e______ MTWTFS

$$y = mx + c$$

slope — y-intercept

$$h(x) = w_1 x + w_0$$

→ w_0 and w_1 are unknown coefficients

→ w_1 has multiplicative or dividing effect on input x, whereas w_0 is an additive term.

→ We can also calculate these constants values from Inversion

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

$n \times 1$ $m \times n$

$$\vec{y} = A x$$

$$A = \vec{y} x^{-1}$$

Loss Function

$y \rightarrow est_i$ $y \rightarrow$ original value

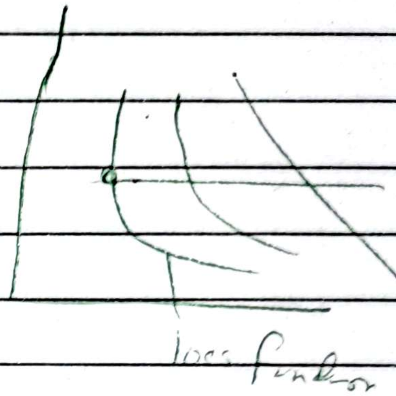
$h(x) \rightarrow$ estimated value

$$y - (\theta_0 + \theta_1 x_1 + \theta_2 x_2) = 0$$

$$y - h(x) = 0$$

or

$$h(x) - y = 0$$



So

$$L(y, h(x)) = ((y - h(x))^2$$

Other loss functions:

$(y - h(x))$ error

$|y - h(x)|$ absolute error

$\sqrt{(y - h(x))^2}$ root squared error

For multiple values
total loss or total cost

$$\frac{1}{N} \sum_{j=1}^{N-1} (y_j - h(x_j))^2$$

Expanded form

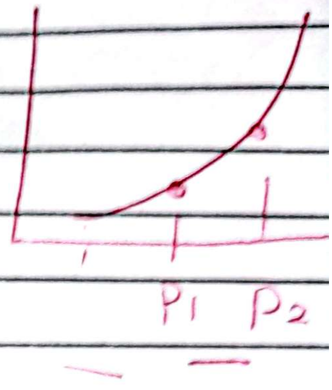
$$\frac{1}{N} \sum_{j=1}^{N-1} (y_j - (w_1 x_j + w_0))^2$$

Gradient Descent

ignore the limit ... equal

$$\frac{dp}{dt} = \lim_{\Delta t \to 0} \frac{P_2 - P_1}{t_2 - t_1} = rP_1$$

$$\frac{dp}{dt} = \frac{P_2 - P_1}{\Delta t} = \gamma_t P_1$$



$$\frac{dp}{dt} = \frac{P_2 - P_1}{t_2 - t_1} = \gamma_t P_1$$

$$t_1 - t_0 = t_2 - t_1 = \Delta t$$

$$P_2 = \Delta t \, \gamma P_1 + P_1$$

$$P_2 = P_1 + \Delta t \, \gamma P_1$$

$$P_2 = P_1 + \Delta t \frac{dp}{dt}$$

new value = oldvalue + learning rate (slope)

for ε

$$\varepsilon_j = \varepsilon_{j-1} + \alpha \frac{d\varepsilon}{d\,steps}$$

$$\varepsilon_j = \varepsilon_{j-1} + \alpha \frac{d\varepsilon}{d\,\theta_1 \theta_2}$$

$$\varepsilon_j = \varepsilon_{j-1} + \alpha \frac{d\varepsilon}{\partial w_1 w_2}$$

$$\frac{\partial}{\partial w_1} (y - h(x))^2$$

$$\frac{\partial}{\partial w_1} (y - (w_1 x + w_0))^2 = \partial (w_1 x + w_0)$$

$$= 2(y - (w_1 x + w_0))(-x)$$

$$\frac{\partial}{\partial w_0}\left(y-(w_1x+w_0)\right)^2 =$$

$$2\left(y-(w_1x+w_0)\right)$$

$$w_0 = w_0 + \alpha\, 2\left(y-(w_1x+w_0)\right)$$

$$w_1 = w_1 + \alpha\, 2\left(y-(w_1x+w_0)\right)(-x_i)$$

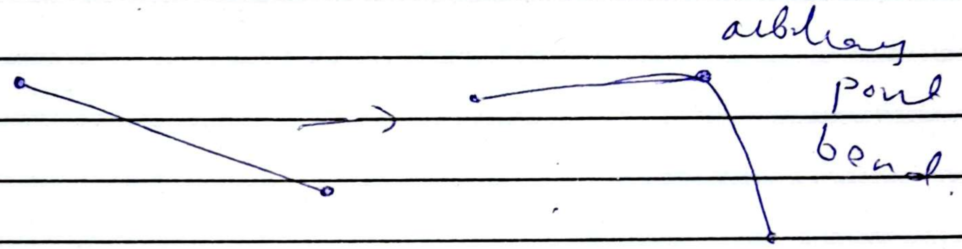
~~Regression with~~

```
For i in range (num_iter):
    m_grad = 0
    b_grad = 0
    N = float (len (points))
    total_cost = 0
    for j in range (0, len(points)):
        x_j = points [j,0]        y_j = points [j,1]
        m_gradient += (1/N) * 2 * x_j (y_j - (m * x_j + b))
        ~~b_gradient += 1/N * 2 * (y_j - (m* x_j + b))~~
        total cost += (y_j - (m* x_j + b)) ** 2
    cost = total_cost / N
    m = m + learning rate * m_gradient
    b = b + learning rate * b_gradient
```

erivative)

Regression with polynomial Models.

linear equation → change → arbitary point



$$x \rightarrow x^2$$

arbitary points depend on bend

polynomial equation

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \cdots\cdots$$

Taylor series

$$\theta_0 + \theta_1 \frac{dx}{d\theta_1} + \theta_2 \frac{dx^2}{d\theta_2^2} + \theta_3 \frac{dx^3}{d\theta^3} +$$

$$\frac{dp}{dt} = \gamma p$$

$$\frac{dp^2}{dt^2} = \gamma$$

$$\frac{d^3}{dt} = 0$$

Date______ M T W T F S

Castelli

Derivative

$$h(x) = w_0 + w_1 x + w_2 x^2$$

$$\frac{\partial}{\partial w_0} \left(y - (w_0 + w_1 x + w_2 x^2)\right)^2 =$$

$$= \partial \left(y - (w_0 + w_1 x + w_2 x^2)\right)$$

$$\frac{\partial}{\partial w_1} \left(y - (w_0 + w_1 x + w_2 x^2)\right)^2 =$$

$$\partial \left(y - (w_0 + w_1 x + w_2 x^2)\right) x$$

$$\frac{\partial}{\partial w_2} \left(y - (w_0 + w_1 x + w_2 x^2)\right)^2 = \partial \left(y - (w_0 + w_1 x + w_2 x^2)\right) x^2$$

$$w_0 = w_0 + \alpha * \partial \left(y - (w_0 + w_1 x + w_2 x^2)\right)$$

$$w_1 = w_1 + \alpha * \partial \left(y - (w_0 + w_1 x + w_2 x^2)\right) x$$

$$w_2 = w_2 + \alpha * \partial \left(y - (w_0 + w_1 x + w_2 x^2)\right) x^2$$

```
For i in range (num_iteralions):
    w0_gradient = 0
    w1_gradient = 0
    w2_gradient = 0
    N = float (len (points))
    total_cost = 0

    For j in range (0, len (points)):
        x_j = points [j, 0]
        y_j = points [j, 1]
        w0_gradient += 1/N * 2 * (y_j - (w0 + w1 * x_j + w2 * x_j ** 2))

        w1_gradient += 1/N * 2 * (y_j - (w0 + w1 x_j + w2 x_j ** 2) * x_j )

        w2_gradient += 1/N * 2 * ((y_j - (w0 + w1 x_j + w2 - x_j ** 2)) * x_j ** 2

        ~~cost~~ +
        total cost += (y_j - (w0 + w1 * x_j + w2 x_j ** 2)) ** 2

    cost = total cost / N

    w0 = w0 + learningrate * w0_gradient
    w1 = w1 + learning rate * w1_gradient
    w2 = w2 + learning rate * w2_gradient
```

Date____ MTWTFS

$$\begin{bmatrix} \sum x^0 & \sum x^1 & \sum x^2 \\ \sum x^1 & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \sum x^0_i \\ \sum x^1_i y_i \\ \sum x^2_i y_i \end{bmatrix}$$

$$\begin{bmatrix} \sum x^0 & \sum x^1 & \sum x^2 \\ \sum x^1 & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \sum x^0 * y \\ \sum x^1 * y \\ \sum x^2 * y \end{bmatrix}$$

Univarient

$$h(x) = w_0 + w_1 x$$

multivarient

$$h(x) = w_0 + w_1 x + w_2 x_2 + w_3 x_3 + \dots$$

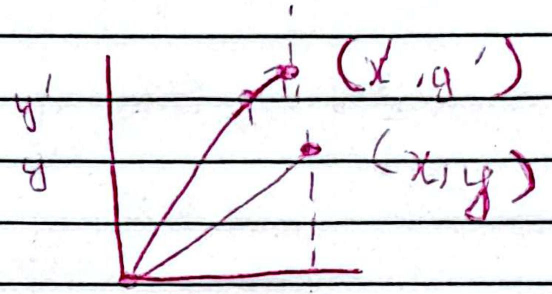
$$= w_0 + \sum_{i=0}^{n} w_i x_i$$

$$h(x) = \sum_{i=0}^{n} w_i x_i$$

Σ

Σ_x

Date______ M T W T F S



$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \; \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$x' = w_{0,0}x + w_{1,0}y$$

$$A \cdot B = |A||B| \cos\theta$$

$\cos(90) = 0$

$\cos(0) = 1$

no effect

$h(x) = w_0 + w_1 x \cos(0)$

$h(x) = w_0 + w_1 x + w_2 x^2 \cos(0)$

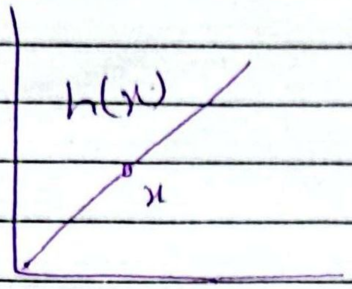
angle value adjusted in the weight we assume that the angle is 0

slope → coefficient value btw 0, 1

$W = [0, 1]$ because angle

Classification

if else



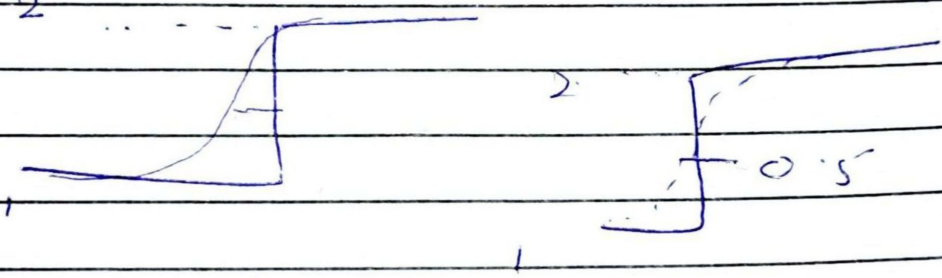
$$h(x) = \begin{cases} 0 & \text{if } x < h(x) \\ 1 & \text{if } x = h(x) \\ 2 & \text{otherwise} \end{cases}$$

Montecarlo Method.

Activation Functions.

Non-linearity

choose based on smooth boundary



Softmax	Sigmoid	tanh
$\cancel{z' = \frac{\lvert z \rvert}{\max(z)}}$	$z' = \frac{1}{1+e^{-z}}$	$z' = \frac{e^{+z} - e^{-z}}{e^{+z} + e^{-z}}$
$z' = \frac{e^{z_i}}{\sum_{j=1} e^{z_i}}$		

Relu $z' = 0, \max(0,z)$

activation function in linear scenario

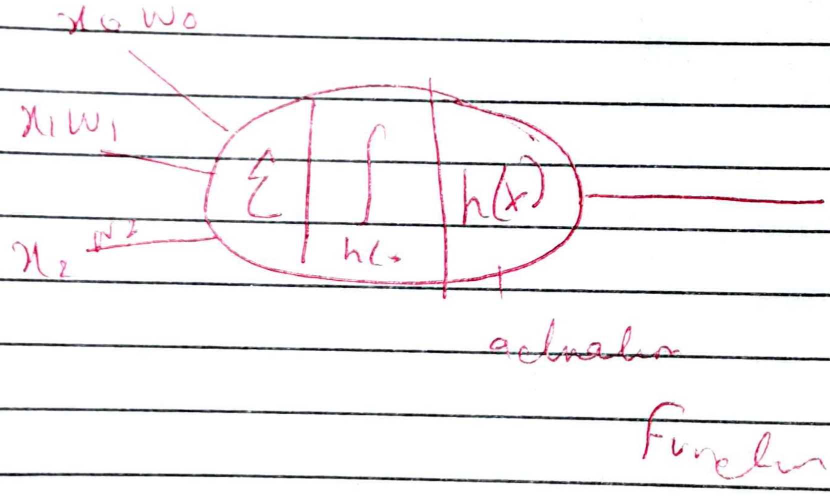
$$h(x) = mx + c$$

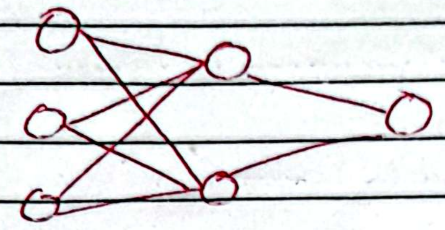
$$= \frac{1}{1+e^{-h(x)}}$$

$$= \frac{1}{1+e^{-(mx+c)}}$$

$$\frac{1}{1+e^{-(w_0 + w_1 x_1 + w_2 x_2 + \cdots)}}$$

Perceptron





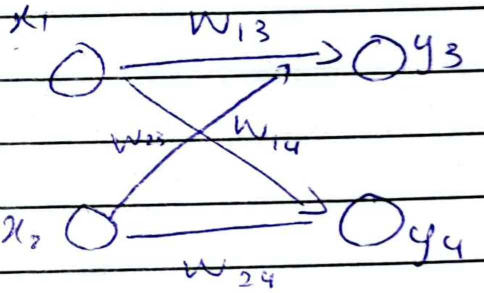
input layer — hidden layer — output layer

neural architecture graph
→ adjacency matrix
→ incident matrix
linked list (slow)
→ tree (complex)

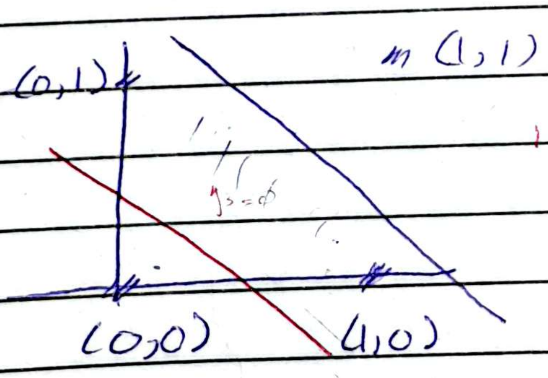
NN architecture — Feed forward (matrix multiplication)
↘ Feed back ward (learning)

Example

x_1	x_2	y_3	y_4
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

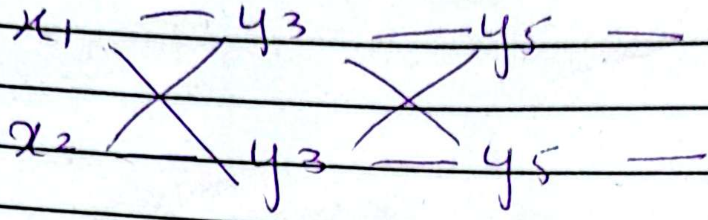


$$\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} w_{13} & w_{23} \\ w_{14} & w_{24} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



$y_3 = w_{13} x_1 + w_{23} x_2$

$y_4 = w_{14} x_1 + w_{24} x_2$



$$\begin{bmatrix} y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} W_{35} & W_{45} \\ W_{36} & W_{46} \end{bmatrix} \underset{2\times 2}{\begin{bmatrix} W_{13} & W_{23} \\ W_{14} & W_{24} \end{bmatrix}} \underset{2\times 1}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$\downarrow$$

$$\begin{bmatrix} w'_{1+2,\,3+2} & W_{2+2,\,3+2} \\ W_{1+2,\,4+2} & W_{2+2,\,4+2} \end{bmatrix}$$

$\cancel{W_{35}}$ W

$$= W_{35}\left[w_{13}x_1 + w_{23}x_2\right]$$

$$+ W_{45}\left[w_{14}x_1 + w_{24}x_2\right)$$

$$\begin{bmatrix} w_{13}x_1 & W_{23}x_2 \\ w_{14}x_1 & w_{24}x_2 \end{bmatrix}$$

Date

required				obtained	
x_1	x_2	y_3'	y_4'	y_3	y_4
0	0	0	0	0	0
0	1	0	1	0.5	0.5
1	0	0	1	0.5	0.5
1	1	1	0	1	1

$\frac{1}{4}[(0-0.5)^2 + (1-0.5)^2$

$E =$

$$y_3 = W_{13}x_1 + W_{23}x_2$$

$$y_4 = W_{14}x_1 + W_{24}x_2$$

$$W = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\boxed{E = \frac{1}{2}(y' - y_s)^2 \qquad E = \frac{1}{2}[(y_3' - y_3)^2 + (y_4' - y_4)^2]}$$

$y_3 \Rightarrow E = \frac{1}{4}[(0-0)^2 + (0-0.5)^2 + (0-0.5)^2 + (0-1)^2]$

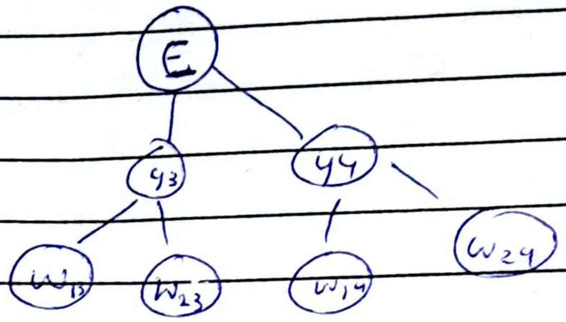
$y_4 \Rightarrow E = \frac{1}{4}[(0-0)^2 + (1-0.5)^2 + (1-0.5)^2 + (0-1)^2]$

$\frac{1}{8}$

$\frac{1}{5}($ y_3 $\quad \frac{0.25+0.25+0}{4} \quad$ $\frac{0+0.25+0.25+1}{4}$

$\frac{1}{8}(0 \cdots \quad 0.125 \quad 0.375$

Gradient Descent



$$\frac{\partial E}{\partial y_3} \qquad \frac{\partial E}{\partial y_4}$$

$$\frac{\partial y_3}{\partial w_{13}}, \quad \frac{\partial y_3}{\partial w_{23}} \qquad \frac{\partial y_4}{\partial w_{14}}, \quad \frac{\partial y_4}{\partial w_{24}}$$

$$\frac{\partial E}{\partial y_3} = \frac{\partial}{\partial y_3}\left(\frac{1}{2}\left((y_3' - y_3)^2 + (y_4' - y_4)^2\right)\right)$$

$$\frac{\partial E}{\partial y_3} = \frac{2}{2}(y_3' - y_3)(-1) \qquad \frac{\partial E}{\partial y_4} = (y_4' - y_4)(-1)$$

$$\boxed{\frac{\partial E}{\partial y_3} = -(y_3' - y_3) \qquad \frac{\partial E}{\partial y_4} = -(y_4' - y_4)}$$

$$\frac{\partial E}{\partial w_{13}} = \frac{\partial E}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{13}} \qquad \frac{\partial E}{\partial w_{14}} = \frac{\partial E}{\partial y_4} \cdot \frac{\partial y_4}{\partial w_{14}}$$

$$\frac{\partial E}{\partial w_{23}} = \frac{\partial E}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{23}} \qquad \frac{\partial E}{\partial w_{24}} = \frac{\partial E}{\partial y_4} \cdot \frac{\partial y_4}{\partial w_{24}}$$

$$\frac{\partial y^3}{\partial w_{13}} = \frac{\partial}{\partial w_{13}} \left(w_{13} x_1 + w_{23} x_2 \right)$$

$$= x_1$$

$$\frac{\partial y^3}{\partial w_{14}} = \frac{\partial}{\partial w_{14}} \left(w_{14} x_1 + w_{24} x_2 \right)$$

$$\frac{\partial y^3}{\partial w_{14}} = x_1$$

$$\frac{\partial y^3}{\partial w_{23}} = x_2 \qquad \frac{\partial y^3}{\partial w_{24}} = x_2$$

$$\frac{\partial E}{\partial w_{13}} = -(y^{3\prime} - y_3) \cdot x_1 \qquad \frac{\partial E}{\partial w_{14}} = -(y_4' - y_4) \cdot x_1$$

$$\frac{\partial E}{\partial w_{23}} = -(y_3' - y_3) \cdot x_2 \qquad \frac{\partial E}{\partial w_{24}} = -(y_4' - y_4) \cdot x$$

Date________ M T W T F S

	case (0,0)	case (0,1)	case (1,0)	case(1,1)
$\frac{\partial E}{\partial w_{13}}$	0	0	0.5	0
$\frac{\partial E}{\partial w_{23}}$	0	0.5	0	0
$\frac{\partial E}{\partial w_{14}}$	0	0	-0.5	1
$\frac{\partial E}{\partial w_{24}}$	0	-0.5	0	1

$\eta = 1$

$W_{new} = W_{old} + \eta \frac{\partial E}{\partial w_{old}}$

$$W = \begin{array}{cc} W_{13} & W_{23} \end{array} \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \begin{array}{cc} W_{14} & W_{24} \end{array}$$

For case (0,1) For case (1,0) For case (1,1)

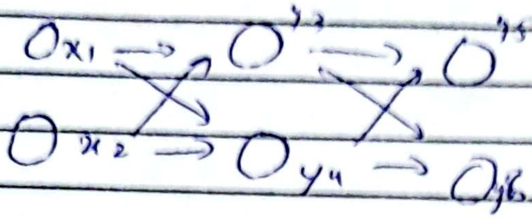
$$W_3 = \begin{bmatrix} 0.5 & 1 \\ 0.5 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1.0 & 1.0 \\ 0.5 & 0 \end{bmatrix} \rightarrow W = \begin{bmatrix} 1.0 & 1.0 \\ 1 & 1 \end{bmatrix}$$

$0.5 + 1(0.5)$

Date ______ M T W T F S

$$\begin{bmatrix} y_1^0 & y_1^1 & y_0^2 & y_1^3 \\ y_2^0 & y_2^1 & y_2^2 & y_2^3 \end{bmatrix} = \begin{bmatrix} w_{13} & w_{23} \\ w_{14} & w_{24} \end{bmatrix} \begin{bmatrix} x_1^0 & x_1^1 & x_1^2 & x_1^3 \\ x_2^0 & x_2^1 & x_2^2 & x_2^3 \end{bmatrix}$$

$$= [W] + \eta \begin{bmatrix} \frac{\partial E}{\partial w_{13}} & \frac{\partial E}{\partial w_{23}} \\ \frac{\partial E}{\partial w_{14}} & \frac{\partial E}{\partial w_{24}} \end{bmatrix}$$



$$\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} w_{13} & w_{23} \\ w_{14} & w_{24} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} W_{35} & W_{45} \\ W_{36} & W_{46} \end{bmatrix} \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}$$

$$y_5 = W_{35} y_3 + W_{45} y_4$$

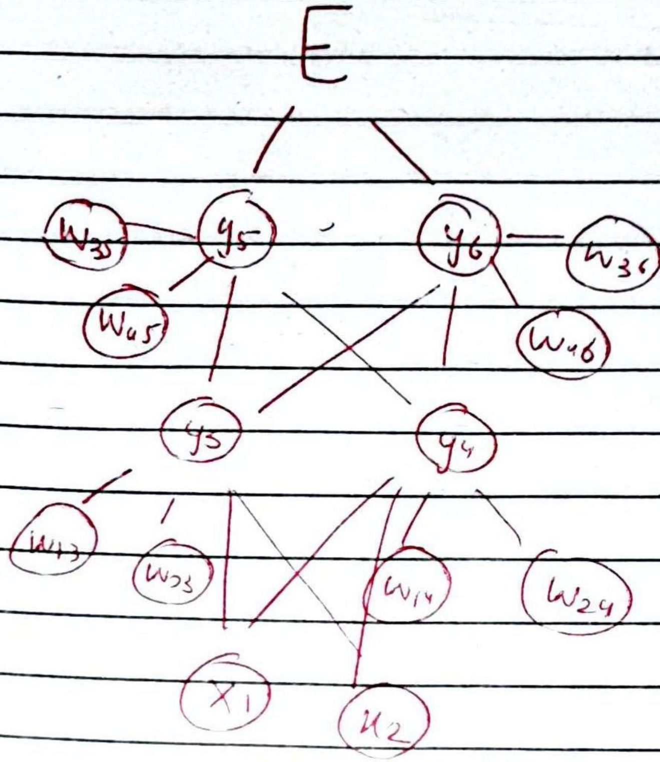
$$y_6 = W_{36} y_3 + W_{46} y_4$$

$$y_5 = W_{35} (W_{13} x_1 + W_{23} x_2) + W_{45} (W_{14} x_1 + W_{24} x_2)$$

$$y_6 = W_{36} (W_{13} x_1 + W_{23} x_2) + W_{46} (W_{14} x_1 + W_{24} x_2)$$

Date____ M T W T F S

$E = \frac{1}{2}\left((y_5' - y_5)^2 + (y_6' - y_6)^2\right)$



$\frac{\partial E}{\partial y_5}, \frac{\partial E}{\partial y_6}, \frac{\partial y_5}{\partial w_{35}}, \frac{\partial y_5}{\partial w_{45}}, \frac{\partial y_6}{\partial w_{36}}, \frac{\partial y_6}{\partial w_{46}}$

$\frac{\partial y_5}{\partial y'}, \frac{\partial y_6}{\partial y_3}, \frac{\partial y_5}{\partial y_3}, \frac{\partial y_6}{\partial y_4}, \frac{\partial y_3}{\partial w_{13}}, \frac{\partial y_3}{\partial w_{23}}, \frac{\partial y_4}{\partial w_{14}}, \frac{\partial y_4}{\partial w_{24}}$

$\cancel{\frac{\partial E}{\partial y_5} = \frac{\partial y_5}{\partial w_{35}} \cdot \partial y_5 \cdot \frac{\partial y_5}{\partial y_3} \cdot \frac{\partial y_3}{\partial y_4} \cdot \frac{\partial y}{\partial w_{13}}}$

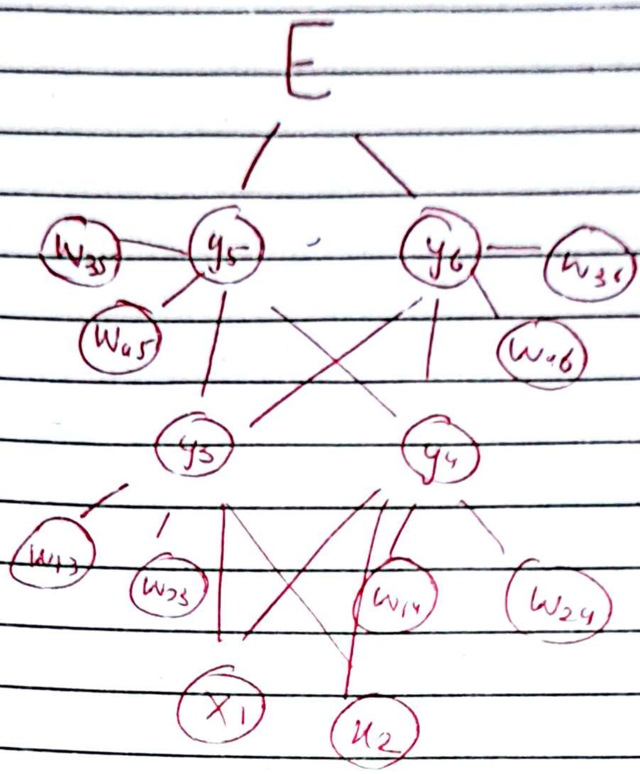
$\frac{\partial E}{\partial y_5} = -(y_5' - y_5)$ $\quad \frac{\partial E}{\partial y_6} = -(y_6' - y_6)$ $\quad \frac{\partial y_5}{\partial w_{35}} = y_3$ $\quad \frac{\partial y_5}{\partial w_{45}} = y_4$

$\frac{\partial y_6}{\partial w_{36}} = y_3$ $\quad \frac{\partial y_6}{\partial w_{46}} = y_4$ $\quad \frac{\partial y_5}{\partial y_3} = w_{35}$ $\quad \frac{\partial y_5}{\partial y_4} = w_{45}$

$\frac{\partial y_6}{\partial y_3} = w_{36}$ $\quad \frac{\partial y_6}{\partial y_4} = w_{46}$ $\quad \frac{\partial y_3}{\partial w_{13}} = x_1,$ $\frac{\partial y_3}{\partial w_{23}} = x_2,$ $\frac{\partial y_4}{\partial w_{14}} = x_1$ $\quad \frac{\partial y_4}{\partial w_{24}} = x_2$

Date______ M T W T F S

$E = \frac{1}{2}\left[(y_5' - y_5)^2 + (y_6' - y_6)^2\right]$



$\frac{\partial E}{\partial y_5}, \frac{\partial E}{\partial y_6}, \frac{\partial y_5}{\partial w_{35}}, \frac{\partial y_5}{\partial w_{45}}, \frac{\partial y_6}{\partial w_{36}}, \frac{\partial y_6}{\partial w_{46}}$

$\frac{\partial y_5}{\partial y_4}, \frac{\partial y_6}{\partial y_3}, \frac{\partial y_5}{\partial y_3}, \frac{\partial y_6}{\partial y_4}, \frac{\partial y_3}{\partial w_{13}}, \frac{\partial y_3}{\partial w_{23}}, \frac{\partial y_4}{\partial w_{14}}, \frac{\partial y_4}{\partial w_{24}}$

$\frac{\partial E}{\partial y_5} = \cancel{\frac{\partial y_5}{\partial w_{35}}} \cdot \cancel{\frac{\partial y_6}{\partial y_3}} \cdot \cancel{\frac{\partial y_5}{\partial y_4}} \cdot \frac{\partial y_3}{\partial w_{13}} \cdot \frac{\partial y}{\partial}$

$\frac{\partial E}{\partial y_5} = -(y_5' - y_5)$ $\frac{\partial E}{\partial y_6} = -(y_6' - y_6)$ $\frac{\partial y_5}{\partial w_{35}} = y_3$ $\frac{\partial y_5}{\partial w_{45}} = y_4$

$\frac{\partial y_6}{\partial w_{36}} = y_3$ $\frac{\partial y_6}{\partial w_{46}} = y_4$ $\frac{\partial y_5}{\partial y_3} = w_{35}$ $\frac{\partial y_5}{\partial y_4} = w_{45}$

$\frac{\partial y_6}{\partial y_3} = w_{36}$ $\frac{\partial y_6}{\partial y_4} = w_{46}$ $\frac{\partial y_3}{\partial w_{13}} = x_1,$ $\frac{\partial y_3}{\partial w_{23}} = x_2,$ $\frac{\partial y_4}{\partial w_{14}} = x_1$ $\frac{\partial y_4}{\partial w_{24}} = x_2$

Date________ M T W T F S

$$\cancel{\frac{\partial E}{\partial w_{35}} = \frac{\partial y_5}{\partial w_{35}} \partial}$$

$$\frac{\partial E}{\partial w_{35}} = \frac{\partial E}{\partial y_5} \cdot \frac{\partial y_5}{\partial w_{35}}$$

$$\frac{\partial E}{\partial w_{45}} = \frac{\partial E}{\partial y_5} \cdot \frac{\partial y_5}{\partial w_{45}}$$

$$\frac{\partial E}{\partial w_{36}} = \frac{\partial E}{\partial y_6} \cdot \frac{\partial y_6}{\partial w_{36}}$$

$$\frac{\partial E}{\partial w_{46}} = \frac{\partial E}{\partial y_6} \cdot \frac{\partial y_6}{\partial w_{46}}$$

$$\frac{\partial E}{\partial w_{13}} = \frac{\partial E}{\partial y_5} \cdot \frac{\partial y_5}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{13}} +$$

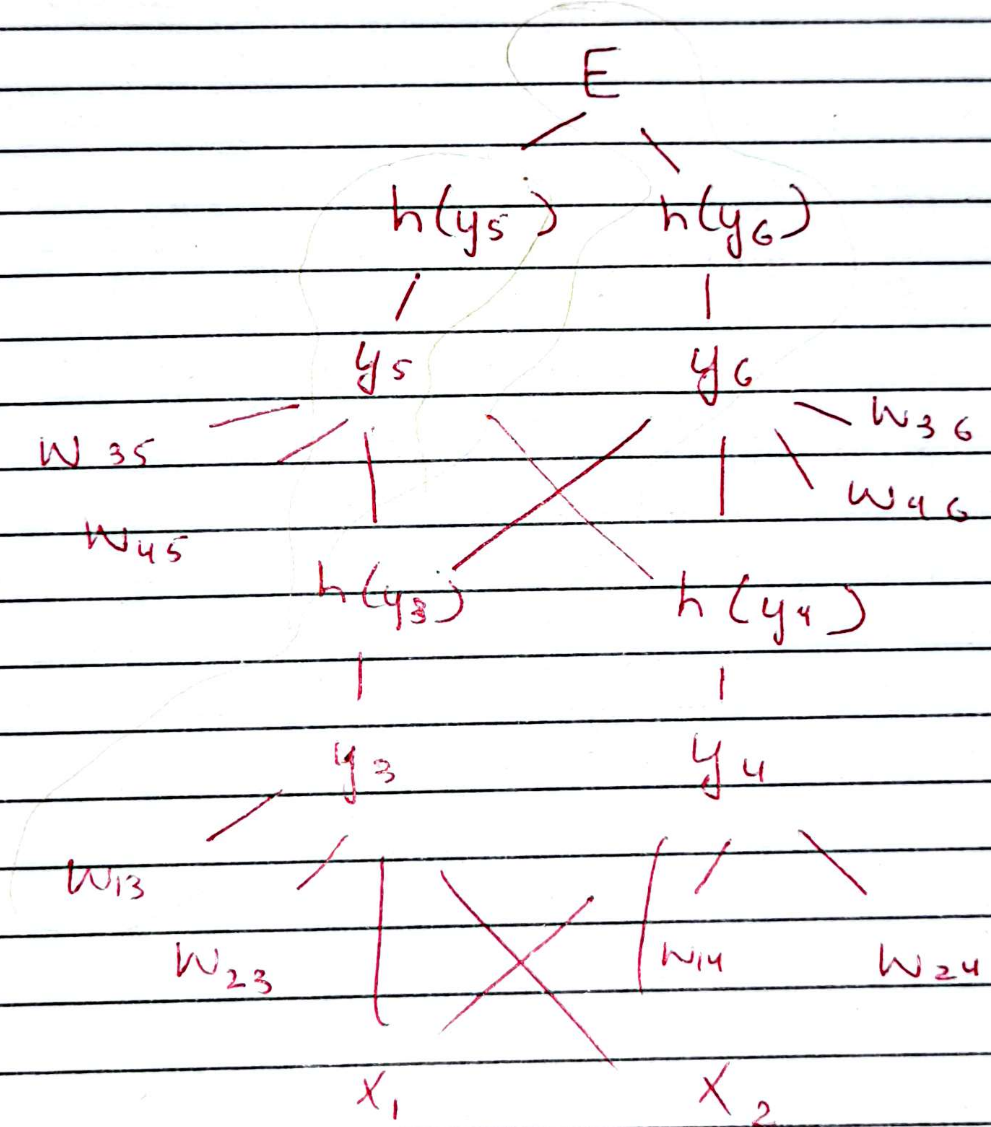
$$\frac{\partial E}{\partial y_6} \cdot \frac{\partial y_6}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{13}}$$

We ignore this approach (tree traversal make calculation exact)

Date______ M T W T F S

$$h(y) = \frac{1}{1+e^{-y}}$$

$$E = \frac{1}{2}\left(h(y)' - h(y)\right)^2$$



Date________ M T W T F S

$$\frac{\partial E}{\partial h(y_5)} = -(h(y_5)' - h(y_5))$$

$$\frac{\partial E}{\partial h(y_6)} = -(h(y_6)' - h(y_6))$$

~~$\partial(h(y$~~

$$\frac{\partial h(y_5)}{y_5} = \frac{e^{-y_5}}{(1+e^{-y_5})^2}$$

$$\frac{\partial h(y_6)}{y_6} = \frac{e^{-y_6}}{(1+e^{-y_6})^2}$$

$y_5 = w_{35} h(y_3) +$
$w_{45} h(y_4)$

$$h(y_k) = \frac{\partial}{\partial y_k} \frac{1}{1+e^{-y_k}}$$

$$= (1+e^{-y_k})^{-1}$$

$$= -1\,(1+e^{-y_k})^{-2}(e^{-y_k})(-1)$$

$$= \frac{(e^{-y_k})}{(1+e^{-y_k})^2}$$

$$= \frac{e^{-y_k}}{(1+e^{-y_k})} \cdot \frac{1}{(1+e^{-y_k})}$$

$\cancel{[1-h(y_k))}$

$[1-y_k]\,[y_k)]$

Subject:

Sr. No.	Date	Topics

$$\frac{\partial E}{\partial w_{35}} = \frac{\partial E}{\partial h(y_5)} \cdot \frac{\partial h(y_5)}{\partial y_5} \cdot \frac{\partial y_5}{\partial w_{35}}$$

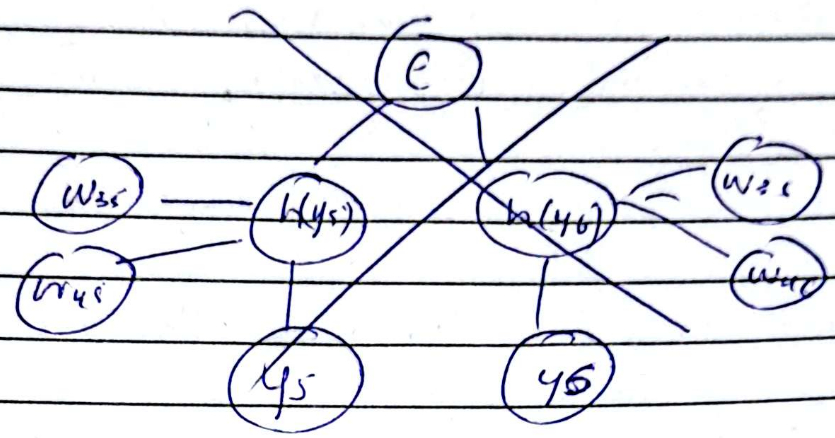
$$\frac{\partial E}{\partial w_{13}} = \frac{\partial E}{\partial h(y_5)} \cdot \frac{\partial h(y_5)}{\partial y_5} \cdot \frac{\partial y_5}{\partial h(y_3)} \cdot \frac{\partial h(y_3)}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{13}} + \frac{\partial E}{\partial h(y_6)} \cdot \frac{\partial h(y_6)}{\partial y_6} \cdot \frac{\partial y_6}{\partial h(y_3)} \cdot \frac{\partial h(y_3)}{\partial y_3} \cdot \frac{\partial y_3}{\partial w_{13}}$$

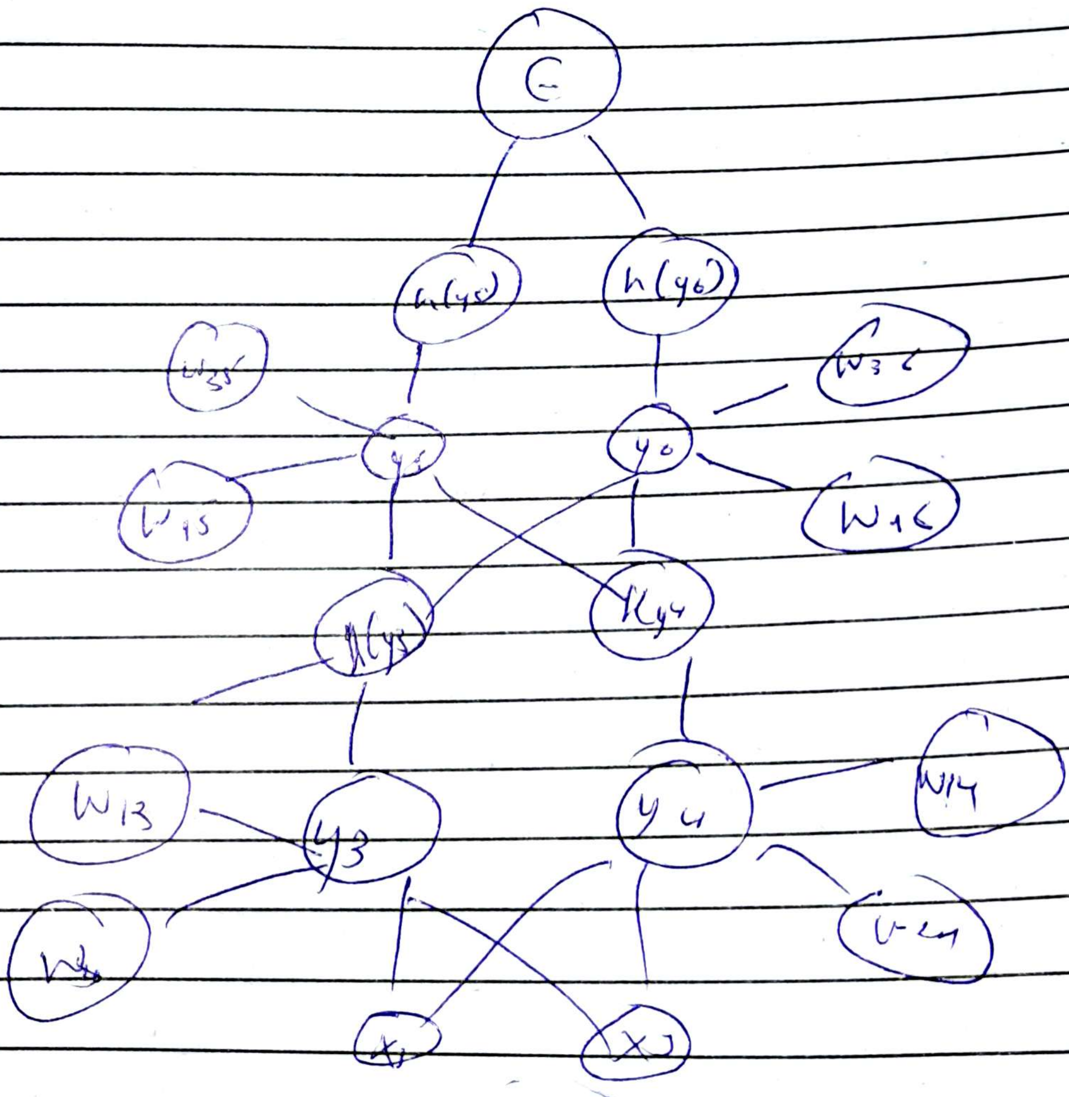
$$= -\cancel{(h(y_5)' - h(y_5))} \cdot \frac{e^{-y_5}}{(1+e^{-y_5})^2}$$

$$= -(h(y_5)' - h(y_5))(y_5)(1-y_5) \cdot w_{35}$$

Date

M T W T F S





Date________ M T W T F S

CNN

$x_0, \ldots\ldots x_7$

w_0, w_1, w_2

Convolve

$$\text{convolve}(x, w) = \begin{bmatrix} x_0 w_0 + x_1 w_1 + x_2 w_2 \\ x_1 w_0 + x_2 w_1 + x_3 w_2 \\ x_2 w_0 + x_3 w_1 + x_4 w_2 \\ + \ldots\ldots\ldots\ldots \\ x_5 w_0 + x_6 w_1 + x_7 w_2 \end{bmatrix}$$

Use padding to give importance to corners

x

$2p$ zeros add on both side

p^l left side

p^r right side

Formula to calculate new size after padding and convolve operation

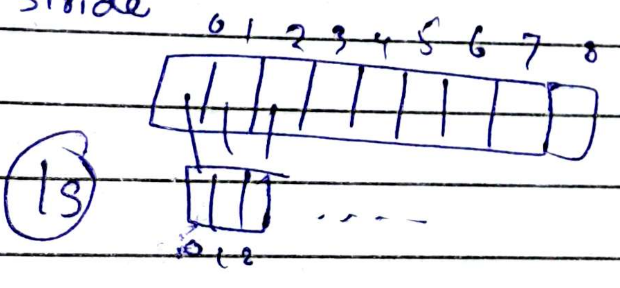
original matrix, kernel matrix

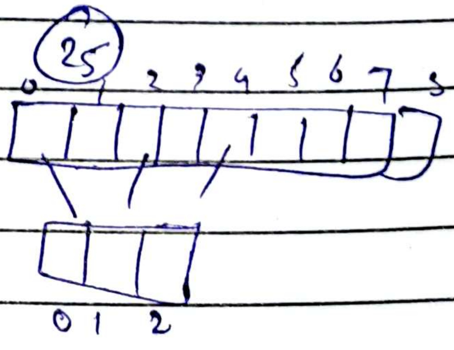
$$\text{size}\left(\frac{\text{size}(X) - \text{size}(K)}{s} + 2p\right) + 1$$

s — stride

padding

stride





Date ______ M T W T F S

Convolution. $I^{n\times n} * k^{m\times n}$

Without padding and stride

$S_I - S_k + 1$

With padding and stride

$\left[\frac{S_I - S_k + 2p}{s}\right] + 1$

different kernels and matrix mult examples

Kernel matrix → just like operator

$x_i - x_{i+1}$

$x_i \; [0 \quad +1]$

tells position

$x_i + x_{i+1}$

$x_i \; [0 \quad +1]$

Required

Output matrix → $O^{256\times 256}$

$I^{1600\times 920} \quad K^{3\times 3} \quad O^{256\times 256}$

-Calcu

Date________ M T W T F S

$$\frac{1600-3}{2}+1$$

$$\frac{920-3}{2}+1$$

adjust padding and stride to find an even number for equally divisible.

$$\left[\frac{1600-3+2p}{s}\right]+1=600$$

$$\frac{1600-3+2p}{s}=600-1$$

$$\frac{1600-3+2p}{s}=599$$

$$\frac{1597+2p}{s}=599$$

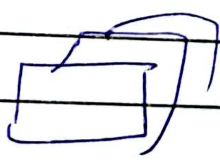
$$1597+2p=599s$$

$$2p=599s-1597$$

$$p=\frac{599s}{2}-\frac{1597}{2}$$

$$\left(\frac{S_I-S_K+2p}{S}\right)+1$$

$$\left(\frac{16-3+2(1)}{2}\right)+1$$



$$H_K \times W_K \times C_K \times N_K$$

if we want to adjust it according
leNet / AlexNet

which take input Image as 32x32x1

1600 x 920

3

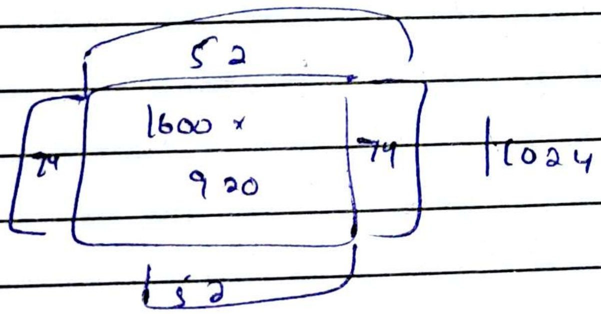
$32 \times 32 \times 50 \times$

$\frac{1600}{32} \times \frac{920}{32}$

3 28.75

50 28.75

or if we add padding



2048

: 1600 x 920 → 32 x 32 x 2048

1600 x 920 resizing

$\frac{1600}{32}$, $\frac{920}{32}$

50 28.5

$(32 \times 32 \times 50 \times 28.75)$

Date ______ M T W T F S

Ker

$$H_k \times W_k \times C_k \quad \times N_k$$

$$H_\circ = \left[\frac{H - H_k + 2p}{S} \right] + 1$$

$$W_\circ = \left[\frac{W - W_k + 2p}{S} \right] + 1$$

$\boxed{3 \times 3}$ ⊗ $\boxed{3 \times 3}$

dot product

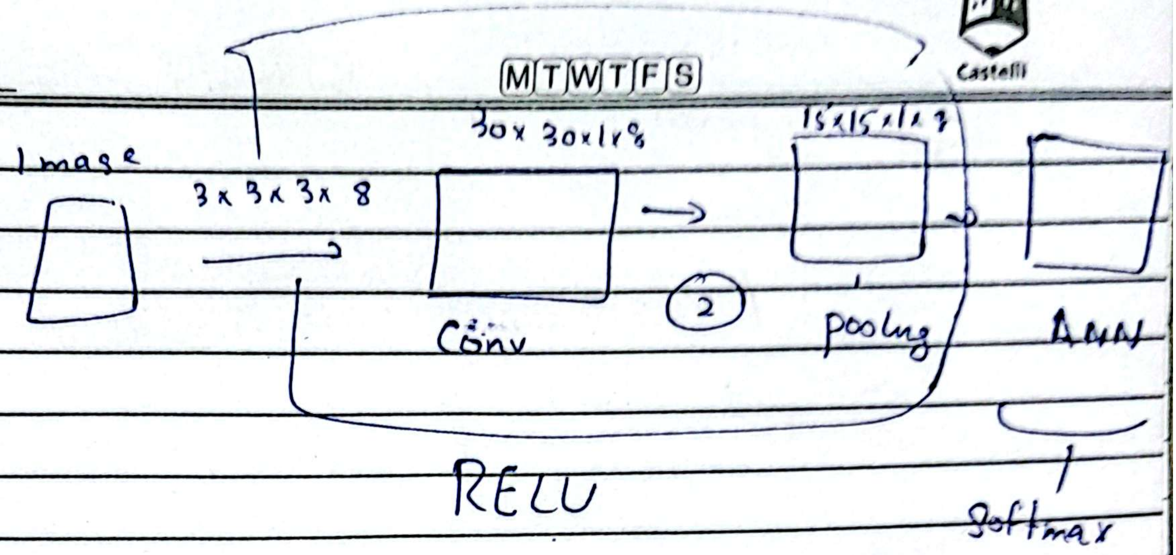
One value

$3 \times 3 \times 3 \times 8$

$$\frac{32 - 3 \times 8}{1} + 1 = 30$$

$$\left[\frac{1600 - 3 + 2p}{S} \right] + 1 = 600$$

1



back propagation

$$\frac{\partial E}{\partial k_1} + \frac{\partial E}{\partial k_{11}} + \frac{\partial E}{k_{13}} + \cdots \frac{\partial E}{\partial k_{19}}$$

Pooling
- Max
- Avg
- Min

$15 \times 15 \times 1 \times 8 \rightarrow 10 \times 10 \times 4$

$$10 = \left[\frac{15 - H_k + 2(0)}{1}\right] + 1 \quad = \quad 6 \times 6$$

$6 \times 6 \times 1 \times 8$ pooling

$6 \times 6 \times 1 \times 8 / 2$ Pooling on last column

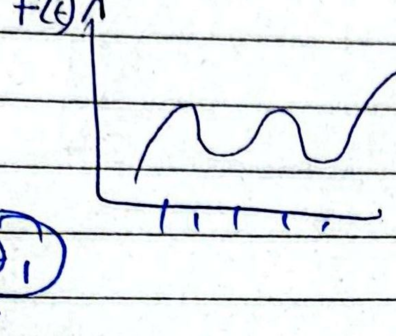
$6 \times 6 \times 1 \times 4$

Day:________ Date:________

Time Series Plots

ECG signals

finance

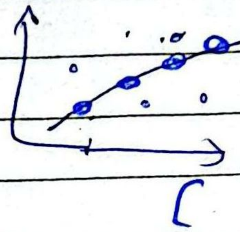


measurement

work

θ_1

GASF



$t = \ldots$ ASS

$i, j = \cos(\theta_i + \theta_j)$

$f(t) = \tanh(f(x))$

$\theta = \cos^{-1}(f(t))$

GADF transformation → Image

$i, j = \sin(\theta_i + \theta_j)$

$i, j = \cos(\theta_i + \theta_j)$

Stin

ANN

2 techniques

(GASF)

- Grammian Angular Summation field

(GADF)

" " Difference field

~~f(t)~~ $f(t) = \tanh(f(t))$

~~Angle~~ → cos

$\theta = \cos^{-1}(f(t))$

Date:

① Normalized Form $(-1, +1)$ tanh

$G(t) = \tanh(f(t))$

② Angle $(0, 2\pi)$

$\theta = \cos^{-1}(G(t))$

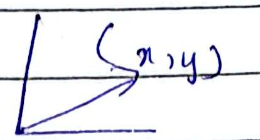
polar coordinates are superior

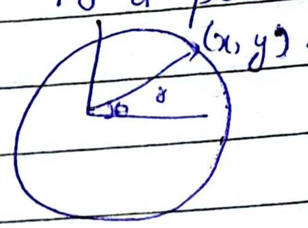
accuracy in case

noise free data

Transformation

Cartesian space to a polar space

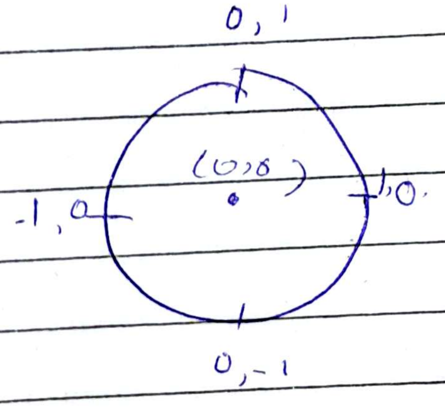




$(x, y) \to (r, \theta)$

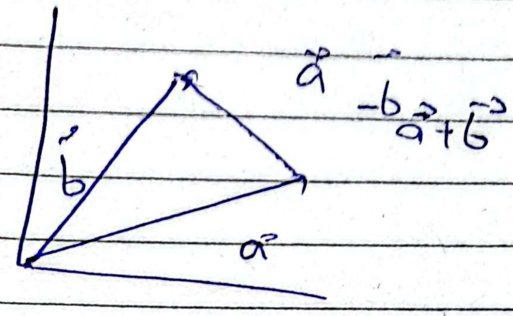
Polar coordinates

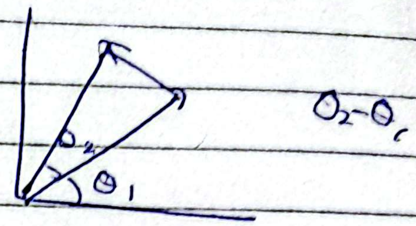
---

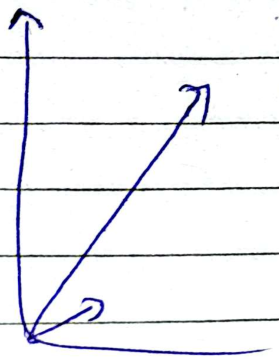
background	
	$x = \cos(\theta)$ $y = -\sin(\theta)$ $t = \cos(\theta)$ $f(t) = -\sin\theta$

Day:

Date:

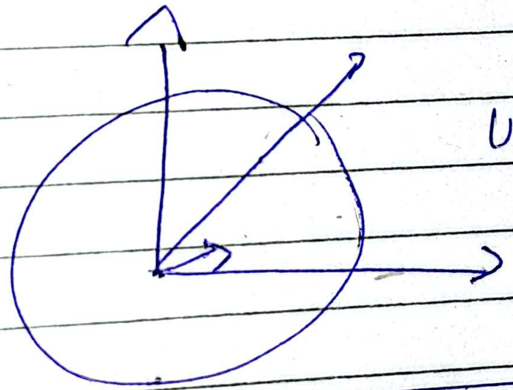






angle difference less

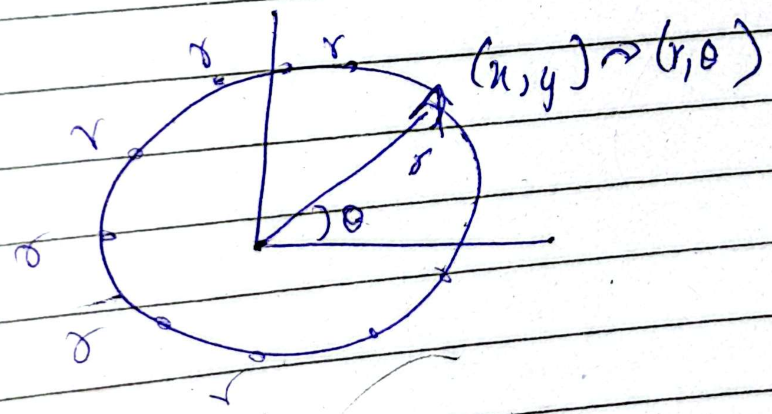
r diff difference more



In

Unit circle

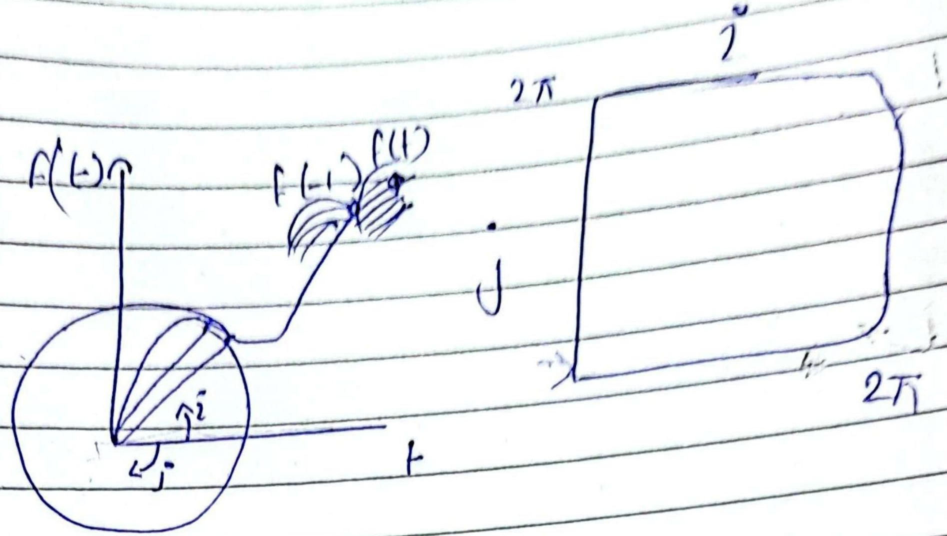
representation



Day:

Date:

$f(t) = f(t-1) +$



(3) Construct Field (GASF)

for loop → $(i, j) = \cos(\theta_i + \theta_j)$ base level & pattern

global

GADF

$(i, j) = \sin(\theta_i - \theta_j)$ state level & patterns

minutes

GASF | → CNN → ANN

Image → F.F. | → CNN → ANN

Spectral Convolution

Positive effect on accuracy

Day:

Date:

obtained required

O O—H

O O—H

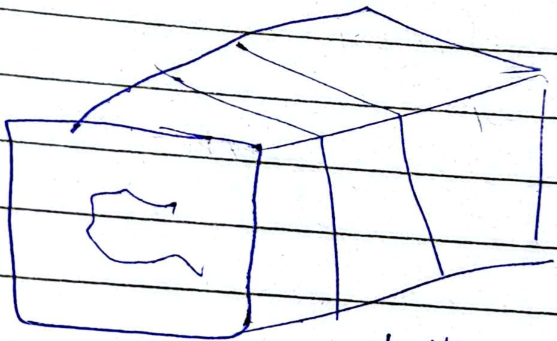
O O—H

a+a

a+b

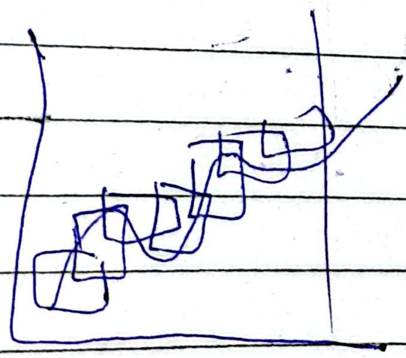
bearish

bullish

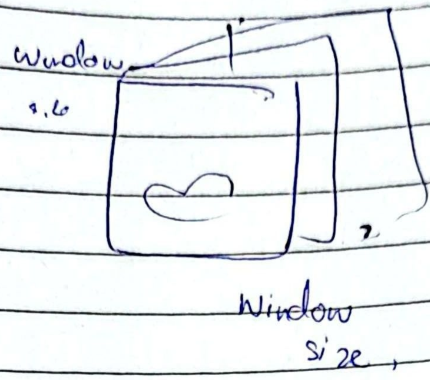


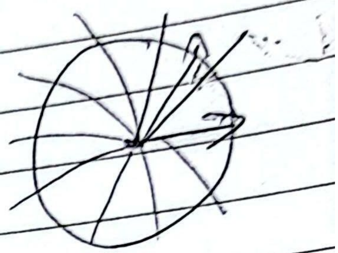
bullish

bearish



Day: ______ Date: ______





① data $f(x)$

$\downarrow$

② $f(t) = \tanh(f(x))$

$[a, b, c, d,$

$f(t$

$\theta = \cos^{-1}(f(t))$

θ_1, θ_2

$G_{i,j} = \cos(\theta_i + \theta_j)$

Date GASF

GADF

(1) Normalized form $[-1, 1]$

(2) $G(t) = \tanh(f(t))$

(3) Angle $(0, 2\pi)$

$\theta = \cos^{-1}(G(t))$

(4) Construct field (GASF)

for loop

$(i, j) = \cos(\theta_i + \theta_j)$

$\sum_{0} \sum^{i} (0, i)$

$-3 \times -1 \times 3$

$3 \times 3 \times 1$

$-3 \times 1 \times -3$

$= 0 + 9 + 9 + 9$

$= 27$

Day:________ Date:________

~~RNN~~

Hopfield Neural Networks

2024 Nobel Prize Physics → AI

Unsupervised

Energy <=> Weights

(Ising Model) -1 $\boxed{\uparrow \mid \uparrow}$ spineffect magnetic spin

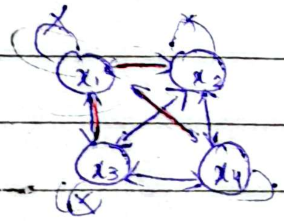
both energy are in same direction
positive effect
no need to spin

$+1$ $\boxed{\uparrow \mid \downarrow}$ → need energy to align them

Energy represented as -1 and $+1$

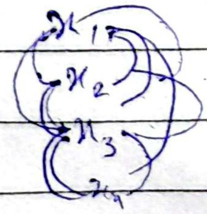
1. 1 layer
2. ~~Ac~~ Sign(~) $\{-1, +1\}$ - bipolar
   $\{0, 1\}$ - binary

$$g(z) = \begin{cases} +1 & z \geq 0 \\ -1 & z < 0 \end{cases}$$

2x2 Image → 

↳ $W^{4 \times 4}$

$$\begin{bmatrix} w_{11} & & & \\ & w_{22} & & \\ & & w_{33} & \\ & & & w_{44} \end{bmatrix}$$

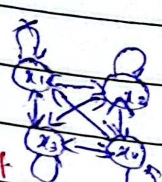


Day: images add Outer product Date:

(3)

$$W = \begin{cases} x_i^k (x_j^k) & i \neq j \\ 0 & i = j \end{cases}$$

$$= \begin{cases} x_i^k (x_i^k)^T - \\ \text{Diagonal} \end{cases}$$



$x_1 \times x_2$

$x_2 \times x_1$

so

Input $w^{4\times4}$

$$\begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \times \begin{bmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{bmatrix}$$

$i, j = -1 \times 1 = -1$

$\Sigma W_{ij} X_{ij}$

this could be an image

neural network one directional

$W = [x_i^k \; x_j^k \quad i =$

$x_i^k (x_i)^T$

Here bidirectional

(4) Stabilitiy Test

$S = 8 ((\Sigma W_{ij} X_i) + \beta I)$

$S = [-3 \; 3 \; 3 \; -3]$

$$\begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \times \begin{bmatrix} 0 & -1 & -1 & 1 \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

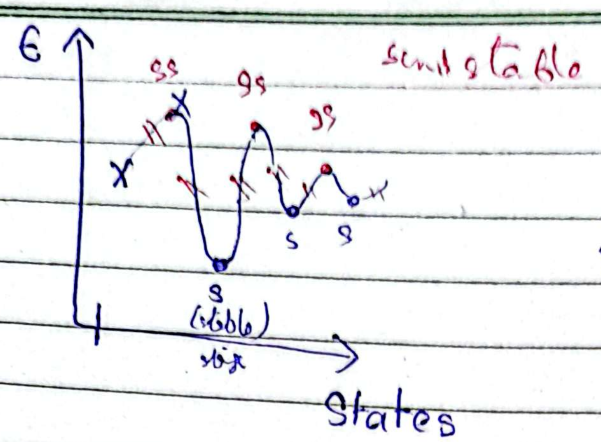
(5) Energy

$-1 \; -1 \; -1$

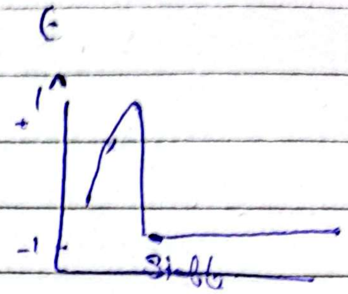
$$E = \sum_{i,j} W_{ij} S_i S_j$$

$$= \sum_i \sum_j W_{ij} S_i S_j$$ Ising equation

Day:________ Date:________



Want



Energy is increasing in a system, more distortion in the system (between the images)

> Min amount of variation

Min amount will always at stable points

– local minima

$$E = \sum_i \sum_j W_{ij} S_i S_j$$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \end{bmatrix}$$

$i = 0$

$j = 0, \boxed{1}, 2, 3$

$= 0 + (-1) \times (-3) \times (3) + \dots$

$= 0 + 9 + 9 + 9$

$= 27$

+1 Answer

Some case positive and negative

Day:

Date:

$i = 1$
$j = 0, 1, 2, 3$

$9 + 0 + 9 + \dots$

$$\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \quad W \begin{bmatrix} 0 & 1 & -1 & 1 \\ 1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{bmatrix}$$

3

add the weight of both images

$$\begin{bmatrix} 0 & 0 & -2 & 2 \\ 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & -2 \\ 2 & 0 & -2 & 0 \end{bmatrix}$$

getting small

15 % ~~train~~ train data

High energy

total

Stability

$S = [4 \ 0 \ -4 \ 4]$
$\quad\quad\quad i \quad\quad j$

$0 + 0 + 32 + 32$

High acc low energy

Energy Based Machine learning

Quadratic unconstrained binary

We want to find

$$\boxed{\min \left(E = \sum_i \sum_j W_{ij} S_i S_j \right)}$$

$-8 + -8$

QUBO

Quantum

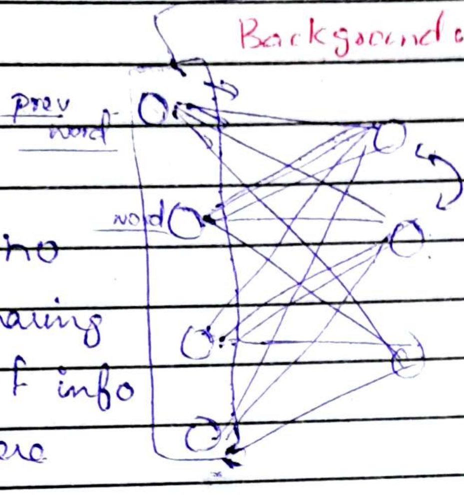
$(-2)(4)(-4) + (2)(4)(4)$

Date________ MTWTFS

RNN

Backgrounds

back propagation → (dust learning)



$f(\sum \omega x + b)$

want that relation

→ Input → output (1 directional)

no sharing of info here

Image → pixels not related

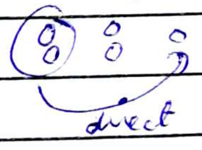
but text are

Image	$P_1 \leftrightarrow P_2$ → ANN/CNN → good result
text/ Speech, stock market	$W_1 \leftrightarrow W_2$ relation → not capture relation

Recurrent → recall → previous

Two approaches
- Inter
- beyond

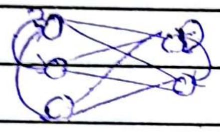
(technical reason
avoid vanishing gradient
obtain some of info)



attention take something from a very older layer and take that info relate to a farther layer

Date ______ MTWTFS

hope field



connected
Everybody with everybody

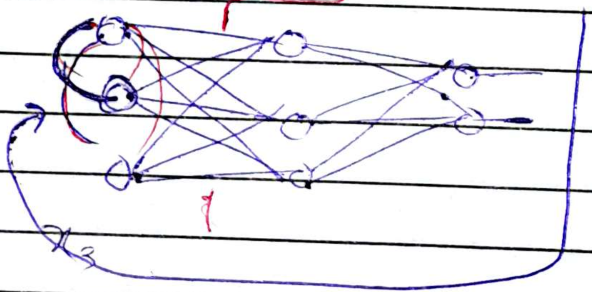
RNN (immodiate neighbour)

$\sigma(Wx + b)$

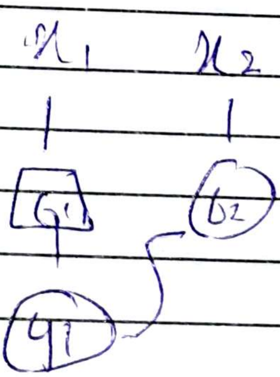
~~(bidirection)~~ RNN

$\cancel{\sigma(W_x x + W_h}$

$\sigma(Wx + U x_{t-1} + b)$



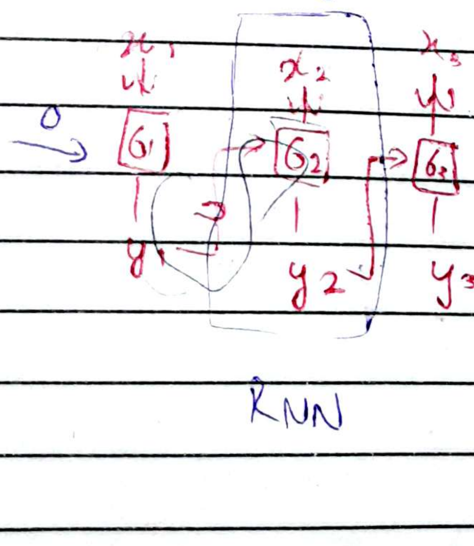
→ (Just select the current previous element/neighbour)



$W_x + U_{x_{t-1}} + b$

$x \rightarrow \boxed{\sigma} \rightarrow y$

$W_x + U^{\beta}_{x_{t-1}} + U_{x_{t-1}} + b$



→ Dependency

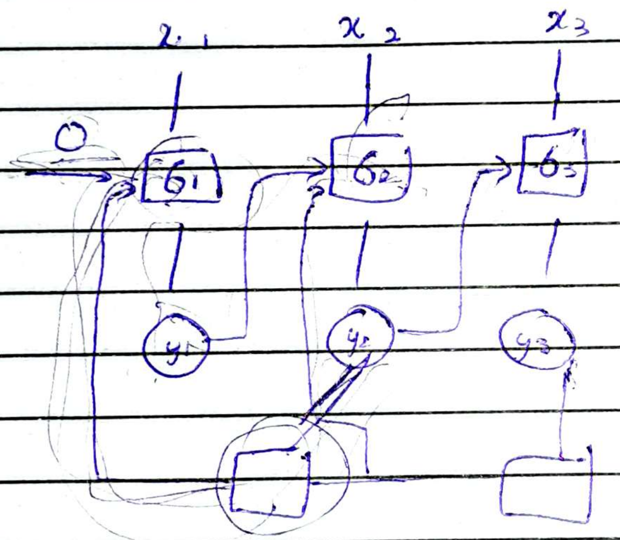
↳ slow

RNN

Date______ M T W T F S

bidirectional RNN

$$\sigma \left(Wx + U^{B}{}_{x_{t-1}} + U^{f}{}_{x_{t+1}} + b \right)$$



The cat ate a mouse

The ← cat → ate

previous in future

verb

S-O relation

O-S relation

(positioning depend on previous and future ~~of~~ objects)

bidirectional adding more context.