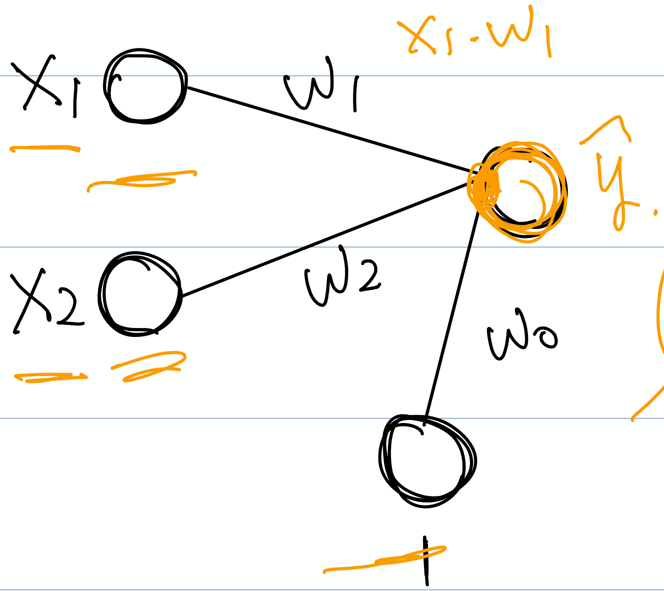
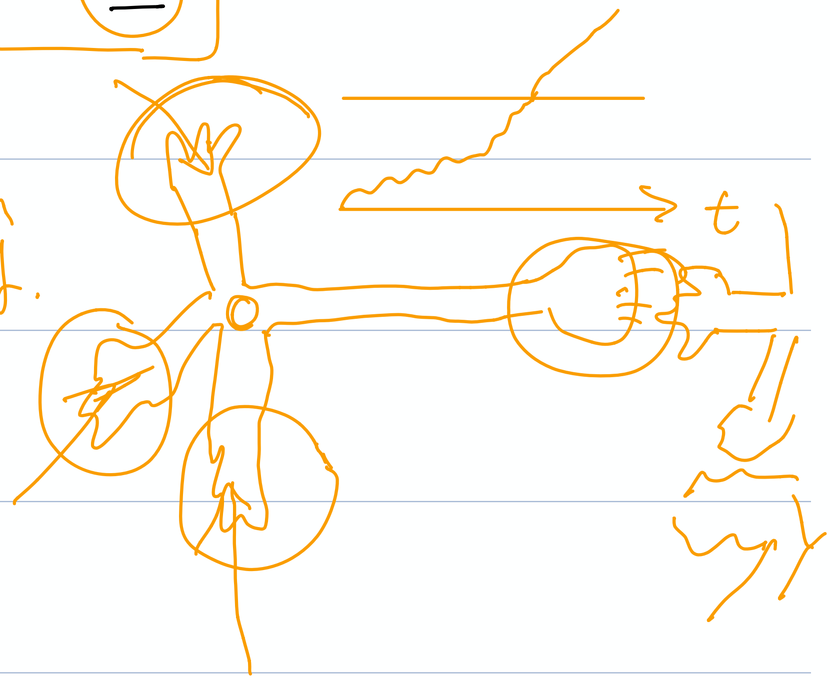
linear regression $\hat{y} = \boxed{w_0 \cdot 1 + w_1 x_1 + w_2 x_2}$



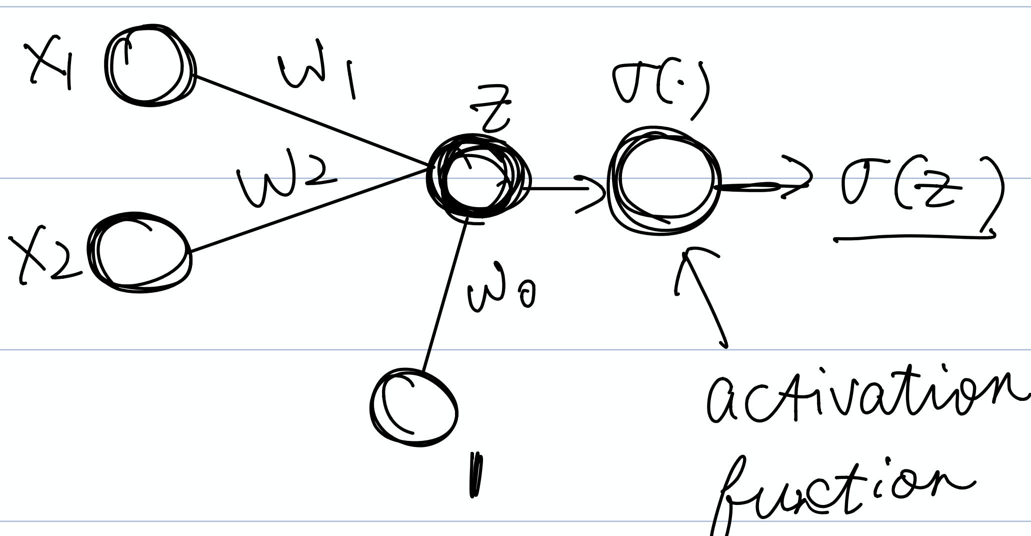


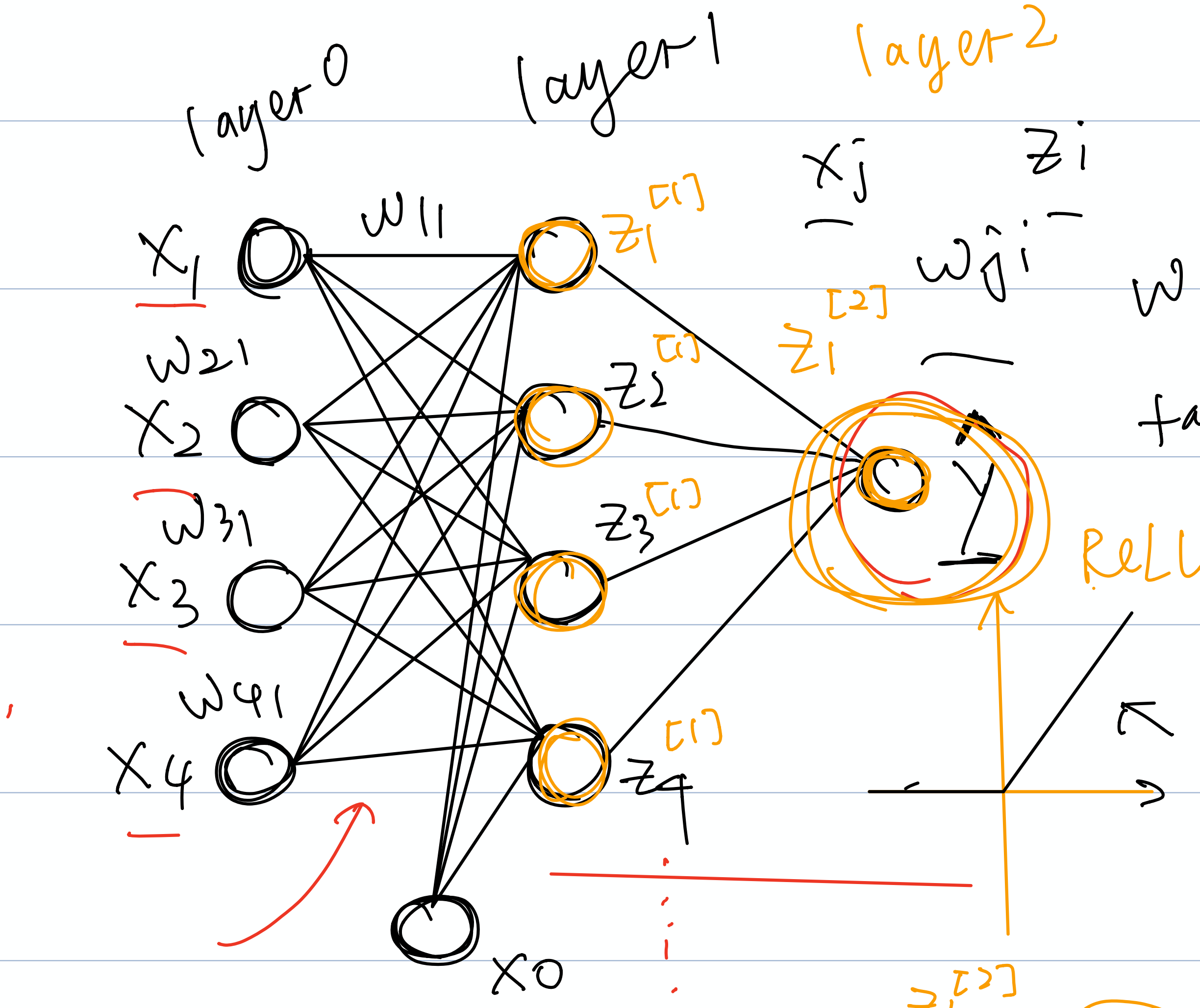
logistic regression: $\sigma(x) = \dfrac{1}{1+e^{-(\underbrace{w_0+w_1x_1+w_2x_2}_{z})}}$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \to \sigma(\cdot) \to \begin{bmatrix} \sigma(z_1) \\ \sigma(z_2) \\ \sigma(z_3) \end{bmatrix}$$

$z = w_0 + w_1 x_1 + w_2 x_2$

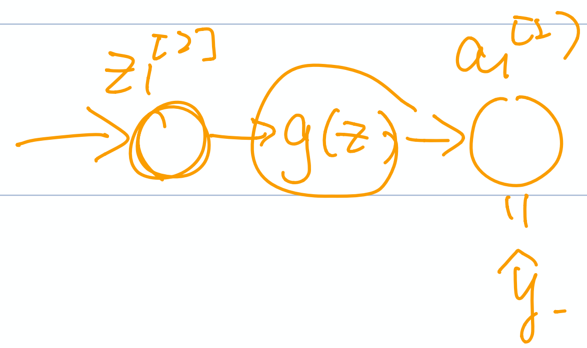
$\sigma(z) = \dfrac{1}{1+e^{-z}}$

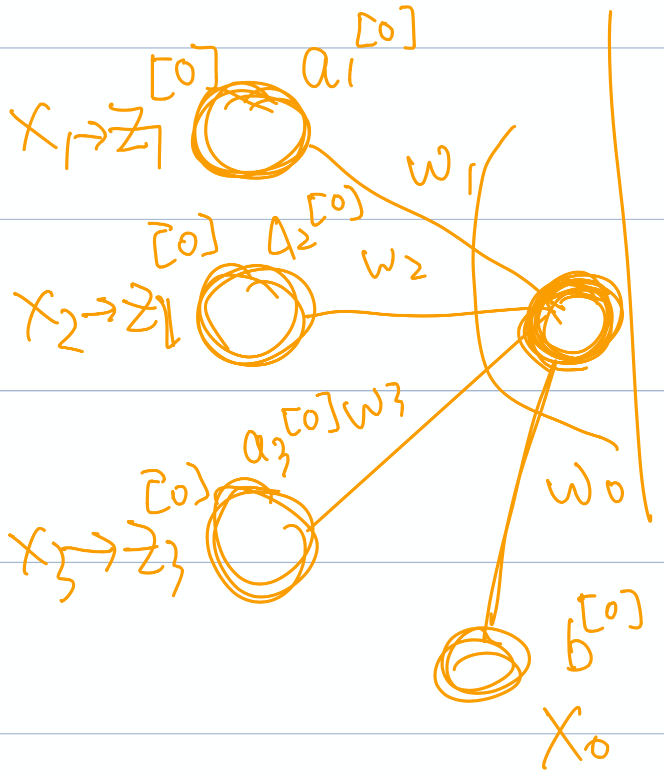




$$\begin{cases} 0 & x \le 0 \\ x & x > 0 \end{cases}$$

$$\hat{y} = g(z_1^{[2]}) = a_1^{[2]}$$

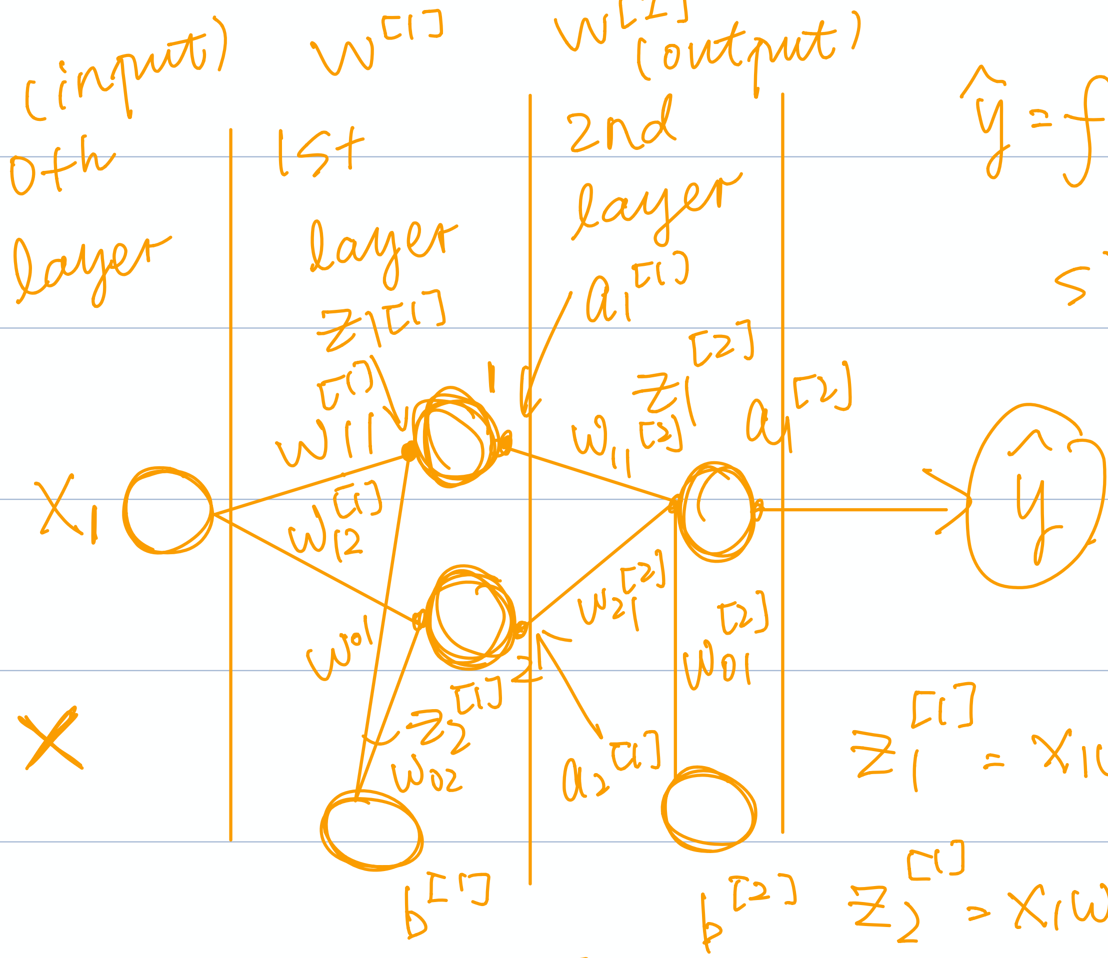




$$a_1^{[0]} w_1^{[1]} + a_2^{[0]} w_2^{[1]} + a_3^{[0]} w_3^{[1]} + w_0^{[1]} \cdot a_0^{[0]}$$

$$= z_1^{[1]}$$

$$\sigma(z_1^{[1]}) = a_1^{[1]} = \hat{y}.$$



$\hat{y} = f(x)$

sin cos
e^x

$$z_1^{[1]} = x_1 w_{11}^{[1]} + w_{01} \cdot b^{[1]}$$

$$z_2^{[1]} = x_1 w_{12}^{[1]} + w_{02} b^{[1]}.$$

$$g(z_1^{[1]}) = a_1^{[1]}$$

$$g(z_2^{[1]}) = a_2^{[1]}$$

$$z_1^{[2]} = a_1^{[1]} \cdot w_{11}^{[2]} + a_2 w_{21}^{[2]} + b^{[2]} w_{01}^{[2]}$$

$$h(z_1^{[2]}) = a_1^{[2]} = \hat{y}.$$

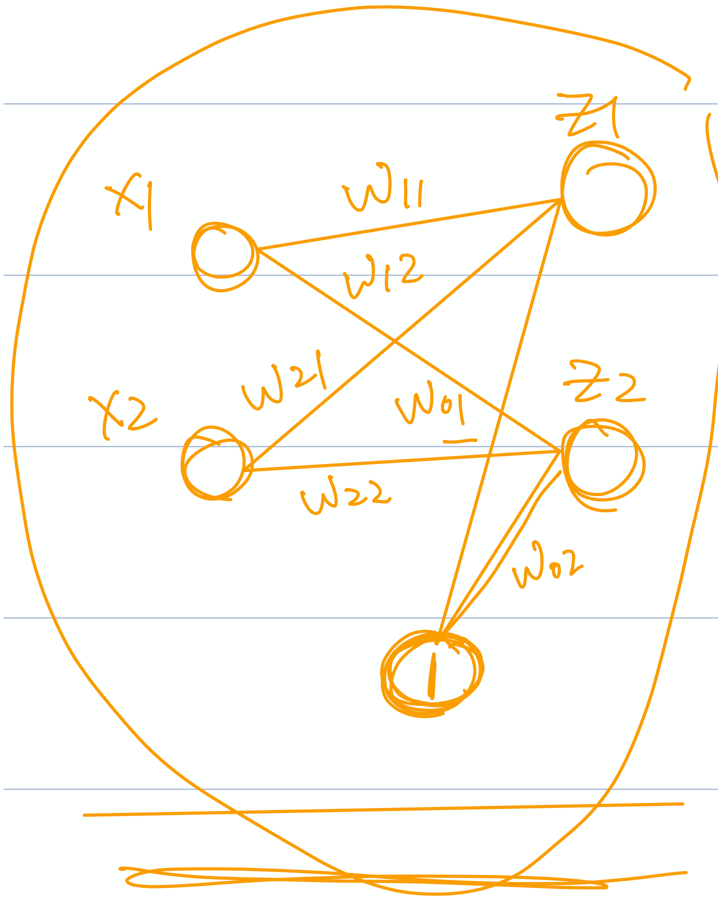
$$Z^{[1]} = W^{[1]} X + \mathbb{1} \cdot b^{[1]}$$

$$g(Z^{[1]}) = A^{[1]}$$

$$g(Z^{[2]}) = A^{[2]} = \hat{y}.$$

$$Z^{[2]} = W^{[2]} A^{[1]} + \mathbb{1} b^{[2]}$$

$$\underline{2}\times\underline{2}\times 2\times 1 = 2\times 1$$



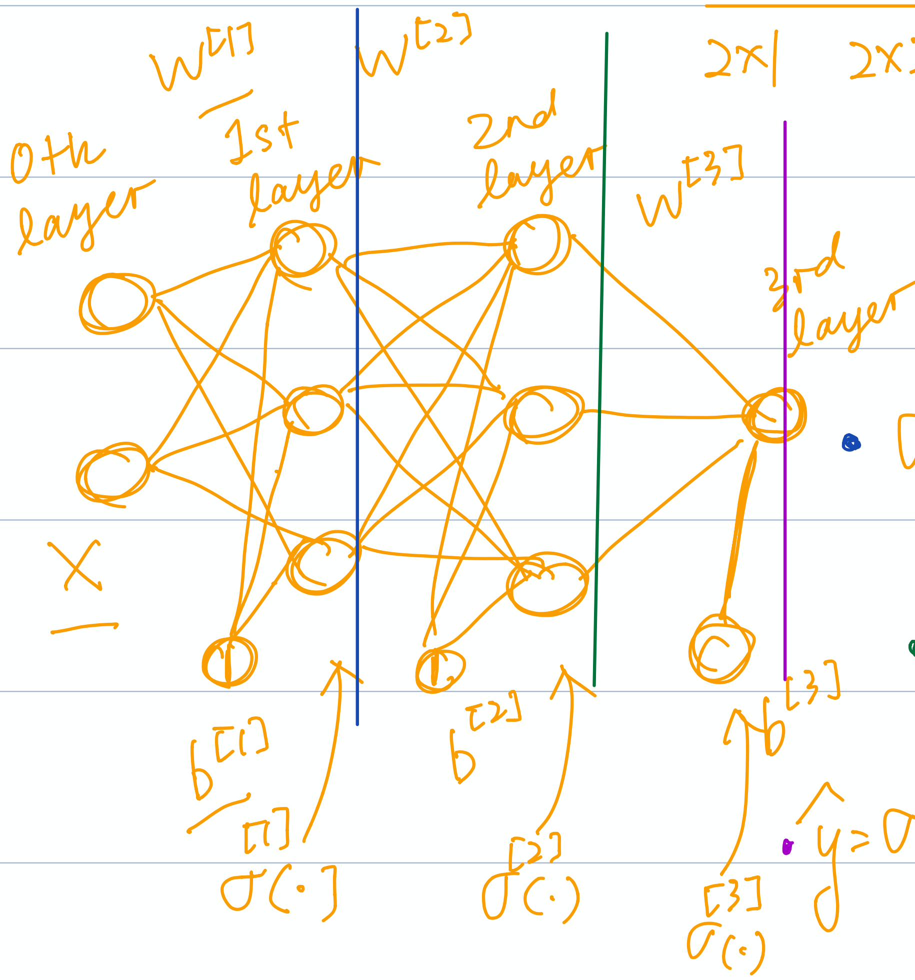
$$\underset{2\times 1}{\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}} = \underbrace{\underset{2\times 2}{\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}}}_{W} \underbrace{\underset{2\times 1}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}}_{X}$$

$$Z \qquad + \underbrace{\begin{bmatrix} w_{01} \\ w_{02} \end{bmatrix}}_{b,\ 2\times 1} \cdot 1$$

$$\underline{Z = WX + b}$$

$$2\times 1 \quad 2\times 2 \quad 2\times 1 \quad 2\times 1$$

$W^{[1]}$ X $b^{[1]}$



- $\underline{\sigma^{[1]}(W^{[1]}X + b^{[1]})}$ " $A^{[1]}$
- $\sigma^{[2]}(W^{[2]}A^{[1]} + b^{[2]})$ " $A^{[2]}$
- $\hat{y} = \sigma^{[3]}(W^{[3]}A^{[2]} + b^{[3]})$

$$\hat{y} = \sigma^{[3]}(W^{[3]}A^{[2]} + b^{[3]})$$

$$= \sigma^{[3]}(W^{[3]}\sigma^{[2]}(W^{[2]}A^{[1]} + b^{[2]}) + b^{[3]}).$$

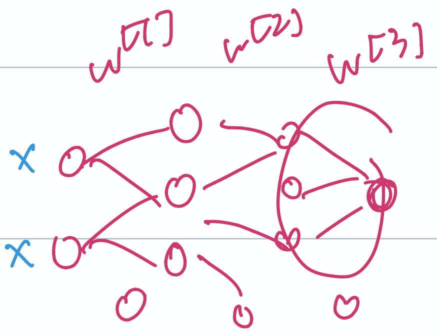
$$= \sigma^{[3]}(W^{[3]}\sigma^{[2]}(W^{[2]}\sigma^{[1]}(W^{[1]}X + b^{[1]}) + b^{[2]}) + b^{[3]}).$$

$$(W^{[3]}\cdot(W^{[2]}\cdot(W^{[1]}X + b^{[1]}) + b^{[2]}) + b^{[3]})$$

$$= (W^{[3]}(W^{[2]}W^{[1]}X + W^{[2]}b^{[1]} + b^{[2]}) + b^{[3]})$$

$$= \underbrace{W^{[3]}W^{[2]}W^{[1]}X}_{1\times3\times3\times3\times3\times2} + \underbrace{W^{[3]}W^{[2]}b^{[1]}}_{1\times1} + \underbrace{W^{[3]}b^{[2]} + b^{[3]}}_{1\times1}$$

1×2



$$\underline{1\times3\times3\times1\times1\times1}$$

$$z^{[3]} = W^{[3]}A^{[2]} + b^{[3]}$$

1×1   1×3   3×1   1×1

$$z^{[2]} = W^{[2]}A^{[1]} + b^{[2]}$$

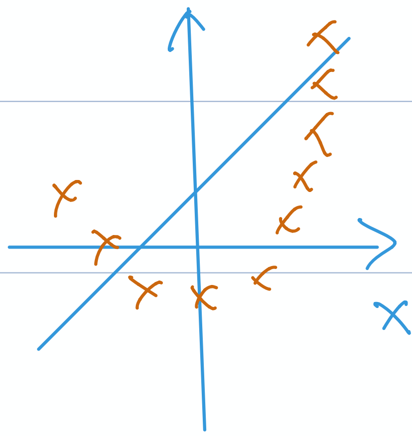
3×1   3×3   3×1   3×1

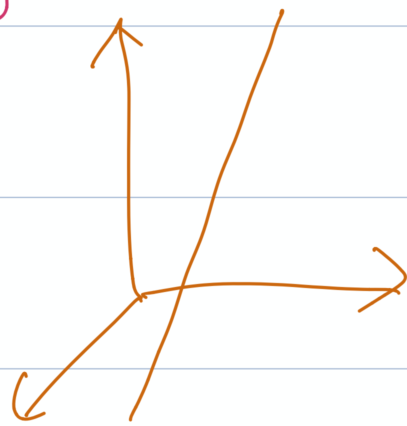
$$z^{[1]} = W^{[1]}X + b^{[1]}$$

3×1   3×2   2×1   3×1

$$\underline{W^{[3]}W^{[2]}W^{[1]}X} + \underbrace{W^{[3]}W^{[2]}b^{[1]} + W^{[3]}b^{[2]} + b^{[3]}}_{b}$$

$$= WX + b \rightarrow \text{linear regression.}$$





64 32

128

256

1024

100,000

dataset

1024

mini batch

$\rightarrow \hat{y} \rightarrow$ loss

MSE

gradient

backpropagation

adam

rmsprop

sgd