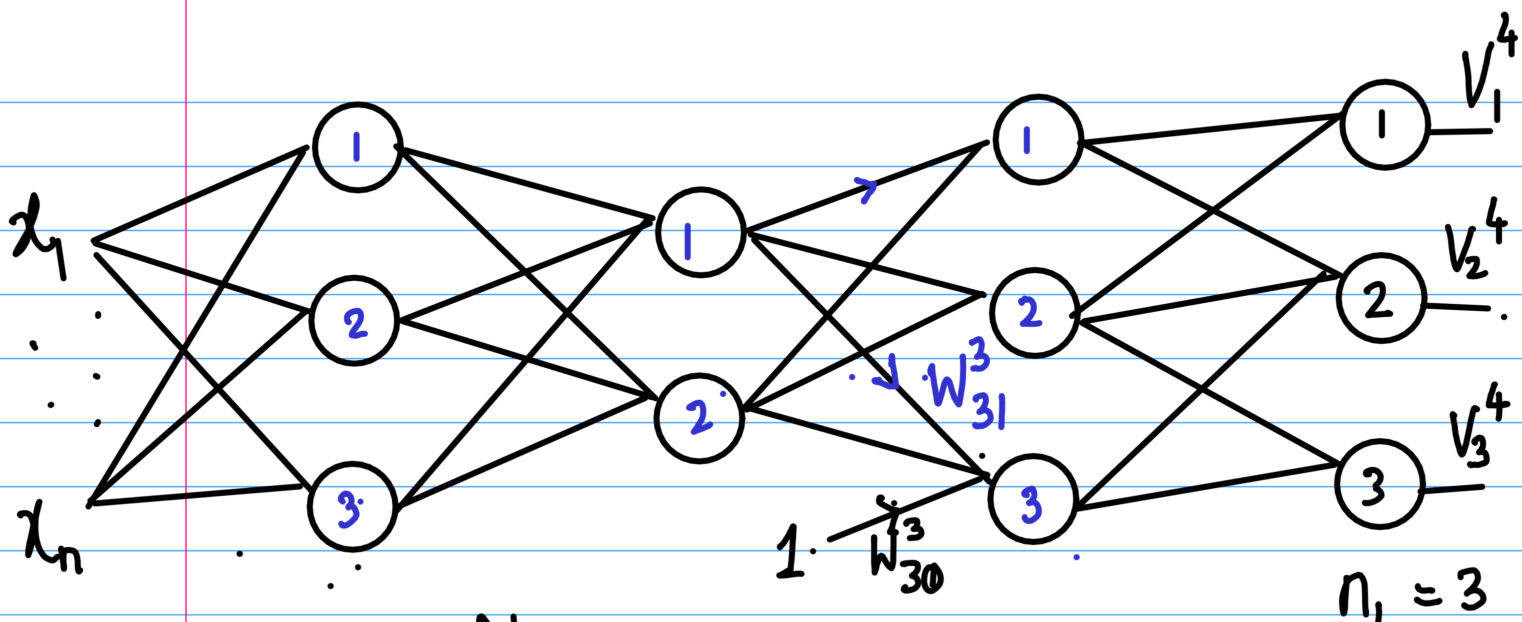



$L$ : no. of layers $\quad L = 4$

$n_i$ : no. of neurons in layer 'i'

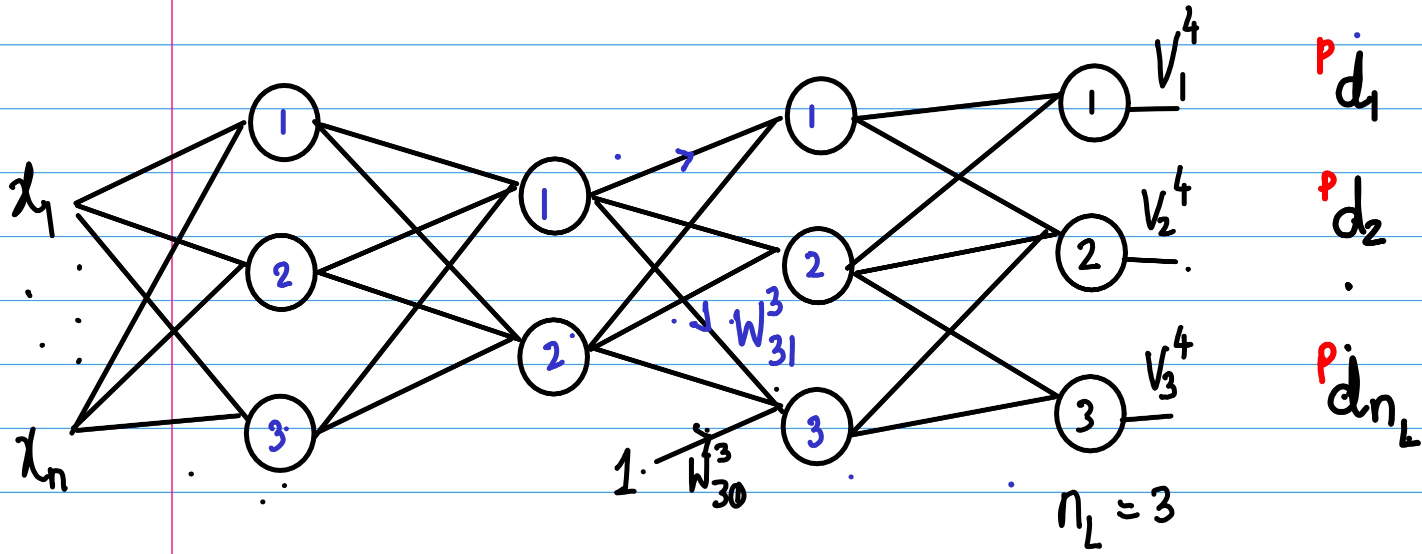
${}^pV^i_j$ : output of neuron j in layer i when $x^p$ is presented at the input

$x^p$ : $p^{th}$ pattern

$x^p_i$ : $i^{th}$ feature/attribute/co-ordinate of the p-th pattern

1 1 / \ 9 9

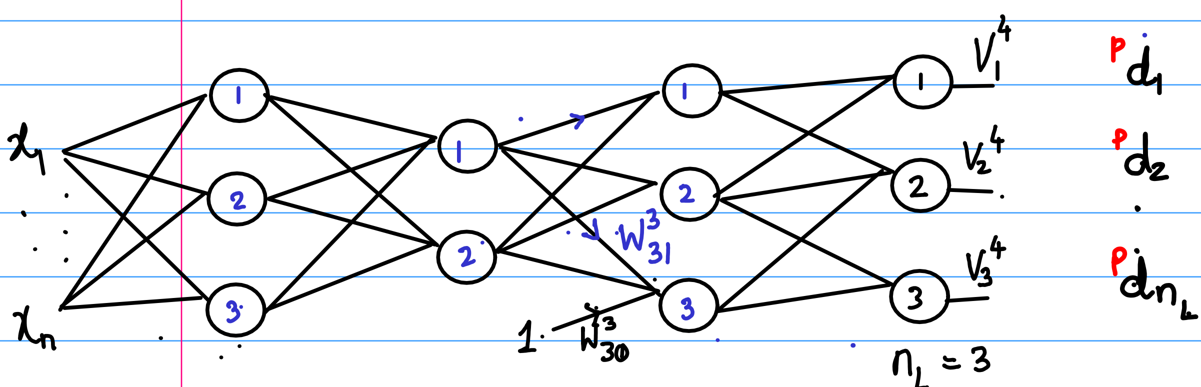
6 6 6 - . .




$$\min \frac{1}{2}\frac{1}{P}\frac{1}{n_L}\sum_{p=1}^{P}\sum_{i=1}^{n_L}\left({}^p d_i - {}^p V_i^L\right)^2$$

loss function
MSE
squared loss

cross entropy

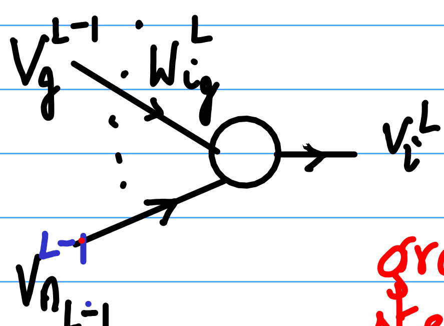




$$\min \frac{1}{2}\frac{1}{P}\frac{1}{n_L}\sum_{p=1}^{P}\sum_{i=1}^{n_L}\left({}^p d_i - {}^p V_i^L\right)^2$$

loss function
MSE
squared loss

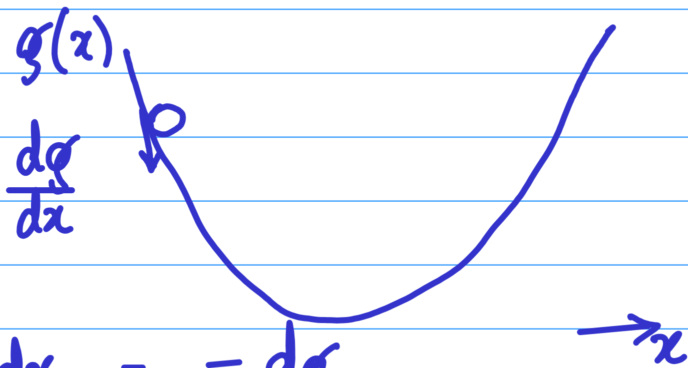

$W_{ig}^L$




gradient descent
steepest descent

$$ {}^{p}V_i^{L} = f\left({}^{p}net_i^{L}\right) $$

$$ E = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left({}^{p}d_i - {}^{p}V_i^{L}\right)^2 $$




$$ \dot{x} = \frac{dx}{dt} = -\frac{dg}{dx} $$

$$ x = (x_1, x_2)^T \qquad g(x_1, x_2) $$

$$ \frac{dx_1}{dt} = -\frac{\partial g}{\partial x_1} $$

$$ \frac{dx_2}{dt} = -\frac{\partial g}{\partial x_2} $$

$$ \frac{d\underline{x}}{dt} = \underline{\dot{x}} = -\nabla g(x) $$

$$ {}^{p}V_i^{L} = f\left({}^{p}net_i^{L}\right) $$

$$ E = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left({}^{p}d_i - {}^{p}V_i^{L}\right)^2 $$

$$ {}^{p}net_i^{L} = \sum_{q=1}^{n_{L-1}} W_{iq}^{L}\, {}^{p}V_q^{L-1} + W_{i0}^{L} $$

$$ {}^{p}net_i^{L} = \sum_{q=0}^{n_{L-1}} W_{iq}^{L}\, {}^{p}V_q^{L-1} $$

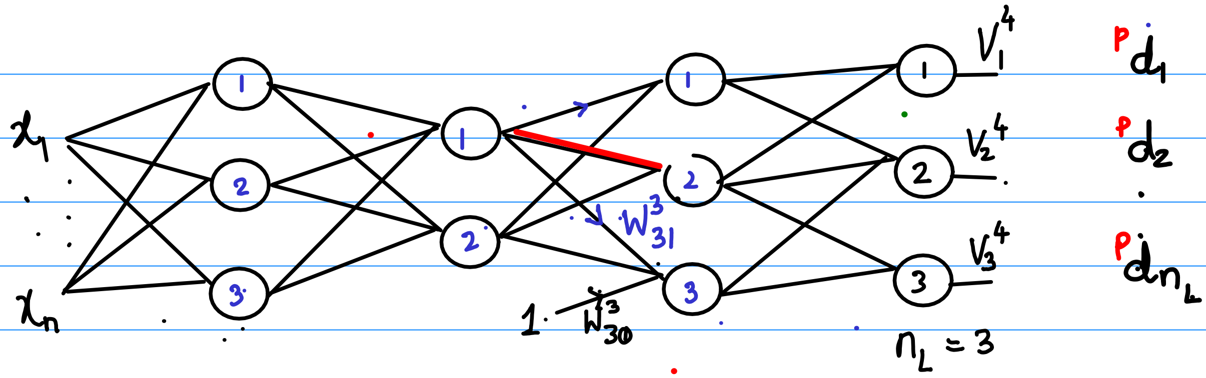
$${}^{p}V_i^L = f({}^{p}net_i^L)$$

$$E = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left({}^{p}d_i - {}^{p}V_i^L\right)^2$$

$${}^{p}net_i^L = \sum_{g=0}^{n_{L-1}} W_{ig}^L \; {}^{p}V_g^{L-1}$$

$$\frac{dW_{kg}^L}{dt} = -\frac{\partial E}{\partial W_{kg}^L}$$

$$= \frac{1}{2} \cdot 2 \cdot \sum_{p=1}^{P} \left({}^{p}d_k - {}^{p}V_k^L\right) \cdot f'({}^{p}net_k^L) \cdot {}^{p}V_g^{L-1}$$




$$E = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left( {}^p d_i - {}^p V_i^L \right)^2$$

$$\frac{d W_{21}^3}{dt} = -\frac{\partial E}{\partial W_{21}^3} \qquad {}^p V_i^L = f({}^p net_i^L)$$

$$= \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left( {}^p d_i - {}^p V_i^L \right) \cdot \frac{d\, {}^p V_i^L}{d W_{21}^3}$$

$$= \sum_{p=1}^{P} \sum_{i=1}^{n_L} \left( {}^p d_i - {}^p V_i^L \right) \cdot f'({}^p net_i^L) \cdot \boxed{\frac{d\, {}^p net_i^L}{d W_{21}^3}}$$

$${}^p net_i^L = W_{i1}^L \cdot {}^p V_1^{L-1} + W_{i2}^L \, {}^p V_2^{L-1} + W_{i3}^L \, {}^p V_3^{L-1}$$

$${}^p net_i^4 = W_{i1}^4 \, {}^p V_1^3 + W_{i2}^4 \, {}^p V_2^3 + W_{i3}^4 \, {}^p V_3^3 + W_{i0}^4$$

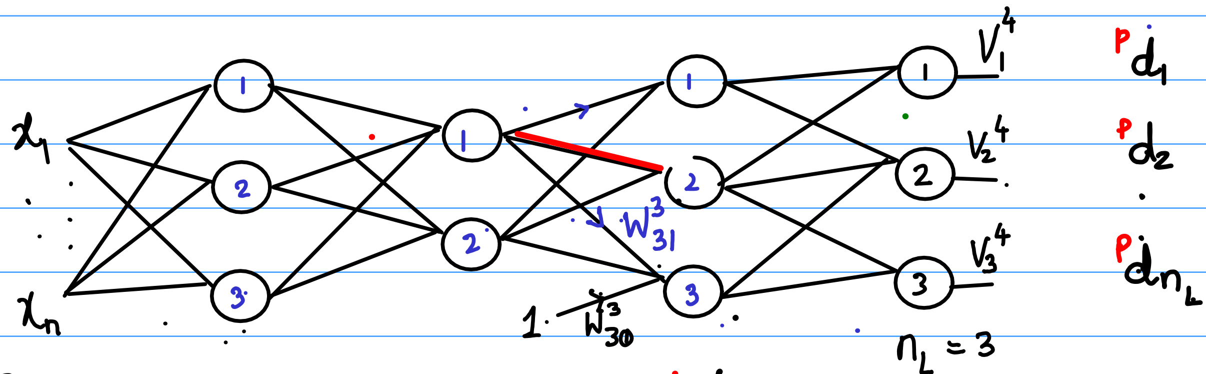
$$\frac{d\, {}^p net_i^L}{d W_{21}^3} = W_{i2}^4 \cdot \frac{d\, {}^p V_2^3}{d W_{21}^3} = W_{i2}^4 \cdot f'({}^p net_2^3) \cdot \frac{d\, {}^p net_2^3}{d W_{21}^3}$$

$${}^p V_2^3 = f({}^p net_2^3)$$

$${}^p net_2^3 = W_{21}^3 \, {}^p V_1^2 + W_{22}^3 \, {}^p V_2^2 + W_{20}^3$$

$$^{p}net_2^3 = W_{21}^3\, {}^{p}V_1^2 + W_{22}^3\, {}^{p}V_2^2 + W_{20}^3$$

$$\frac{d\, {}^{p}net_2^3}{d\, W_{21}^3} = {}^{p}V_1^2$$




$$\delta_2^3 = (d_1 - V_1^4)\, f'(net_1^4) \cdot W_{12}^4$$
$$+ (d_2 - V_2^4)\, f'(net_2^4) \cdot W_{22}^4$$
$$+ (d_3 - V_3^4)\, f'(net_3^4) \cdot W_{32}^4$$

$$\frac{d\, W_{21}^3}{dt} = \delta_2^3 \cdot f'(net_2^3) \cdot V_1^2$$

$$\frac{d W_{21}^3}{dt} \approx \frac{W_{21}^3(t+\Delta t) - W_{21}^3(t)}{\Delta t} = \frac{-\partial E}{\partial W_{21}^3}$$

$$\frac{d W_{ij}}{dt} \approx \frac{W_{ij}(t+\Delta t) - W_{ij}(t)}{\Delta t} = -\frac{\partial E}{\partial W_{ij}}$$

$$\Rightarrow W_{ij}(t+\Delta t) = W_{ij}(t) + \Delta t \left[ \frac{-\partial E}{\partial W_{ij}} \right]$$

$$W_{vy}(k+1) = W_{vy}(k) + \eta \left[ \frac{-\partial E}{\partial W_{vy}} \right]$$

$\eta \rightarrow$ learning rate.