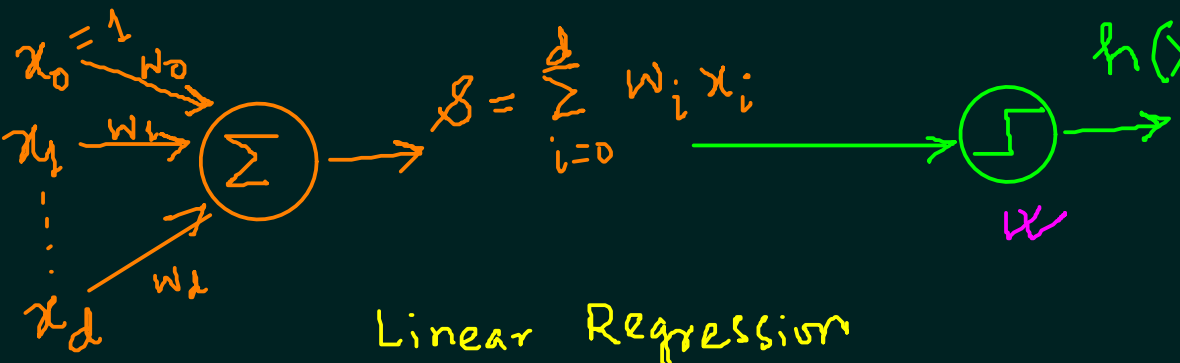


Linear Regression

Linear Classification

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} e\left(h(x_n), y_n\right)$$

$$\hookrightarrow E_{in}(w) = \frac{1}{N}\sum_{n}\left(y_n - w^T x_n\right)^2$$

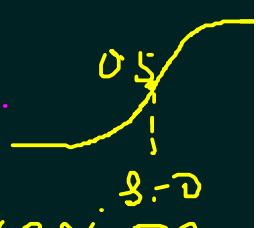
- Directly 1-step: $(X^T X)^{-1} X^T \leftarrow X'$
- ① Or Iteratively ⇝ G.D. (batch)

② $$\frac{\nabla E_{in}(w)}{\partial w_k} = 0 \Rightarrow \frac{2}{N}\sum_{n}\left(y_n - w^T x_n\right)\cdot\left(-x_{nk}\right)$$

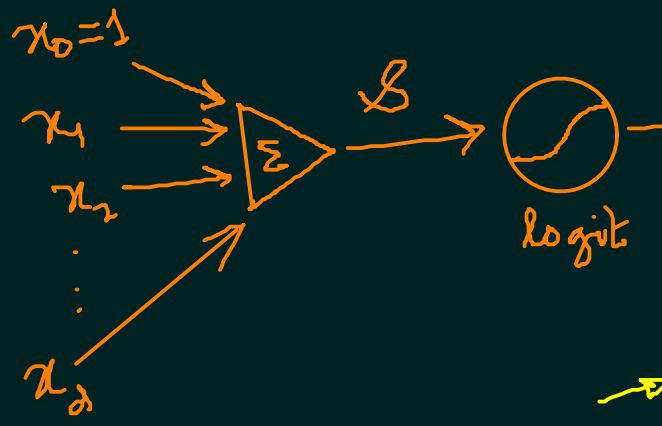
$$w_k \leftarrow w_k^{old} - \eta\,\frac{\partial \nabla E_{in}(w)}{\partial w_k} \leftarrow \text{weight update}$$ (LMS, Delta)

↳ (x_i, y_i) ← one ex at a time ⇝ Stochastic G.D.

← Logistic Regression

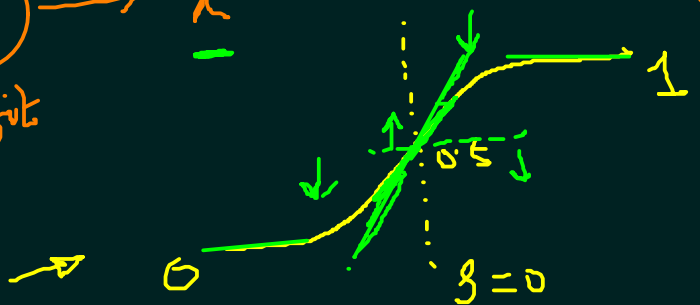


$\Rightarrow P(y=0) = ?\quad P(y=1)?$



$x = \theta(s) \rightarrow \frac{1}{1+e^{-s}}$

$\tanh(s) = \frac{e^{s}-e^{-s}}{e^{s}+e^{-s}}$

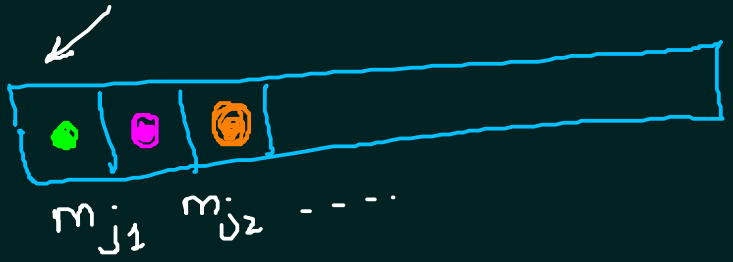


Ex: Predict H. Attack

Ex: Movie Rating

u_i User choice

m_j Movie Attr.



0 - 1 ← Rating r_{ij} ↔

$$e_{ij} = \left(r_{ij} - \sum_{k=1}^{K} u_{ik} v_{jk}\right)^2$$

← error for each training data (u_i, v_j, r_{ij})

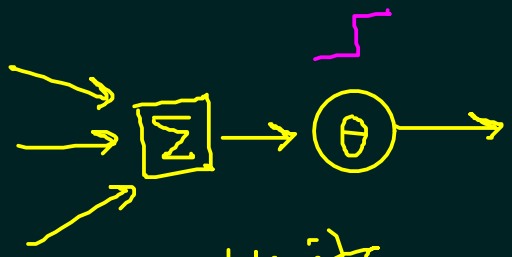
→ Minimize (SGD)/(GD)

→ "epoch" → $x_1 \cdots x_2 \cdots x_3 \cdots$ } Cross-validation

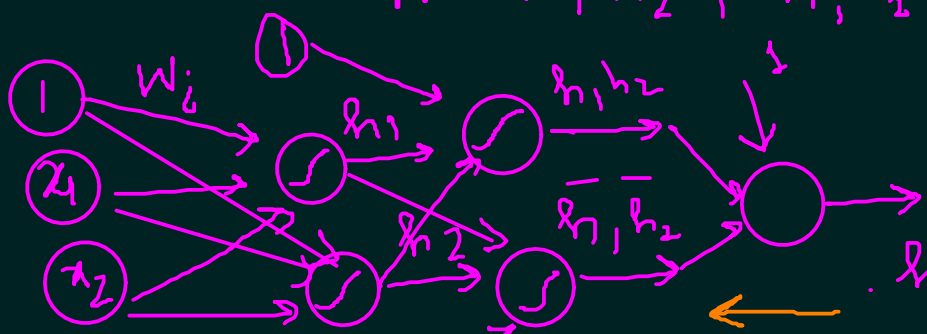
Possible ij prob.

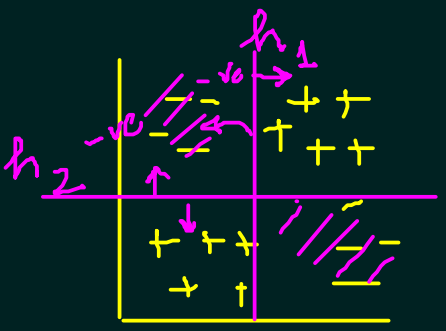
$x_1 x_4 x_5$ $x_2 x_4 x_6$

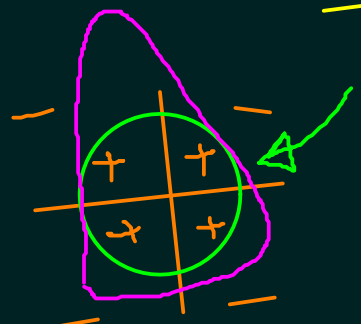
$\Delta w_{ij} \approx 0$ ⇝ $\eta = 0.1, 0.01$ ⇄

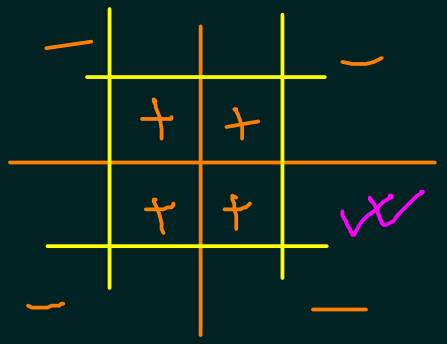


units of NN

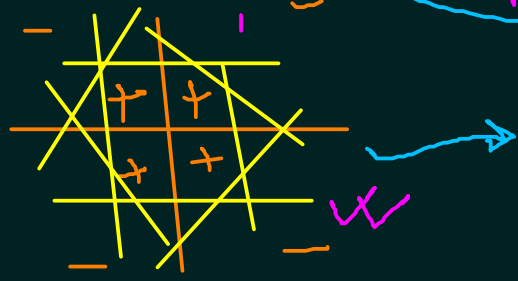


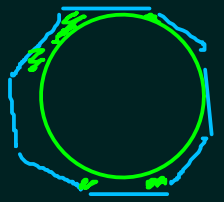


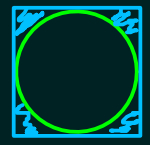




MLP







FWD

4 LS/PT ⇝ 8 PT/LS

$$\frac{\partial E}{\partial s} \cdot \frac{\partial s}{\partial w_k}$$

opt. difficult

generalization diff

$$\frac{E_{in}(w)}{\partial w_k} = \frac{2}{N} \sum_n \left(y_n - \theta(s_n) \right) \left(-\theta'(s_n) \right) \{ x_{nk} \}$$

$$\theta(s) = \frac{1}{1+e^{-\sum w_i x_i}}$$

$x = \theta(s)$

(x_n, y_n)

$$\left[w_k = w_k - \eta \left[\frac{\partial L}{\partial w_k} \right] \right]$$

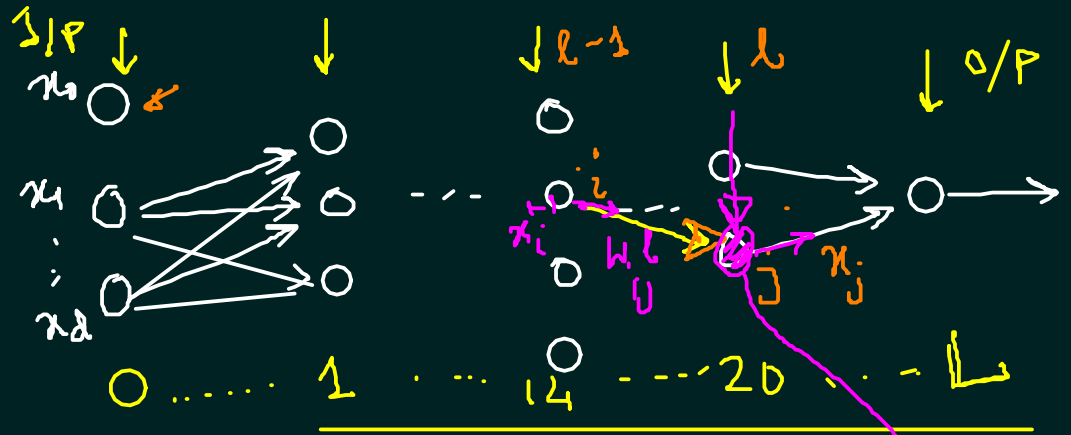
$$\theta(s) = \frac{1}{1+e^{-s}}$$

$$\theta'(s) = \frac{1}{1+e^{-s}} \cdot \frac{e^{-s}}{1+e^{-s}}$$

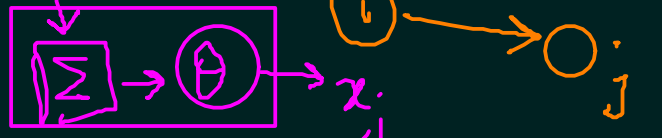
$$\theta(s)\left[1-\theta(s)\right]$$

one unit

$$err = \frac{1}{N} \sum_n \left(y_n - \theta(s_n) \right)^2$$



$$w_{ij}^{(l)} \qquad \begin{cases} 1 \le l \le L \\ 0 \le i \le d^{(l-1)} \\ 1 \le j \le d^{(l)} \end{cases}$$



Notation

$$x_j^{(l)} = \theta\left(s_j^{(l)}\right)_{d^{(l-1)}}$$

$$= \theta\left(\sum_{i=0} w_{ij}^{(l)} \cdot x_i^{(l-1)}\right)$$

Minimize $E_{in}(w)$ for all $i, j, l \leftrightarrow w_{ij}^{(l)}$

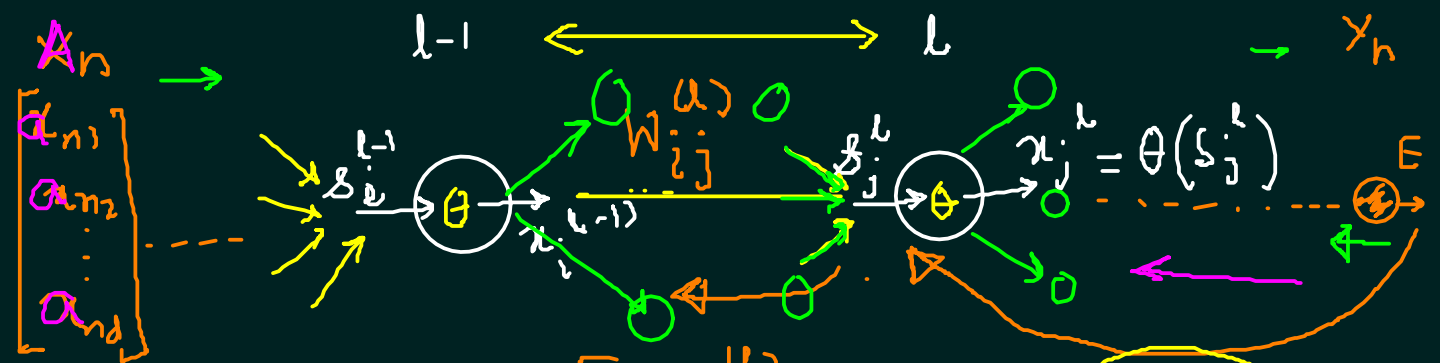
↳ Compute $\dfrac{\partial E_{in}(w)}{\partial w_{ij}^{(l)}}$ $\forall\, i, j, l$

$\Rightarrow$ $w_{ij}^{l} \leftarrow w_{ij}^{l}(\text{old}) - \eta \dfrac{\partial E_{in}(w)}{\partial w_{ij}^{l}}$

→ Fwd e

← Bck Δw_{ij}^{l}

Compute $\frac{\partial E(w)}{\partial w_{ij}^{(l)}}$



$$\frac{\partial E(w)}{\partial w_{ij}^{(l)}} = \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} * \frac{\partial E(w)}{\partial s_j^{(l)}}$$

$$\frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = x_i^{(l-1)}$$

$\delta_j^{(l)}$ (Recursion)

(X_n, Y_n)

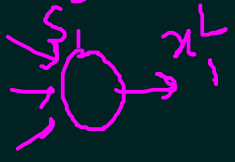
→ For final layer $l = L$, $j = 1$

BASE

$$\delta_1^{(L)} = \frac{\partial E(w)}{\partial s_1^{(L)}} = \frac{\partial}{\partial s_1^L}\left[\frac{1}{2}\sum(\theta(s_1^L) - y_n)^2\right]$$

$$E(w) = \frac{1}{2}(x_1^L - y_n)^2$$

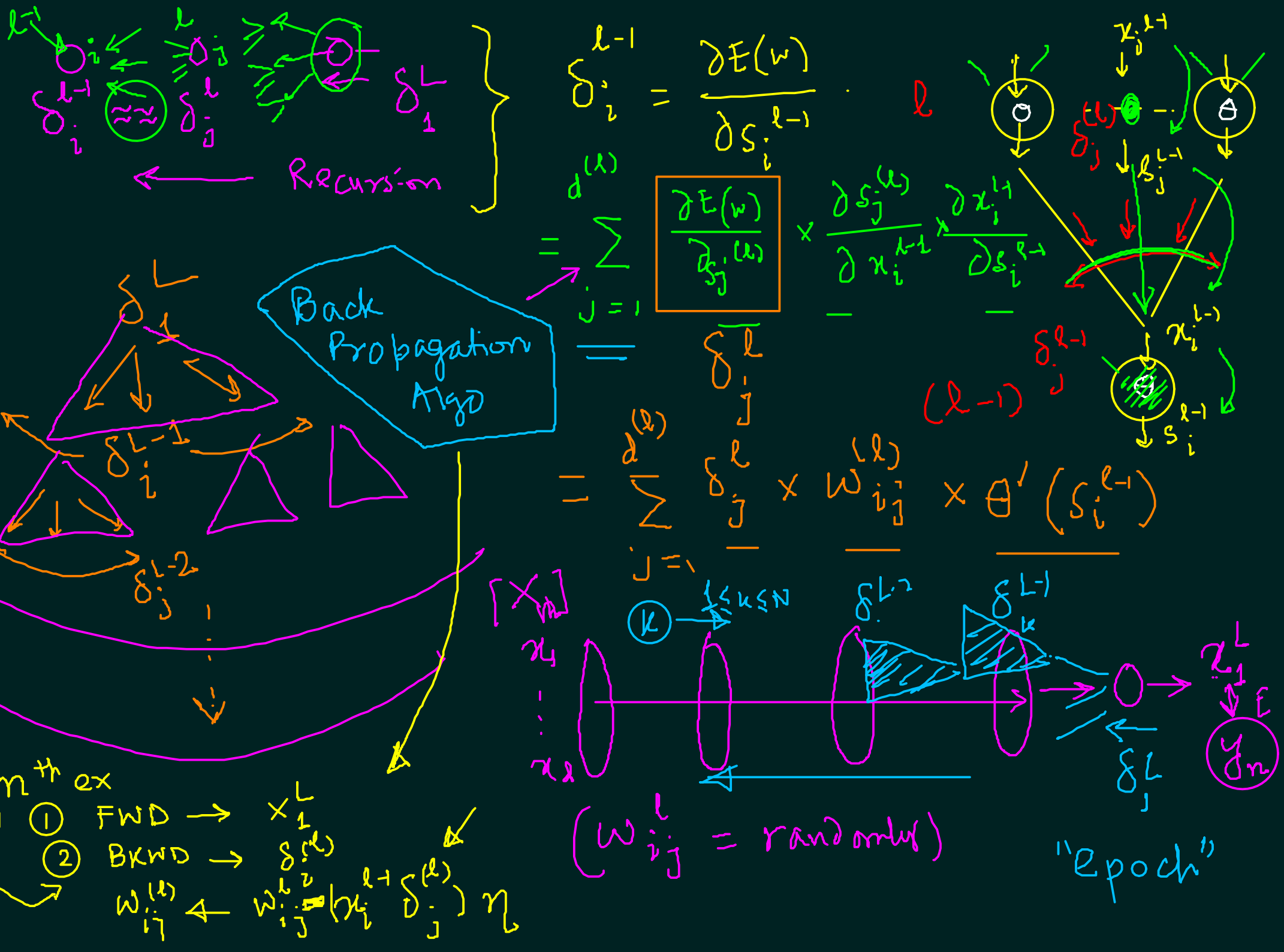
$$x_1^L = \theta(s_1^L)$$



last layer (direct)

$$= (\theta(s_1^L) - y_n)\,\theta(s_1^L)(1 - \theta(s_1^L))$$

$$= (\theta(s_1^L) - y_n)(x_1^L)(1 - x_1^L)$$



$$\delta_i^{l-1} = \frac{\partial E(w)}{\partial s_i^{l-1}}$$

$$= \sum_{j=1}^{d^{(l)}} \frac{\partial E(w)}{\partial s_j^{(l)}} \times \frac{\partial s_j^{(l)}}{\partial x_i^{l-1}} \times \frac{\partial x_i^{l-1}}{\partial s_i^{l-1}}$$

$$= \sum_{j=1}^{d^{(l)}} \delta_j^{l} \times W_{ij}^{(l)} \times \theta'(s_i^{l-1})$$