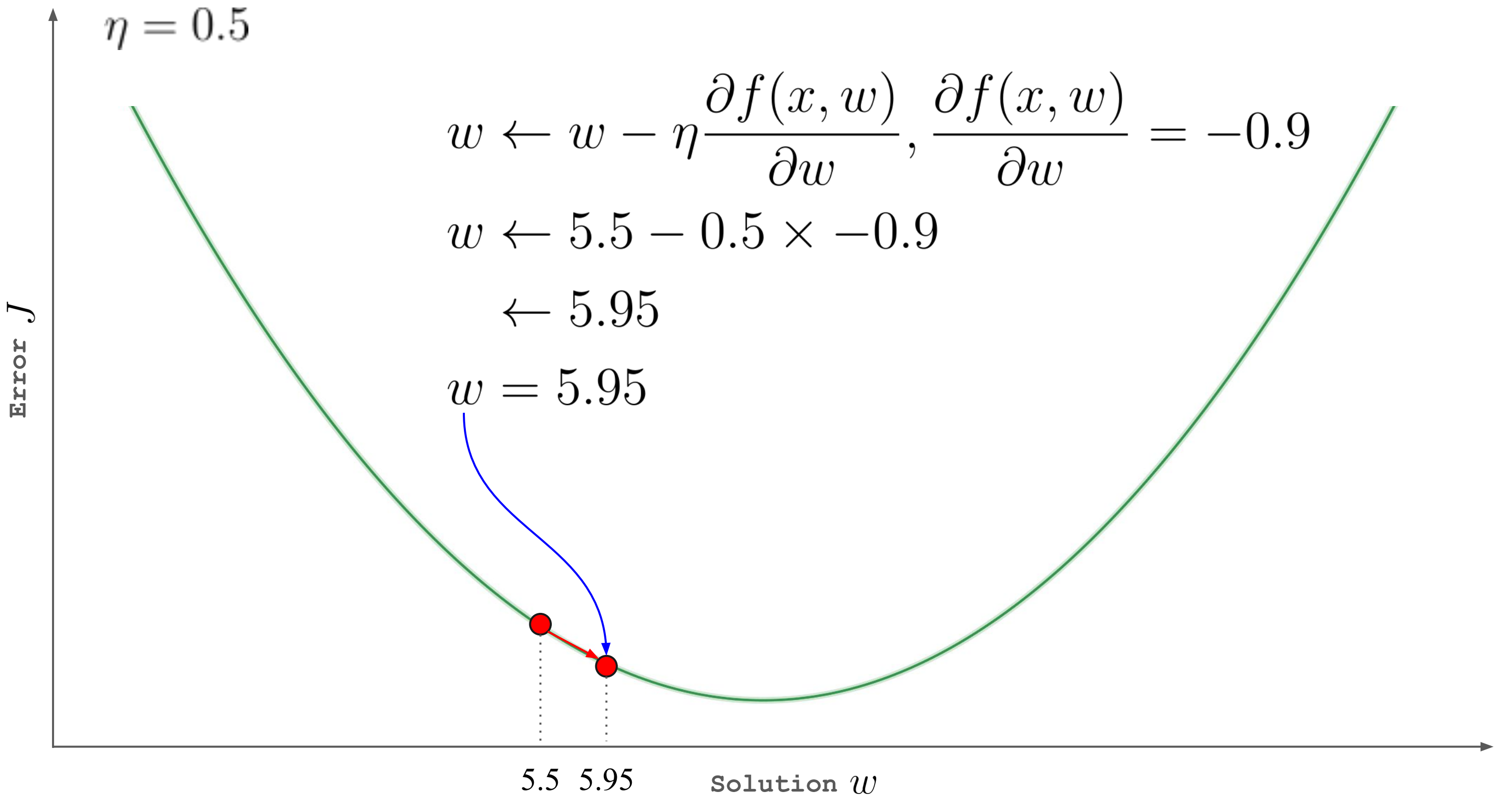
$$w = 5.5$$



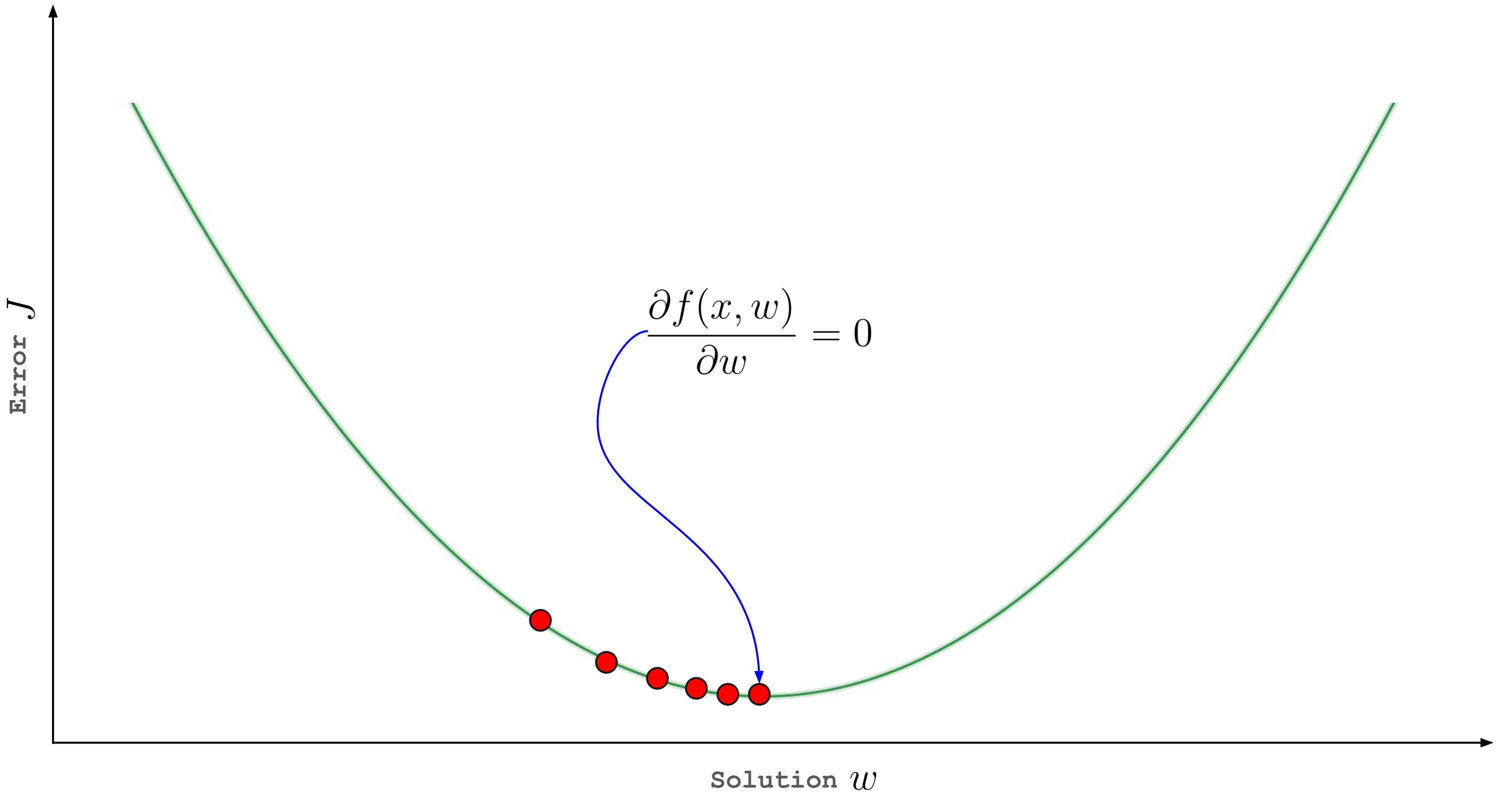
Gradient-descent



Gradient-descent



Gradient-descent



Delta Learning Rule

The learning rule in Adaline

Delta Learning Rule

- *Delta Learning Rule* was developed by Widrow and Hoff in the early 1960s.
- Also known as the the *Widrow & Hoff Learning Rule* or the *Least Mean Square (LMS) algorithm*.
- The delta rule is used in Adaline networks for training.
- The delta rule uses the difference between target output and obtained activation to drive learning, which aims to minimize that difference (error) by using the gradient descent to minimize the error from an adaline network's weights.
- The delta rule employs the *error function* for **Gradient Descent learning**, which involves *the modification of weights along the most direct path in weight-space to minimize error*.

Delta Learning Rule

- Given

$$E_j = y_j - \hat{y}_j$$ is the **error** for the j-th sample

$$\hat{y}_j = \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)$$ is the **output** for the j-th sample

$$J_j = E_j^2$$ is the **squared error** for the j-th sample

- The learning of delta rule use the ***Gradient Descent Learning*** of E_j

$$w_i \leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i}$$

Derivative of Error

Derivative of Error

$$\frac{\partial J_j}{\partial w_i}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative formulars: https://www.math.wustl.edu/~freiwald/131derivativetable.pdf

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\frac{\partial J_j}{\partial w_i}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'} \quad \frac{\partial J_j}{\partial w_i} = \frac{1}{2}\frac{\partial E_j^2}{\partial w_i} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\frac{\partial E_j^2}{\partial w_i} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}$$

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'} \quad \frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \quad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$= \frac{1}{2}\frac{\partial E_j^2}{\partial E_j}\frac{\partial E_j}{\partial w_i}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'} \quad \frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \quad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}} \frac{\partial E_j}{\partial w_i} = \frac{2}{2} E_j \frac{\partial E_j}{\partial w_i}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}fg = fg' + gf'} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x}c = 0} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x}x^n = nx^{n-1}}$$

$$\begin{aligned}
\frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\
&= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i} \\
&= E_j\frac{\partial E_j}{\partial w_i}
\end{aligned}$$

$$\boxed{\begin{aligned}
J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\
&= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2
\end{aligned}}$$

Derivative of Error

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}} \frac{\partial E_j}{\partial w_i} = \frac{2}{2} E_j \frac{\partial E_j}{\partial w_i}$$

$$= E_j \frac{\partial E_j}{\partial w_i} = E_j \frac{\partial (y_j - \hat{y}_j)}{\partial w_i}$$

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

Derivative of Error

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\begin{aligned} \frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\ &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i} \\ &= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial(y_j - \hat{y}_j)}{\partial w_i} \end{aligned}$$

Derivative of Error

$$\boxed{J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 = \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\begin{aligned}
\frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\
&= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i} \\
&= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial (y_j - \hat{y}_j)}{\partial w_i} \\
&= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right)
\end{aligned}$$

Derivative of Error

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 = \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial(y_j - \hat{y}_j)}{\partial w_i}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i}$$

Derivative of Error

$$\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\begin{aligned} \frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\ &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i} \\ &= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial (y_j - \hat{y}_j)}{\partial w_i} \\ &= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i} \\ &= -E_j\frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j\frac{\partial \phi\left(u\right)}{\partial w_i}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i} \end{aligned}$$

Derivative of Error

$$\boxed{J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 = \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'} \qquad \frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j \frac{\partial E_j}{\partial w_i} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$= E_j \frac{\partial E_j}{\partial w_i} = E_j \frac{\partial (y_j - \hat{y}_j)}{\partial w_i} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$= E_j \left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j \left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j \frac{\partial \hat{y}_j}{\partial w_i} \qquad \boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$= -E_j \frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j \boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i}$$

Derivative of Error

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\begin{aligned}
\frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\
&= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j \frac{\partial E_j}{\partial w_i} \\
&= E_j \frac{\partial E_j}{\partial w_i} = E_j \frac{\partial (y_j - \hat{y}_j)}{\partial w_i} \\
&= E_j \left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j \frac{\partial \hat{y}_j}{\partial w_i} \\
&= -E_j \frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j \boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i} \\
&= -E_j \frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i}
\end{aligned}$$

Derivative of Error

$$J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2$$
$$= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i}$$

$$= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial (y_j - \hat{y}_j)}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$= -E_j\frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j\boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i}$$

$$= -E_j\frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i} = -E_j\frac{\partial \phi(u)}{\partial u}\frac{\partial\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i}$$

Derivative of Error

$$J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 = \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i}$$

$$= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial (y_j - \hat{y}_j)}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$= -E_j\frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j\boxed{\frac{\partial \phi(u)}{\partial w_i}};\, u = w_0 + \sum_{i=1}^{n} w_i x_{j,i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= -E_j\frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i} = -E_j\frac{\partial \phi(u)}{\partial u}\boxed{\frac{\partial\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i}}$$

Derivative of Error

$$\boxed{\begin{aligned} J_j &= \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2 \\ &= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2 \end{aligned}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} c = 0}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)}$$

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'}$$

$$\begin{aligned}
\frac{\partial J_j}{\partial w_i} &= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i} \\
&= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j \frac{\partial E_j}{\partial w_i} \\
&= E_j \frac{\partial E_j}{\partial w_i} = E_j \frac{\partial (y_j - \hat{y}_j)}{\partial w_i} \\
&= E_j \left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j \frac{\partial \hat{y}_j}{\partial w_i} \\
&= -E_j \frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j \boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i} \\
&= -E_j \frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i} = -E_j \frac{\partial \phi(u)}{\partial u}\boxed{\frac{\partial\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i}} \\
&= -E_j \frac{\partial \phi(u)}{\partial u}\left(\frac{\partial w_0}{\partial w_i} + \ldots + \frac{\partial w_i x_{j,i}}{\partial w_i} + \ldots + \frac{\partial w_n x_{j,n}}{\partial w_i}\right)
\end{aligned}$$

Derivative of Error

$$J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2$$
$$= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$= E_j\frac{\partial E_j}{\partial w_i} = E_j\frac{\partial(y_j - \hat{y}_j)}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$= -E_j\frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j\boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= -E_j\frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i} = -E_j\frac{\partial \phi(u)}{\partial u}\boxed{\frac{\partial\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i}}$$

$$= -E_j\frac{\partial \phi(u)}{\partial u}\left(\frac{\partial w_0}{\partial w_i} + \ldots + \boxed{\frac{\partial w_i x_{j,i}}{\partial w_i}} + \ldots + \frac{\partial w_n x_{j,n}}{\partial w_i}\right)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

Derivative of Error

$$J_j = \frac{1}{2}\left[y_i - \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)\right]^2$$
$$= \frac{1}{2}\left[y_j - \hat{y}_j\right]^2 = \frac{1}{2}E_j^2$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} c = 0$$

$$\frac{\partial J_j}{\partial w_i} = \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial w_i}} + E_j^2 \frac{\partial \frac{1}{2}}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}$$

$$= \frac{1}{2}\boxed{\frac{\partial E_j^2}{\partial E_j}}\frac{\partial E_j}{\partial w_i} = \frac{2}{2}E_j\frac{\partial E_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= E_j \frac{\partial E_j}{\partial w_i} = E_j \frac{\partial (y_j - \hat{y}_j)}{\partial w_i}$$

$$= E_j\left(\frac{\partial y_j}{\partial w_i} - \frac{\partial \hat{y}_j}{\partial w_i}\right) = E_j\left(-\frac{\partial \hat{y}_j}{\partial w_i}\right) = -E_j\frac{\partial \hat{y}_j}{\partial w_i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x))g'(x)$$

$$= -E_j \frac{\partial \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i} = -E_j \boxed{\frac{\partial \phi(u)}{\partial w_i}}; u = w_0 + \sum_{i=1}^{n} w_i x_{j,i}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f \pm g = f' + g'$$

$$= -E_j \frac{\partial \phi(u)}{\partial u}\frac{\partial u}{\partial w_i} = -E_j \frac{\partial \phi(u)}{\partial u}\boxed{\frac{\partial\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)}{\partial w_i}}$$

$$= -E_j \frac{\partial \phi(u)}{\partial u}\left(\frac{\partial w_0}{\partial w_i} + \ldots + \boxed{\frac{\partial w_i x_{j,i}}{\partial w_i}} + \ldots + \frac{\partial w_n x_{j,n}}{\partial w_i}\right)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} fg = fg' + gf'$$

$$= -E_j \frac{\partial \phi(u)}{\partial u} x_{j,i} = -(y_j - \hat{y}_j)\frac{\partial \phi(u)}{\partial u} x_{j,i}$$

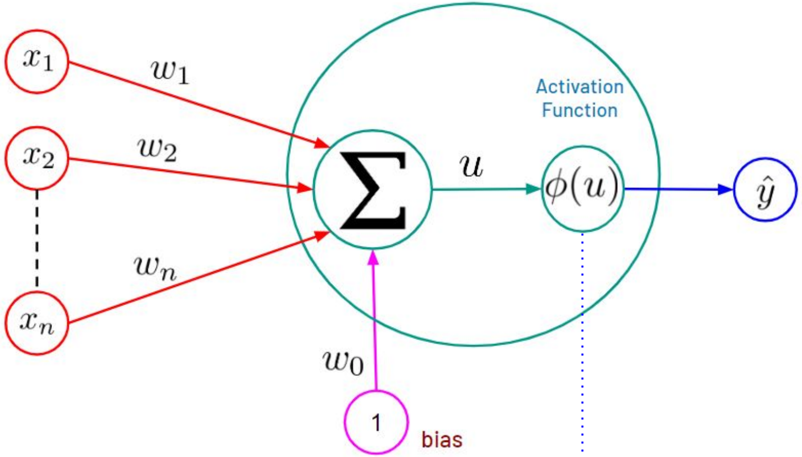
Derivative of Error

$$w_i \leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i} \quad \frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

- According to the convergence theory, **during the training phase, the value of error will decrease**.
- Then when using that error to adjust the weights of model based on the *Gradient descent learning*, **the weight change** can be formulated as
  - $$\boxed{w_i \leftarrow w_i + \eta \left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}}$$
  - when η is the learning rate which it value is $\eta = (0, 1]$
  - $\dfrac{\partial \phi(u)}{\partial u}$ is **the derivative of activation function** used in the neural model

Gradient descent with Linear activation Function

Gradient descent with Linear activation



The output of model when using linear activation function

$$\hat{y}_j = \phi\left(w_0 + \sum_{i=1}^{n} w_i x_{j,i}\right)$$
$$= \phi\left(w_0 + w_1 x_{j,1} + \cdots + w_i x_{j,n}\right)$$

ϕ is linear activation function

Gradient descent with Linear activation

Compute the derivative $$\phi(u) = u$$

Gradient descent with Linear activation

Compute the derivative

$$\phi(u) = u$$

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial u}{\partial u} = 1$$

Gradient descent with Linear activation

Compute the derivative

$$\phi(u) = u$$

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial u}{\partial u} = 1$$

$$\frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right)\frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

Gradient descent with Linear activation

Compute the derivative

$$\phi(u) = u$$

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial u}{\partial u} = 1$$

$$\frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

$$w_i \leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i}$$

Gradient descent with Linear activation

Compute the derivative

$$\phi(u) = u$$

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial u}{\partial u} = 1$$

$$\frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

$$w_i \leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i}$$

$$\leftarrow w_i + \eta \left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

Gradient descent with Linear activation

Compute the derivative

$$\phi(u) = u$$

$$\frac{\partial \phi(u)}{\partial u} = \frac{\partial u}{\partial u} = 1$$

$$\frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

Weight adjusting

$$\boxed{\begin{aligned} w_i &\leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i} \\ &\leftarrow w_i + \eta \left(y_j - \hat{y}_j\right) \frac{\partial \phi\left(u\right)}{\partial u} x_{j,i} \\ &\leftarrow w_i + \eta \left(y_j - \hat{y}_j\right) x_{j,i} \end{aligned}}$$

Gradient descent with non-Linear activation Function

Gradient descent with non-linear activation

Sigmoid function

$$
\begin{aligned}
\hat{y}_j &= \phi(u_j) \\
&= \frac{1}{1 + e^{-u_j}} \\
&= \frac{1}{1 + e^{-(w_0 + \sum_{i-1}^{n} w_i x_{j,i})}} \\
&= \frac{1}{1 + e^{-(w_0 + w_1 x_{j,1} + \ldots + w_n x_{j,n})}}
\end{aligned}
$$

Gradient descent with non-linear activation

$$\frac{\partial \phi(u_j)}{\partial u_j} = \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}}$$

$$\boxed{\begin{aligned}
\frac{\partial}{\partial x}\frac{f(x)}{g(x)} &= \frac{g(x)\frac{\partial}{\partial x}f(x) - f(x)\frac{\partial}{\partial x}g(x)}{g^2(x)} \\
\frac{\partial}{\partial u_j}\frac{1}{1+e^{-u_j}} &= \frac{(1+e^{-u_j})\frac{\partial}{\partial u_j}1 - 1\frac{\partial}{\partial u_j}(1+e^{-u_j})}{(1+e^{-u_j})^2} \\
&= \frac{-1\frac{\partial}{\partial u_j}(1+e^{-u_j})}{(1+e^{-u_j})^2} \\
&= -\frac{1}{(1+e^{-u_j})^2}\frac{\partial}{\partial u_j}(1+e^{-u_j})
\end{aligned}}$$

Gradient descent with non-linear activation

$$\begin{aligned}\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j}\frac{1}{1+e^{-u_j}} \\ &= -\frac{1}{(1+e^{-u_j})^2}\frac{\partial}{\partial u_j}(1+e^{-u_j})\end{aligned}$$

$$\boxed{\begin{aligned}\frac{\partial}{\partial x}(f(x)+g(x)) &= \frac{\partial}{\partial x}f(x) + \frac{\partial}{\partial x}g(x) \\ \frac{\partial}{\partial u_j}(1+e^{-u_j}) &= \frac{\partial}{\partial u_j}1 + \frac{\partial}{\partial u_j}e^{-u_j} \\ &= \frac{\partial}{\partial u_j}e^{-u_j} \\ \\ \frac{\partial}{\partial x}e^{x} &= e^{x}\frac{\partial}{\partial x}x \\ \frac{\partial}{\partial u_j}e^{-u_j} &= e^{-u_j}\frac{\partial}{\partial u_j}(-u_j) \\ &= e^{-u_j}(-1) \\ &= -e^{-u_j}\end{aligned}}$$

Gradient descent with non-linear activation

$$\begin{aligned}\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}} \\ &= -\frac{1}{(1+e^{-u_j})^2} \frac{\partial}{\partial u_j} (1+e^{-u_j}) \\ &= \frac{e^{-u_j}}{(1+e^{-u_j})^2}\end{aligned}$$

Gradient descent with non-linear activation

$$
\begin{aligned}
\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}} \\
&= -\frac{1}{(1+e^{-u_j})^2} \frac{\partial}{\partial u_j} (1+e^{-u_j}) \\
&= \frac{e^{-u_j}}{(1+e^{-u_j})^2} \\
&= \frac{1}{1+e^{-u_j}} \frac{e^{-u_j}}{1+e^{-u_j}}
\end{aligned}
$$

Gradient descent with non-linear activation

$$
\begin{aligned}
\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}} \\
&= -\frac{1}{(1+e^{-u_j})^2} \frac{\partial}{\partial u_j} (1+e^{-u_j}) \\
&= \frac{e^{-u_j}}{(1+e^{-u_j})^2} \\
&= \frac{1}{1+e^{-u_j}} \frac{e^{-u_j}}{1+e^{-u_j}} \\
&= \frac{1}{1+e^{-u_j}} \frac{1+e^{-u_j}-1}{1+e^{-u_j}} = \frac{1}{1+e^{-u_j}} \frac{(1+e^{-u_j})-1}{1+e^{-u_j}}
\end{aligned}
$$

Gradient descent with non-linear activation

$$
\begin{aligned}
\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}} \\
&= -\frac{1}{(1+e^{-u_j})^2} \frac{\partial}{\partial u_j} (1+e^{-u_j}) \\
&= \frac{e^{-u_j}}{(1+e^{-u_j})^2} \\
&= \frac{1}{1+e^{-u_j}} \frac{e^{-u_j}}{1+e^{-u_j}} \\
&= \frac{1}{1+e^{-u_j}} \frac{1+e^{-u_j}-1}{1+e^{-u_j}} = \frac{1}{1+e^{-u_j}} \frac{(1+e^{-u_j})-1}{1+e^{-u_j}} \\
&= \frac{1}{1+e^{-u_j}} \left(\frac{1+e^{-u_j}}{1+e^{-u_j}} - \frac{1}{1+e^{-u_j}} \right)
\end{aligned}
$$

Gradient descent with non-linear activation

$$
\begin{aligned}
\frac{\partial \phi(u_j)}{\partial u_j} &= \frac{\partial}{\partial u_j} \frac{1}{1+e^{-u_j}} \\
&= -\frac{1}{(1+e^{-u_j})^2} \frac{\partial}{\partial u_j} (1+e^{-u_j}) \\
&= \frac{e^{-u_j}}{(1+e^{-u_j})^2} \\
&= \frac{1}{1+e^{-u_j}} \frac{e^{-u_j}}{1+e^{-u_j}} \\
&= \frac{1}{1+e^{-u_j}} \frac{1+e^{-u_j}-1}{1+e^{-u_j}} = \frac{1}{1+e^{-u_j}} \frac{(1+e^{-u_j})-1}{1+e^{-u_j}} \\
&= \frac{1}{1+e^{-u_j}} \left(\frac{1+e^{-u_j}}{1+e^{-u_j}} - \frac{1}{1+e^{-u_j}} \right) \\
&= \phi(u_j)\,(1-\phi(u_j))
\end{aligned}
$$

Gradient descent with with non-linear activation

Compute the derivative

$$\phi(u) = \frac{1}{1 + e^{-u_j}}$$

$$\frac{\partial \phi(u_j)}{\partial u_j} = \phi(u_j)\left(1 - \phi(u_j)\right)$$

$$\frac{\partial J_j}{\partial w_i} = -\left(y_j - \hat{y}_j\right)\frac{\partial \phi\left(u\right)}{\partial u} x_{j,i}$$

Weight adjusting

$$\boxed{\begin{aligned} w_i &\leftarrow w_i - \eta \frac{\partial J_j}{\partial w_i} \\ &\leftarrow w_i + \eta\left(y_j - \hat{y}_j\right)\frac{\partial \phi\left(u\right)}{\partial u} x_{j,i} \\ &\leftarrow w_i + \eta\left(y_j - \hat{y}_j\right)\phi(u_j)\left(1 - \phi(u_j)\right) x_{j,i} \end{aligned}}$$

Summary

...