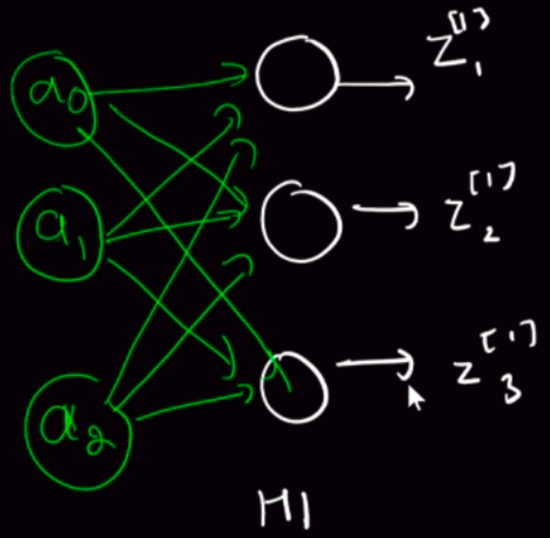
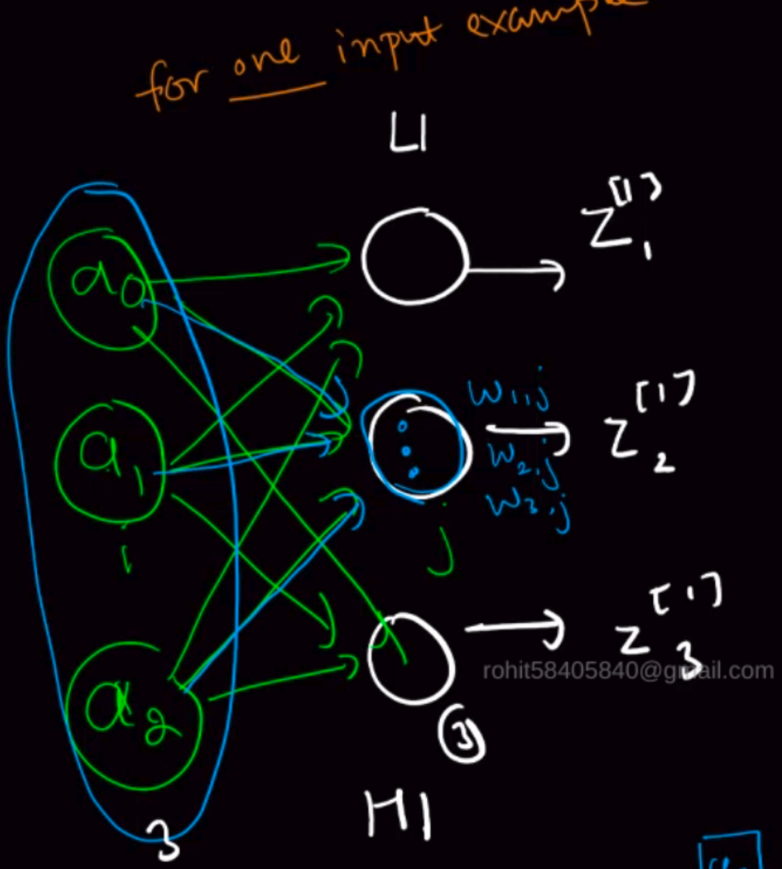
$$Z = \sum w_i x_i + b$$



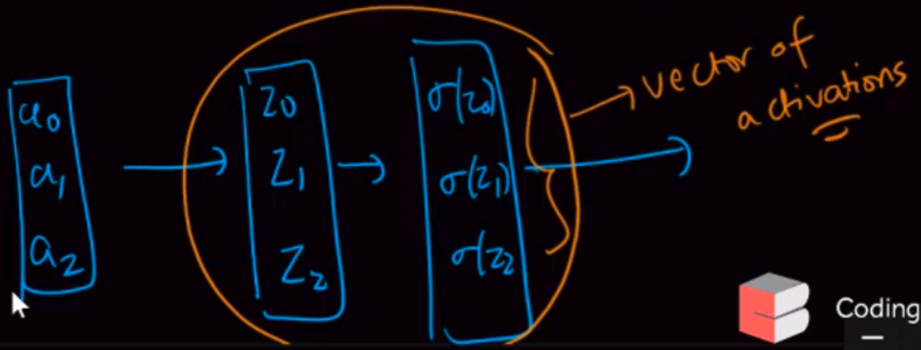
$$Z_j^{[l]}$$

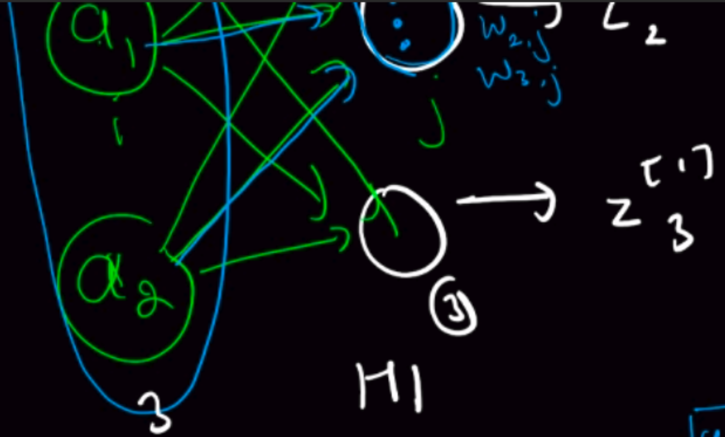
$$Z_j^{[l]} = \sum_i W_{ij}^{l} x_i^{l-1} + b$$



$$W^{l} = \begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \end{bmatrix}_{3 \times 3}$$

$$\boxed{z_j^{[l]} = \sum_i w_{i,j}^{[l]} a_i^{[l-1]} + b}$$





$$\boxed{z_j^{[l]} = \sum_i w_{i,j}^{[l]} a_i^{[l-1]} + b}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \rightarrow \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} \rightarrow \begin{bmatrix} \sigma(z_0) \\ \sigma(z_1) \\ \sigma(z_2) \end{bmatrix} \rightarrow$$

vector of activations

vector Notation?

vector Notation?

$$z^{[l]}$$

$$z^{[l]} = [w^l]^T \boxed{a^{l-1}} + b^l$$

$(l,(l-1)) \times (l-1,1)$

$(w^l)^T \rightarrow (l, l-1)$

vector
Notation?

$$\Downarrow$$

$$z^{[l]}$$

$$z^{[l]} = [w^{l}]^{T} \boxed{a^{l-1}} + b^{l}$$

$$\cancel{(l,(l-1)) \times (l-1,1)}$$

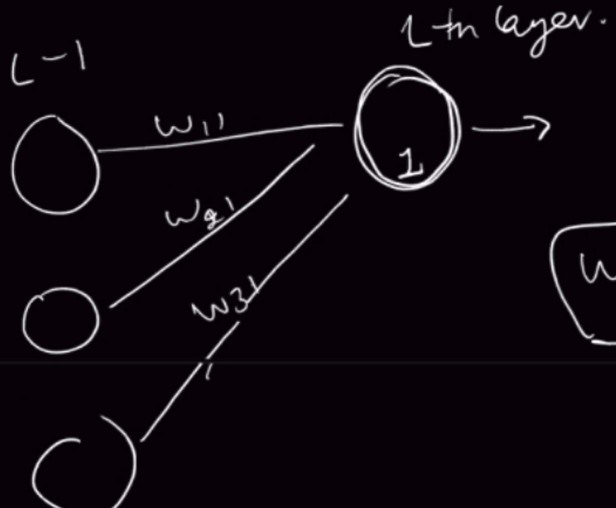
$$(l,1)$$

$$(l,1)$$

$$(w^{l})^{T} \to (l,l-1)$$

: l outputs

$$(W^L)^T = \begin{bmatrix} w^L_{11} & w^L_{21} & w_{31} \\ w_{12} & w_{22} & w_{32} \\ w_{13} & w_{23} & w_{33} \end{bmatrix} \begin{bmatrix} a_1^{l-1} \\ a_2^{l-1} \\ a_3^{l-1} \end{bmatrix} = \begin{bmatrix} z_1^{[l]} \\ z_2^{[l]} \\ z_3^{[l]} \end{bmatrix} = Z^{[l]}$$

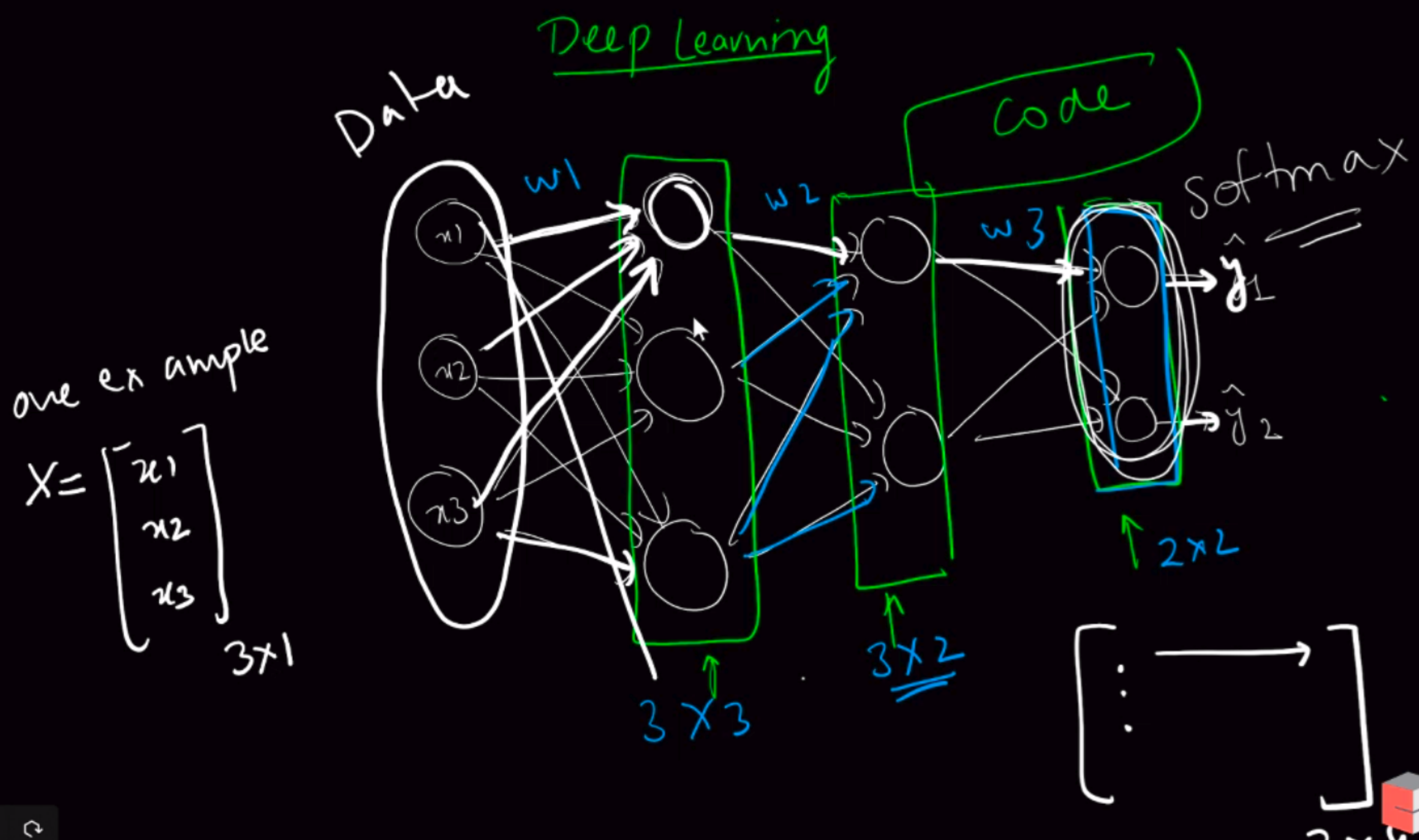


w_{ij}

$$\boxed{w_{11}a_1 + w_{21}a_2 + w_{31}a_3}$$

$$w_{11}a_1 + w_{21}a_2 + w_{31}a_3$$

$$g(z^{[l]}) = \sigma(z^{[l]}) = \begin{bmatrix} \sigma(z_1^{[l]}) \\ \sigma(z_2^{[l]}) \\ \sigma(z_3^{[l]}) \end{bmatrix} \Leftarrow \text{Output after hidden layer}$$


Deep Learning
Data
code
softmax
w1
w2
w3
x1
x2
x3
$\hat{y}_1$
$\hat{y}_2$
one example
X= [x1 x2 x3]
3x1
3 X 3
3x2
2x2

[10 , 10 , 20]

Numbers

Z_3

0.1 0.1 0.8

Softmax

Prob. distribution

$$Z_2 = (W^{[2]})^T a^{[1]} + b^{[2]}$$

$$a_2 = \sigma(Z_2)$$

$$Z_3 = (W^3)^T . a^{[2]} + b^{[3]}$$

$$\hat{y} = \text{softmax}(Z_3)$$