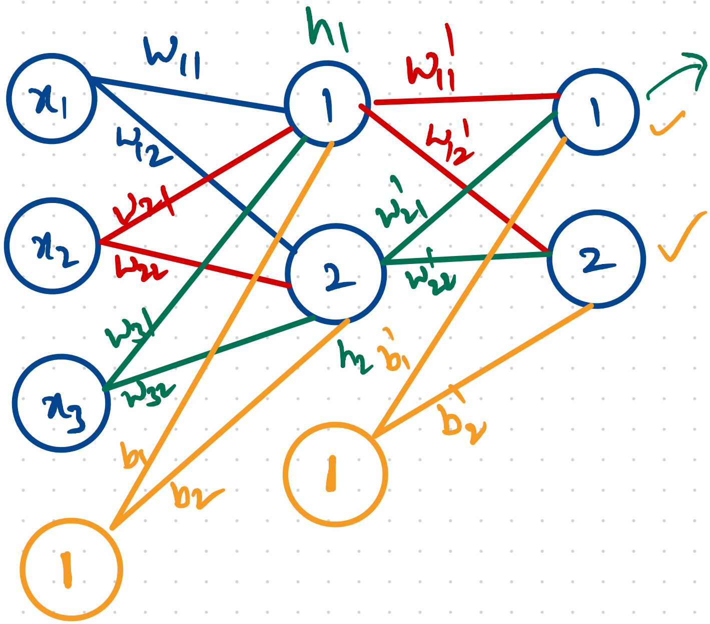
Day-7 : 29 June 2023

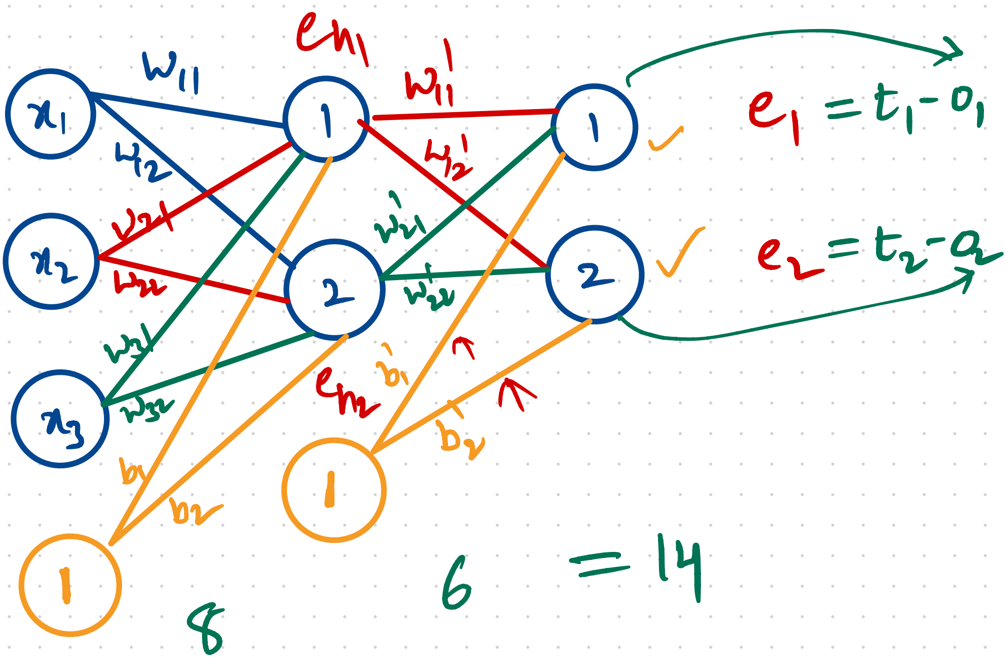


$$\sigma\left(\underbrace{\sum w'_{j1} h_j + b_j}\right)$$

$$I \xrightarrow{W_{ih}} H \xrightarrow{W_{ho\,2\times2}} O$$

$$O = \sigma\left(W_{ho} H_n + B_n\right)$$

$$\boxed{O = \sigma\left(\underset{2\times2}{W_{ho}}\,\sigma\left(\underset{2\times3}{W_{ih}}\,\underset{3\times1}{x_i} + \underset{2\times1}{B_i}\right) + \underset{2\times1}{B_n}\right)}$$



$$I \xleftarrow{W_{ih}^T} H \xleftarrow{W_{ho}^T} O$$

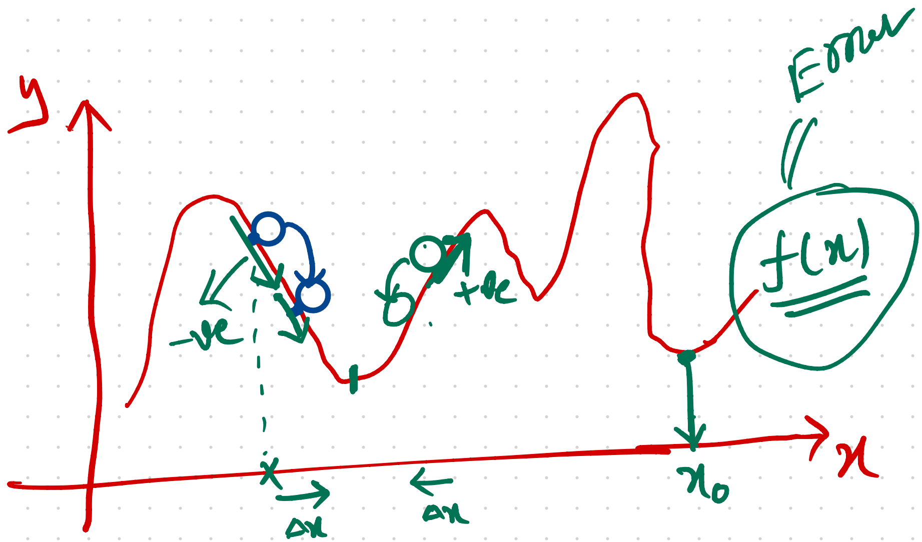
$$e_h = \begin{bmatrix} e_{h_1} \\ e_{h_2} \end{bmatrix} = \begin{bmatrix} w'_{11} & w'_{12} \\ w'_{21} & w'_{22} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = W_{ho}^T e$$

$$e_{h'_1} = b'_1 e_1 \,, \quad e_{h'_2} = b'_2 e_2$$

"Gradient Descent"

$\updownarrow$

"Neural Networks"



$$\frac{dE}{dw_{jk}} = \text{gradient} = \begin{cases} -ve, & a+\Delta x \\ +ve, & a-\Delta x \end{cases}$$

$$\boxed{w_{jk}^{new} = w_{jk}^{old} - \alpha \frac{dE}{dw_{jk}}}$$

$$E = \sum_{k} (t_k - o_k)^2$$

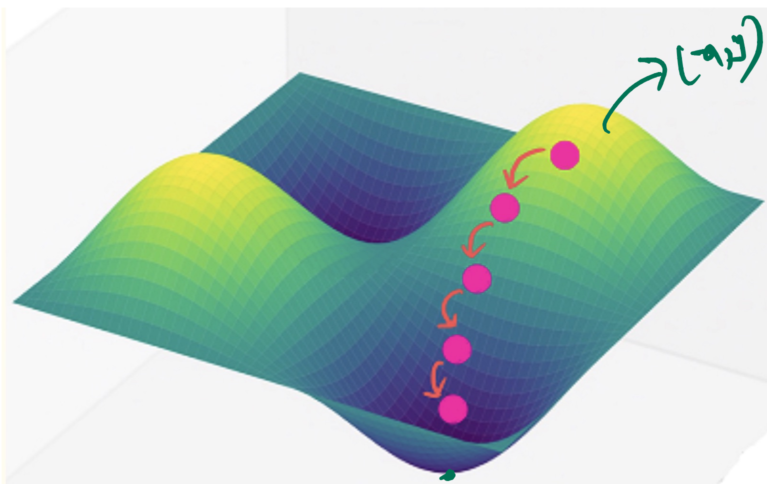
$t_1 - o_1$

$t_2 - o_2$

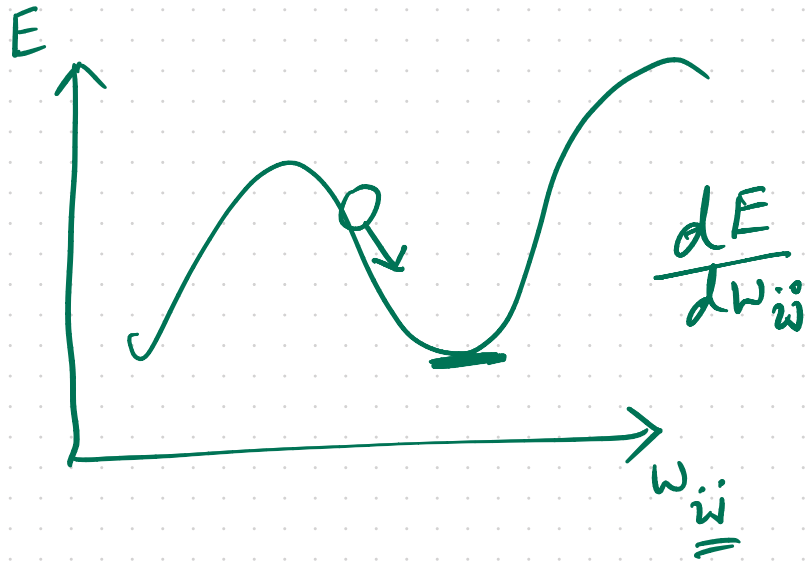
$$\underline{\underline{E}}(w_{11}, w_{12}, w_{21}, w_{22}, w_{31}, w_{32}, w'_{11}, w'_{12}, w'_{21}, w'_{22}, b_1, b_2, b'_1, b'_2) = \underbrace{\sum_{k} (t_k - o_k)^2}$$

F

14-D space



$$\boxed{\underline{\underline{w}} = \text{Argmin}(E)}, \quad \min(E) \approx 0$$



$$\frac{dE}{dw_{ij}}$$

The error function $E: \mathbb{R}^m \to \mathbb{R}$, where m is total number of weights on the edges connected among layers.

E is depends on each weight

$$\overline{w} = (w_{11}, w_{12}, \ldots\ldots w_{21}, w_{22} \ldots)$$

$$\boxed{E(\overline{w}) = \sum_{k=1}^{s} (t_k - o_k)^2}$$

find $\overline{w}_0 \ni E(\overline{w}_0) \approx 0$

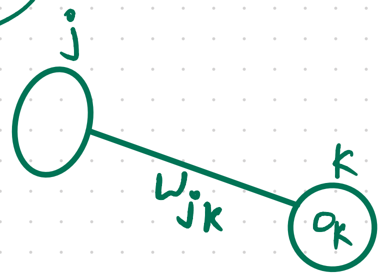
$$\boxed{\overline{w}_{n+1} = \overline{w}_n - \alpha \nabla E(\overline{w}_n)}$$

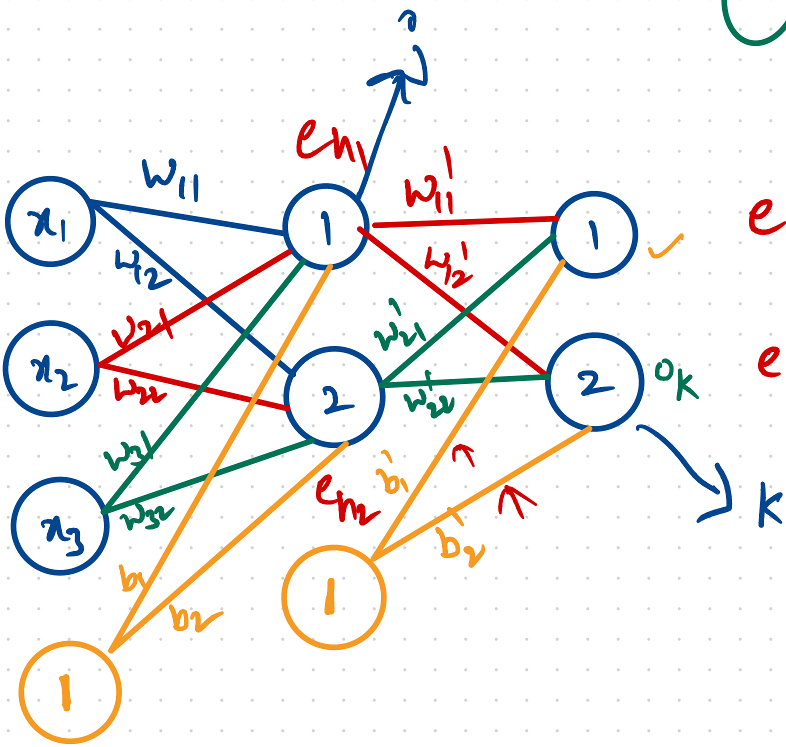
$$E = \sum_{i=1}^{s} (t_i - o_i)^2$$

$$(t_1 - o_1)^2 + \ldots \ldots (t_k - o_k)^2 + \ldots$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{jk}} \quad ✓$$

$$\frac{\partial E}{\partial w_{jk}} = -2(t_k - o_k)\frac{\partial o_k}{\partial w_{jk}}$$





$e_1 = t_1 - o_1$

$e_2 = t_2 - o_2$

~~80~~ 90%

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = -2(t_k - o_k) \cdot \frac{\partial o_k}{\partial w_{jk}}$$

$$o_k = \underbrace{\sigma\left(\sum_{i=1}^{2} w_{ik} h_i + b_k'\right)}_{\sigma_k}$$

$$\frac{\partial o_k}{\partial w_{jk}} =$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

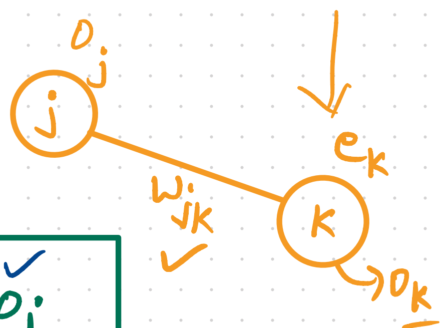
$$\sigma'(x) = \sigma(x)(1-\sigma(x))$$
$$= \sigma(1-\sigma)$$

$$\boxed{\frac{\partial o_k}{\partial w_{jk}} = \sigma_k(1-\sigma_k) \cdot h_j}$$

$$\frac{\partial E}{\partial w_{jk}} = -2(t_k - o_k) \cdot \frac{\partial o_k}{\partial w_{jk}}$$

$$\boxed{\frac{\partial E}{\partial w_{jk}} = -2\,\underbrace{(t_k - o_k)}_{e_1}\, \sigma_k (1-\sigma_k) \cdot h_j}$$

$$\boxed{\frac{\partial E}{\partial w_{jk}} = -(e_k)\, \sigma_k (1-\sigma_k) \cdot h_j}$$



$$\boxed{\frac{\partial E}{\partial w_{jk}} = -\underline{e_k}\, \underline{o_k (1-o_k)} \cdot o_j}$$

$$\boxed{w_{jk}^{new} := w_{jk}^{old} - \alpha \frac{\partial E}{\partial w_{jk}}}$$ ✓

α – learning rate

$$\boxed{\frac{\partial E}{\partial b_k} = -e_k O_k \underline{(1-O_k)} \cdot 1}$$

$$\boxed{b_k^{new} = b_k^{old} - \alpha \frac{\partial E}{\partial b_k}}$$

1. Setup network
2. Input – forward propagation
3. Error – Back propagation
4. Update weights (Gradient Descent)
5. <u>Repeat</u> (2, 3, 4) till reach solution