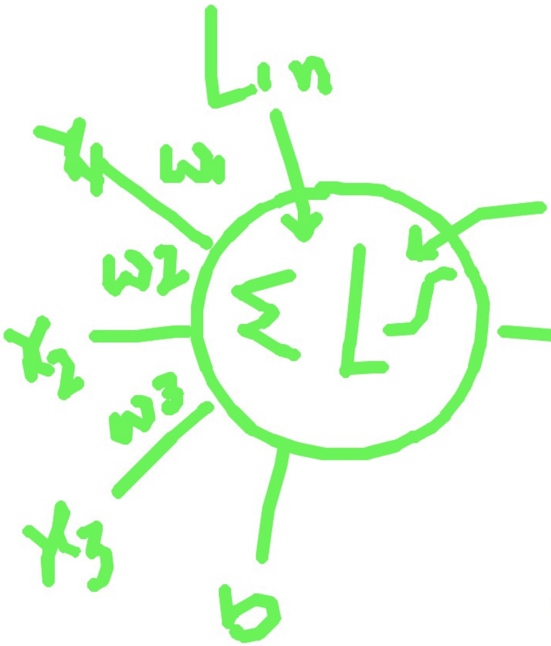
Artificial Neuron



NL (sigmoid .. can be others)

$$\hat{y} = NL(\Sigma w_i x_i + b)$$
$$= Sig(\Sigma w_i x_i + b)$$

$$\boxed{\hat{y} = g(\Sigma w_i x_i + b)}$$

For a given inp obs_o1 = $\{\underline{x}_1, \underline{x}_2 \ldots \underline{x}_n\}$

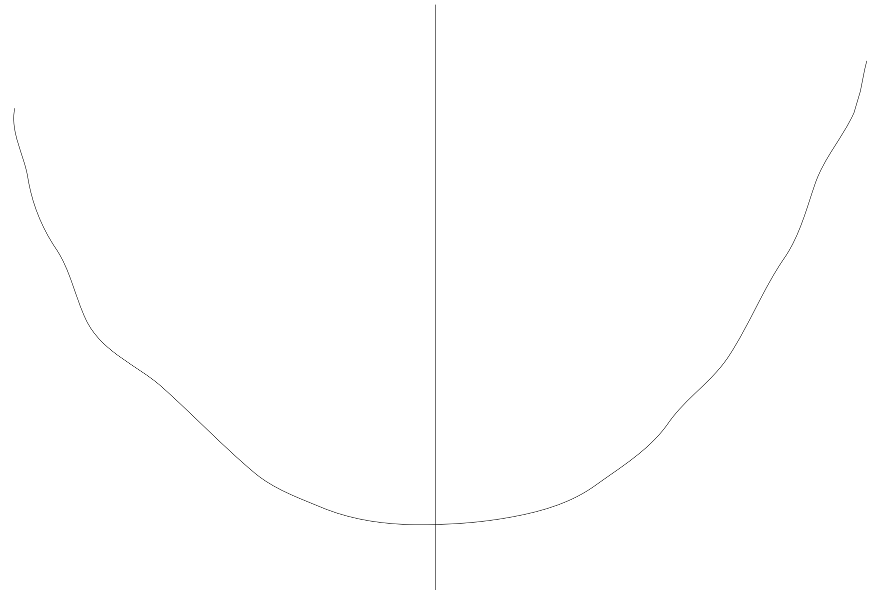
and y :

Dev from Truth $(y - \hat{y})$

$$\boxed{E = \Sigma (y - \hat{y})^2}$$ SSE

Penalizing the larger errors and SSE curve

- Suppose we have two value of y-ypred
- 2, 10
- Intial diff of dev = 10-2 = 8
- 2**2 = 4, 10**2=100 , diff now is 100-4 = 96
- Basically means, by squaring the (y-ypred) , we are penalizing the larger errors and making them even larger errors
- When measuring error we dont want the model to make larger errors and if so we want to fix it quickly
- We fix the larger errors first
- SSE = (Deviation) ** 2 , looks like x**2



Backpropagation of error

$\hat{y} = g(\sum w_i x_i + b)$

$$\boxed{\hat{y} = g(\sum w_i x_i)}$$

$w_0 = b$ is
$x_0 = 1$

$$E = \sum (y - \hat{y})^2 = \sum \left(y - g(\sum w_i x_i)\right)^2$$

$E \to 0$, $v \to 0 \Rightarrow -\frac{\partial v}{\partial t}$ brakes

$E \to 0$, $\boxed{\Delta w_i = -\eta \frac{\partial E}{\partial w_i}}$ Multiple Times

Backpropagation of the error

$$\boxed{\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i}\left(\sum \left(y - g(\sum w_i x_i)\right)^2\right)}$$

Derivative of Sigmoid

$$g(x) = \frac{1}{1+e^{-x}} \qquad g'(x) = \frac{d}{dx} g(x)$$

$$= \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \frac{d}{dx}\left(\underbrace{1+e^{-x}}_{u}\right)^{-1}$$

$$\boxed{\frac{d}{dx} u^n = n u^{n-1} \cdot \frac{du}{dx}}$$

$$= -1 \cdot \left(1+e^{-x}\right)^{-1-1} \cdot \frac{d}{dx}\left(1+e^{-x}\right)$$

$$= \frac{-1}{(1+e^{-x})^2} \cdot \left(0 + e^{-x} \cdot -1\right) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \left(\frac{1}{1+e^{-x}}\right) \cdot \left(\frac{e^{-x}}{1+e^{-x}}\right) = g(x) \cdot \frac{e^{-x}+1-1}{1+e^{-x}}$$

$$= g(x) \cdot \left(1 - \frac{1}{1+e^{-x}}\right) = g(x) \cdot (1 - g(x))$$

$$\boxed{g'(x) = g(x) \cdot (1 - g(x))}$$

Final cal of Derivative

$$\boxed{\frac{d}{dw} u^n = n u^{n-1} \cdot \frac{du}{dw}}$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \Sigma \left(\underline{y - g\left(\Sigma w_i x_i\right)} \right)^2$$

$$= 2 \cdot \Sigma \underbrace{\left(y - g\left(\Sigma w_i x_i\right)\right)}_{\delta}^{2-1} \cdot \frac{\partial}{\partial w_i} \left(y - g\left(\Sigma w_i x_i\right)\right)$$

$$= 2 \cdot \Sigma \delta_i \cdot \left(0 - \frac{\partial}{\partial w_i} \left(g\left(\Sigma w_i x_i\right)\right)\right)$$

$$= -2 \Sigma \delta_i \cdot \left(g\left(\Sigma w_i x_i\right) \cdot \left(1 - g\left(\Sigma w_i x_i\right)\right)\right) \frac{\partial}{\partial w_i} \left(\Sigma w_i x_i\right)$$

$$= -2 \Sigma \delta_i \cdot \mathrm{Der}\left(g\left(\Sigma w_i x_i\right)\right) \cdot x_i$$

$$\boxed{\frac{\partial E}{\partial w_i} = -2 \Sigma \delta_i \cdot \mathrm{Der}\left(\mathrm{Act}\right) \cdot x_i}$$

Finally....phew!, we have the dE/dWi

$$\boxed{\frac{\partial E}{\partial w_i} = -2\sum \left(y - g\left(\sum w_i x_i\right)\right) \cdot \left(g\left(\sum w_i x_i\right)\left(1 - g\left(\sum w_i x_i\right)\right)\right) \cdot x_i}$$

$$\Downarrow$$

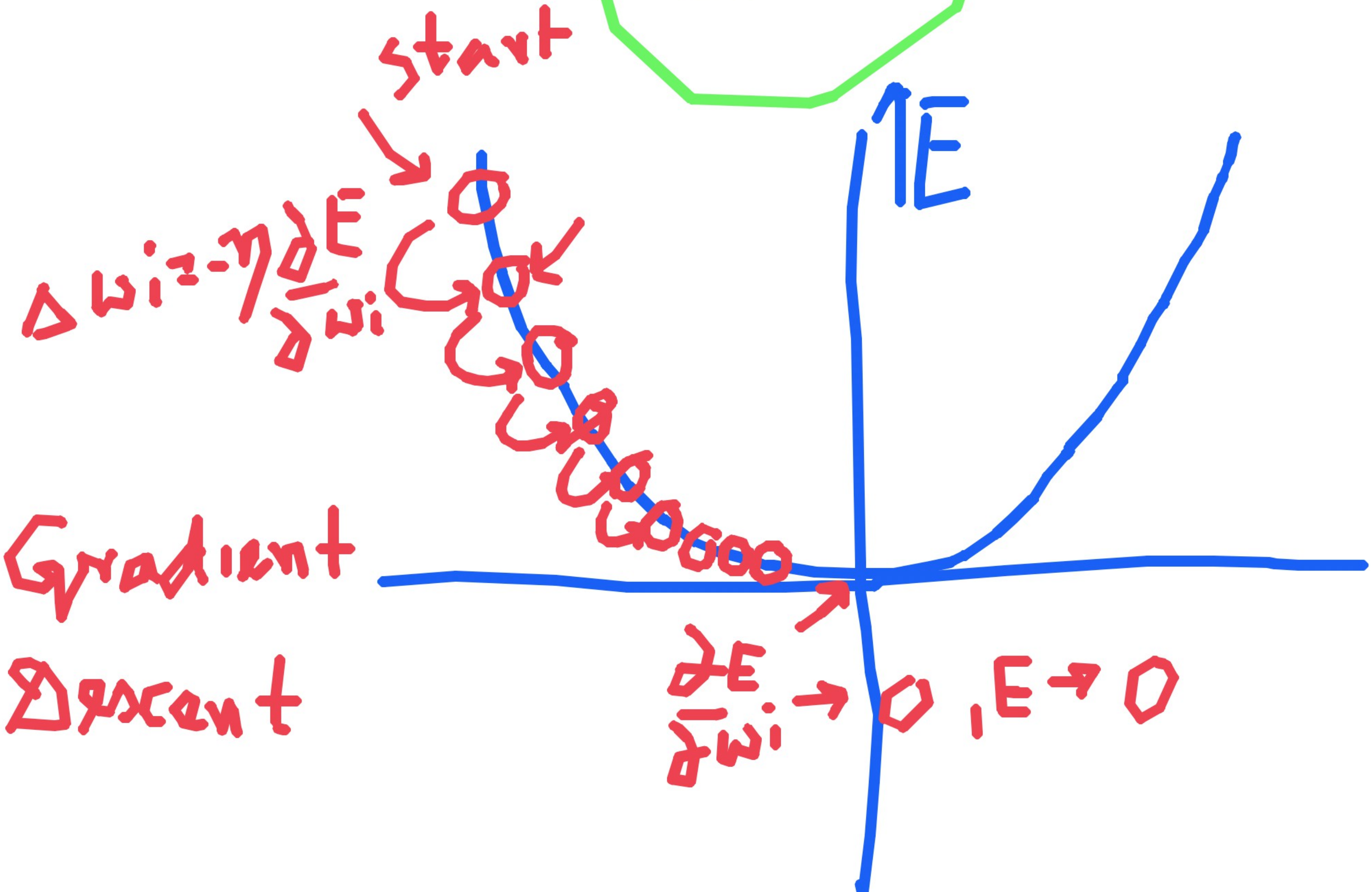
$$\boxed{\frac{\partial E}{\partial w_i} = -2\sum \left(\text{Dev from Truth}\right) \cdot \text{Dev}\left(\text{Act}\right) \cdot x_i}$$

$\frac{\partial E}{\partial w_i} = 0$, if any of these terms is 0.

Apply Backprop

$$\boxed{\Delta w_i = -\eta \frac{\partial E}{\partial w_i}}$$

Learning Rate



- $Y = g(z)$
- $d/dx\,(y\ ) = d/dx(g(z\ )) = g'(z).d/dx(z)$