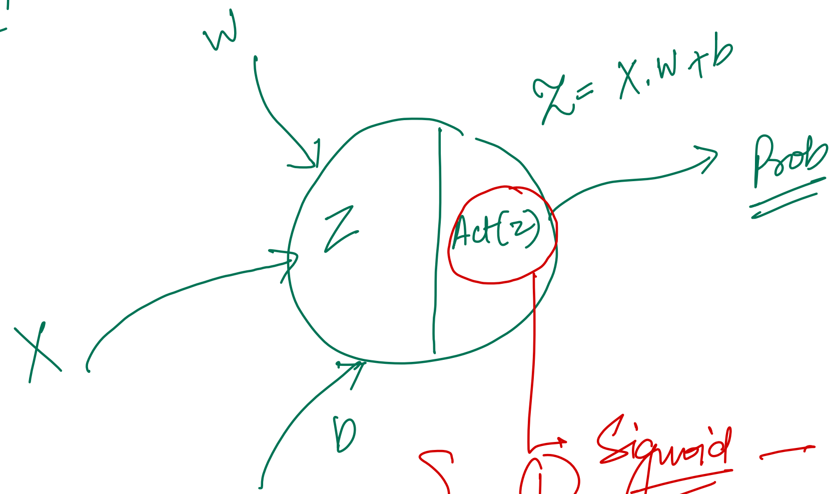
Neuron



1. Sigmoid → Log Reg Unit

$\Downarrow$

2. Softmax

1 hidden → 52%

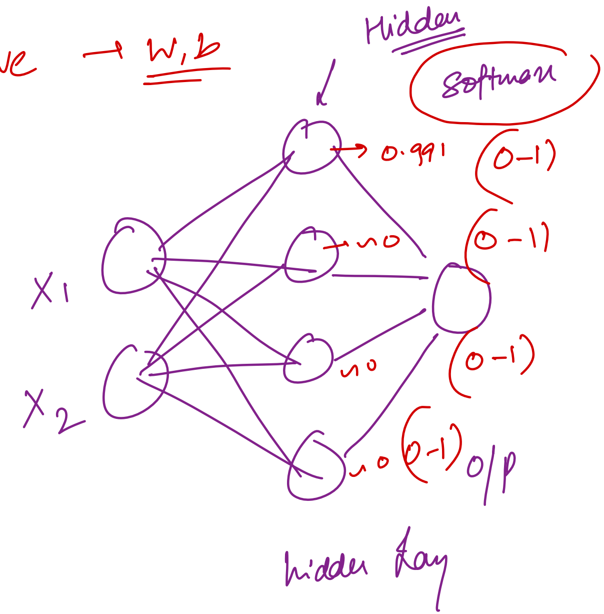
Linear decision boundary

Activation

1. Differentiable (through some hacks)
2. Compute inexpensive → W, b
3. non-linear

$$\frac{\partial a}{\partial z} \Rightarrow \left(\frac{1}{1+e^{-z}}\right)$$

Sigmoid



Activation

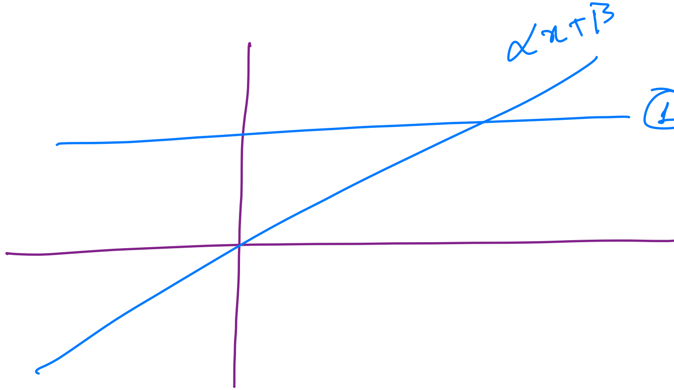
1. Regression
   ↳ linear

Trained

$$act(z) = \alpha x + \beta$$

$z \rightarrow (-\infty, \infty)$

$\hat{y} = (10, \infty)$



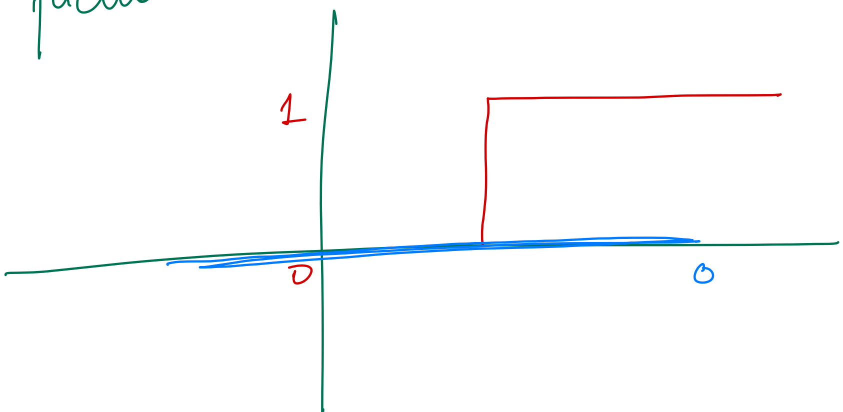
$\alpha x + \beta$

① Derivative

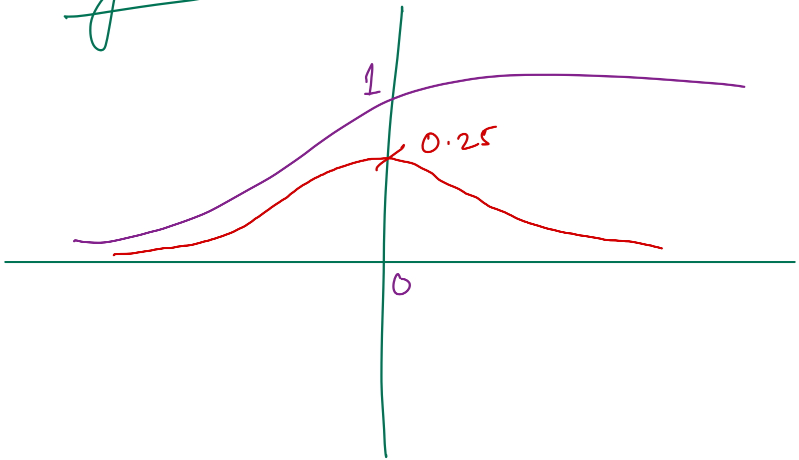
$z > 0 \quad = 1$

$z < 0 \quad = 0$

② Step fuction



③ Sigmoid



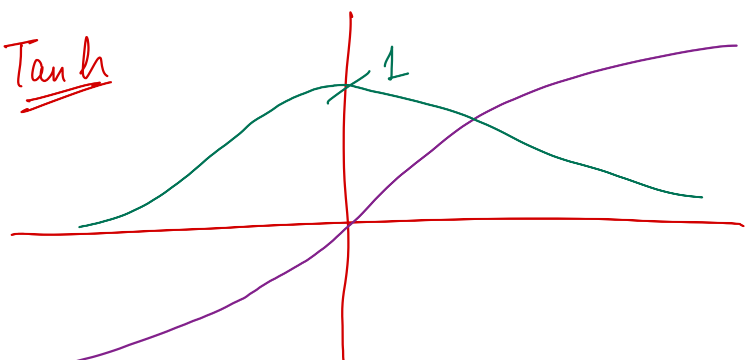
$y =$

$[0, 1]$

$$Act(z) = \frac{1}{1 + e^{-z}}$$

④ Softmax

⑤ Tanh

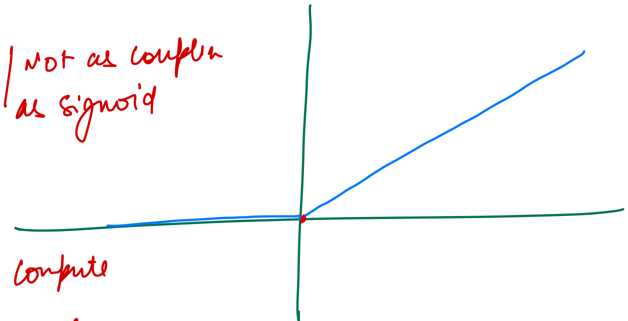


$[-1, 1]$

$$Act(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

6. ReLU → Rectified Linear Unit

1. Simple | Not as complex as sigmoid



$$Act(z) = \begin{cases} z > 0, & z \\ z \le 0, & 0 \end{cases}$$

2. Faster compute
3. Differentiable
4. Simplest Non-linear function → $\max(0, z)$

In hidden $\alpha x + \beta$

Linear Act function
└ Non-linearity ✗

Step function → kills a lot of info.

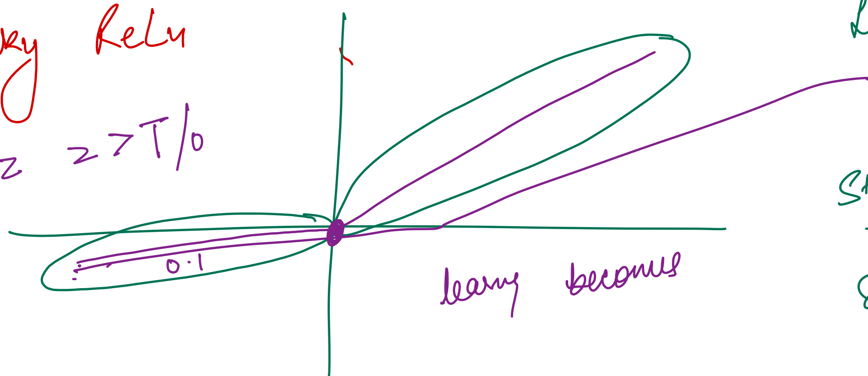
Sigmoid →

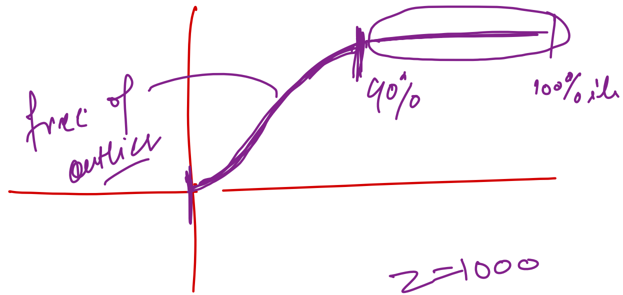
→ Squashing effect

7. Leaky ReLU

$Act(z) = z$ $z > T/0$

$0.1z$ $z < 0$





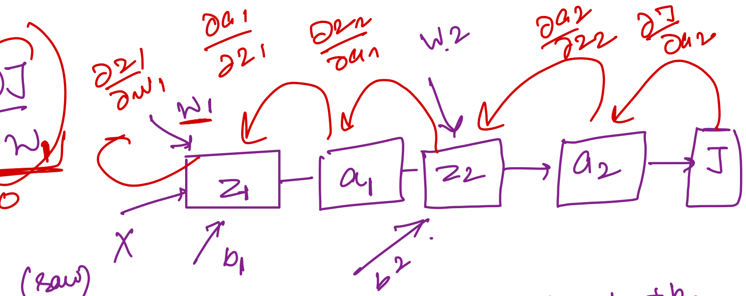
Tanh → similar to sigmoid.

→ use-case dependent.

ReLU → hidden layers.

$z = 1000$

$$w^{new} = w^{old} - \alpha \underbrace{\frac{\partial J}{\partial w}}_{\sim 0}$$



$$\boxed{\frac{\partial J}{\partial w} \sim 0}$$

$z_1 = X \cdot w_1 + b_1$

$z_2 = a_1 \cdot w_2 + b_2$

Sigmoid → 0.25

50 layer

$0.25 \times 0.25 \times 0.25 \ldots$

~ 0

Vanishing gradients

$$\boxed{w^{new} = w^{old}}$$

Tanh $\Rightarrow$ 1

$$\frac{\partial a}{\partial z} = 1$$

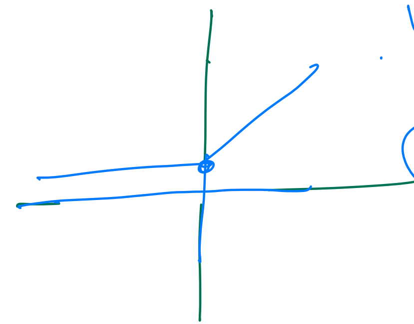
$Z_1 \rightarrow 2 \times 2 \times 2 \times 2$

ReLu $\rightarrow$ max(0, z) $\rightarrow$ exploding gradients

Relu = max(0, z)

$$\frac{\partial a}{\partial z} = 1$$

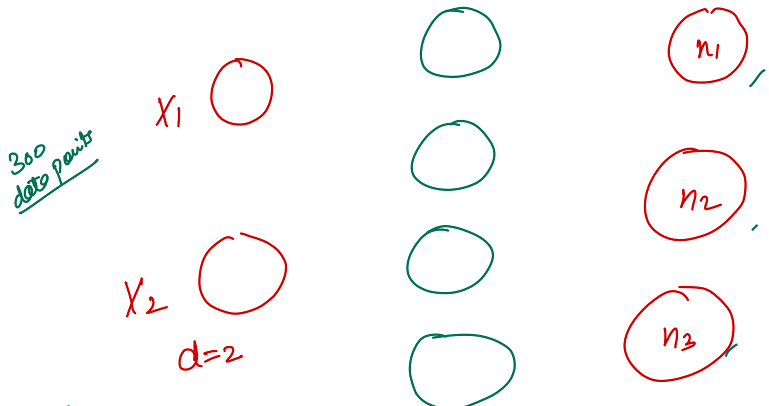
$$\frac{\partial a}{\partial z} \neq 1 \;,\; 0$$



vanishing

= 0

$$\text{Leaky}(z) = z \quad z > 0 \text{ or Threshold}$$
$$= 0.1z \quad z \le 0$$



Shapes

$A^0 = X \ (300, 2)$

$z^1 = (300, 4)$

$w^1 = (2, 4)$

$b^1 = (1, 4)$

Act(z_1)

$A^1 = (300, 4)$

$w^2 = (4 \times 3)$

$b^2 = (1, 3)$

$z^2 = (300, 3)$

$A^2 = \hat{y} = (300, 3)$

$$z^1 = X . w + b$$
$$= \underline{(300 \times 2) . (2 \times 4)} + (1, 4)$$
$$= (300, 4) + (1, 4)$$

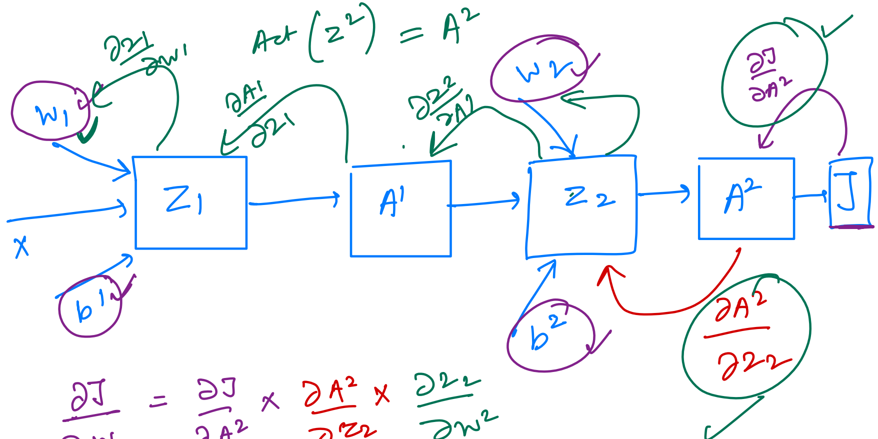
$$z^2 = A^1 . w^2 + b^2$$
$$= (300 \times 4) \ (4 \times 3) + (1, 3)$$
$$\Rightarrow (300, 3) + (1, 3)$$

$$Z^1 = X \cdot W_1 + b_1$$

$$A^1 = Act(Z^1)$$

$$Z^2 = A^1 \cdot W_2 + b_2$$

$$Act(Z^2) = A^2$$



$$\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial A^2} \times \frac{\partial A^2}{\partial Z_2} \times \frac{\partial Z_2}{\partial W^2}$$

$$\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial A^2} \times \frac{\partial A^2}{\partial Z^2} \times \frac{\partial Z2}{\partial A^1} \times \frac{\partial A1}{\partial Z1} \times \frac{\partial Z1}{\partial W^1}$$

$$\text{loss} = \frac{-1}{1} \; y \log(a) + (1-y) \log(1-a)$$

W^2, W_2

$$\frac{\partial J}{\partial w^2} = \underbrace{\frac{\partial J}{\partial A^2} \times \frac{\partial A^2}{\partial z^2}}_{a-y} \times \frac{\partial z^2}{\partial w^2}$$

$$\boxed{\frac{\partial J}{\partial w^2} = (A^2 - y)\cdot A^1}$$

$$\boxed{\frac{\partial J}{\partial b^2} = (A^2 - y)}$$

$$\frac{\partial J}{\partial w^1} = \underbrace{\frac{\partial J}{\partial A^2} \times \frac{\partial A^2}{\partial z^2}}_{} \times \overbrace{\frac{\partial z^2}{\partial A^1}}^{w^2} \times \overbrace{\frac{\partial A^1}{\partial z^1}}^{\text{Relu} \to (0,1)} \times \frac{\partial z^1}{\partial w^1}$$

$$\boxed{\frac{\partial J}{\partial w^1}} = (A^2 - y)\cdot w^2 \cdot 0 \cdot X = 0$$

or

$$\boxed{\frac{\partial J}{\partial w^1} = (A^2 - y)\, w^2 \cdot X}$$ Reduced form

$$\boxed{\frac{\partial J}{\partial b^1} = (A^2 - y)\, w^2}$$ Reduced form

$$z^2 = A^1 . W_2 + b_2$$

$$\frac{\partial z^2}{\partial w^2} = A^1$$

$$z^2 = \cancel{A^1 w^2}^{\,0} + b^2$$

$$\frac{\partial z^2}{\partial b^2} = 1$$

$$z^2 = A^1 w^2 + b^2$$

$$\frac{\partial z^2}{\partial A^1} = W^2$$

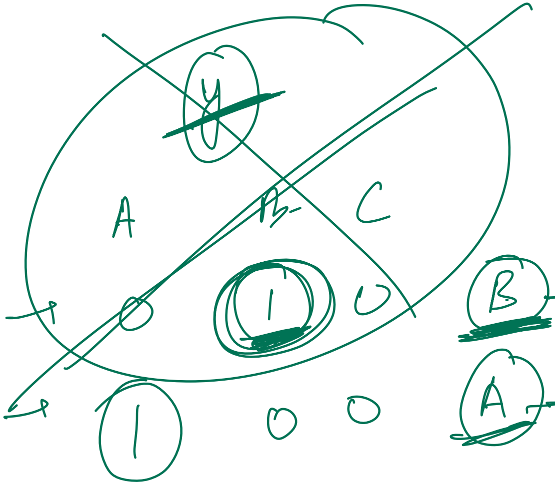
$$z^1 = X . w^1 + b^1$$

$$\frac{\partial z^1}{\partial w^1} = X$$

$$A^1 = \text{Relu}(z^1)$$
$$\max(0, z^1)$$

$$A^1 = \begin{cases} 0 & z^1 \le 0 \\ z^1 & z^1 > 0 \end{cases}$$

$$\frac{\partial A^1}{\partial z^1} = 0,\ 1$$



$\hat{y}$ [0.33 0.33 0.33]

dz(range(m), y)

$\boxed{0.33}$ − 1 =

− 1 =

y (true)

$\hat{y}$ → OHE

A	B	C
0	1	0
1	0	0

0.33 0.33 0.33

0.73 0.54 .

$\boxed{y(\text{true}) - \hat{y}}$