Session-5

Neural Networks - 5
Back Prop - Again

Feb ~~14~~ 16, 2024

Agenda

To Code a full fledged NN From scratch

Lets Talk About Shapes

n neurons

$$\underbrace{\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{m1} & x_{m2} \end{bmatrix}_{m \times d}}_{\text{data matrix}} \cdot \underbrace{\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ \vdots & & \\ w_{21} & w_{22} & w_{23} \end{bmatrix}_{d \times n}}_{\text{weight matrix}}$$

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{m \times n} + \begin{bmatrix} b'_{11} & b'_{12} & b'_{13} \end{bmatrix}_{1 \times n}$$

$$\begin{bmatrix} - & - & - \\ - & - & - \\ - & - & - \\ b'_{11} & b'_{12} & b'_{13} \end{bmatrix}$$

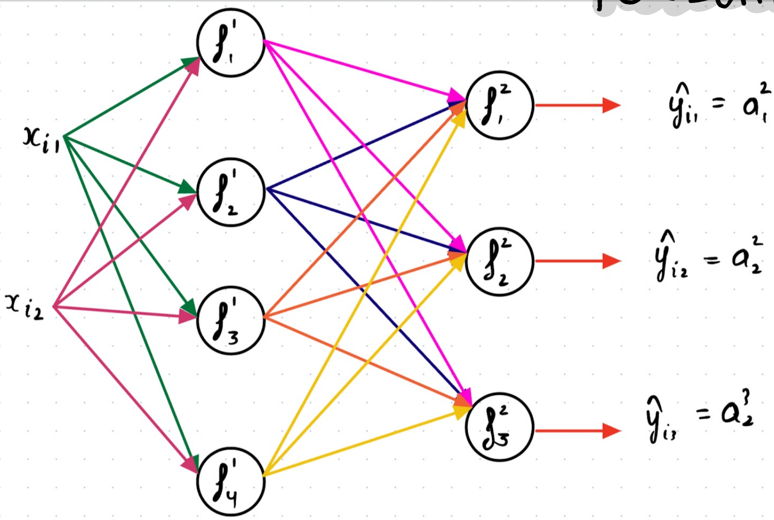
$m \times n$

Z matrix

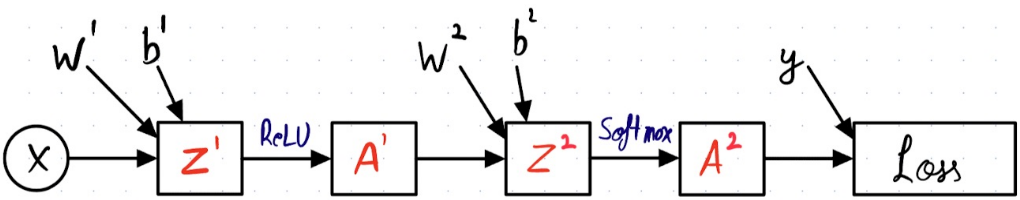
$$\begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ & & & \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix}$$

$m \times n$

Our Neural Network For Tonight



$X = (300, 2)$



d = 2 # dimension of inputs
n = 3 # no " classes
h = 4 # " " neurons in hidden Layer

```
> w1 = np.random.randn(d, h)
> b1 = np.zeros((1, h))
```

```
> w2 = np.random.randn(h, n)
> b2 = np.zeros((1, n))
```

→ **Forward Propagation**

```
> z1 = np.dot(x, w1) + b1  // (300,2) x (2,4)
```

300×4

$+$

$300,4 \quad b_1 (1,4)$

```
> A1 = np.maximum(0, z1)
```

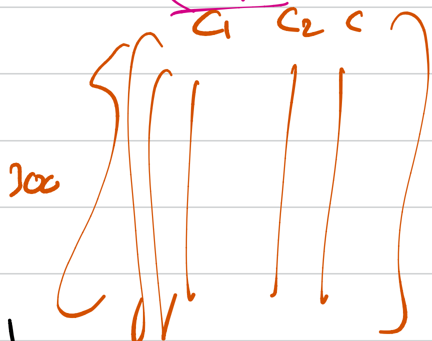
→ Second layer

```
> z2 = np.dot(A1, w2) + b2  // (300,4) x (4,3)
```

$= (300,3) + (1,3)$

$(300,3)$

→ $\text{Softmax} = \dfrac{e^{x_i}}{\sum_{i=1}^{n} e^{x_i}}$




```
> z2_exp = np.exp(z2)
> z2_exp_sum = np.sum(z2_exp,
                      axis=1,
                      keepdims=True)
> Probs = z2_exp / z2_exp_sum
```

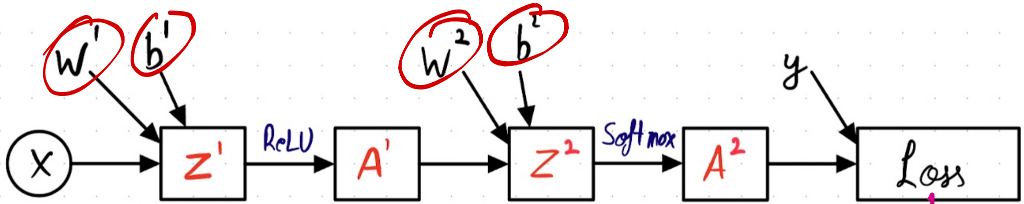
Forward Prop done!!

$$\frac{e^1}{e^1+e^2+e^3}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$\frac{e^4}{e^4+e^5+e^6}$$

→ **LETS DO BACKPROPAGATION**



$$\frac{\partial A2}{\partial Z2} \qquad \frac{\partial L}{\partial A2}$$

$$\frac{\partial L}{\partial w^2} = \frac{\partial L}{A^2} \times \frac{\partial A^2}{Z2} \times \frac{\partial Z2}{\partial w^2}$$

$$\frac{\partial L}{\partial z^2} = \left(P_i - I(i = y) \right)$$

P_i [0.2 0.3 0.5] [0 1 0] y actual

$= [0.2 \quad 0.3-1 \quad 0.5]$

$= [0.2 \quad -0.7 \quad 0.5]$ $\quad dz_2 = \frac{\partial L}{\partial z_2}$

> m = y.shape[0] // No of rows

> temp = Probs

> temp[range(m), y] -= 1

> dz2 = temp

$dz2 \begin{bmatrix} 1 & 2 & \boxed{3} \\ \boxed{4} & 5 & 6 \\ 2 & \boxed{8} & 0 \end{bmatrix}$ m=3

$dz_2[3, y] = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 5 & 6 \\ 2 & 7 & 0 \end{bmatrix}$

dw2

2 $\frac{\partial L}{\partial Z2}$ / $\frac{\partial L}{\partial w2}$

$$\frac{\partial L}{\partial w2} = \frac{\partial L}{\partial Z2} \times \underbrace{\frac{\partial Z2}{\partial w2}}_{A_1}$$

$$Z2 = \underbrace{\frac{\partial (w^2 A' + b2)}{\partial w2}}_{A1}$$

$$\frac{\partial L}{\partial w^2} = \underbrace{\frac{\partial L}{\partial Z^2}}_{(300,3)} \times A_1$$

A_1 → 300, 4

→ Why are we calculating $\frac{\partial L}{\partial w2}$

$$\underbrace{w^2}_{(4,3)} = w^2 - \eta \times \underbrace{\frac{\partial L}{\partial w2}}_{\text{need } (4,3)}$$

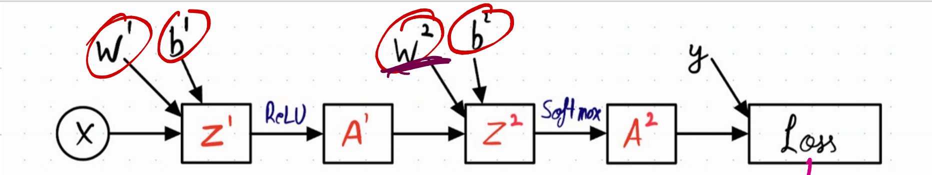
η → 0-3

$$\frac{\partial L}{\partial w^2} = A_1^T \cdot \frac{\partial L}{\partial Z^2}$$

$$w2 = \begin{bmatrix} a & e \\ b & c \\ c & \\ a & \end{bmatrix}_{4\times 3}$$

> $dw_2 = \dfrac{np.dot(A1.T, dZ_2)}{m}$ → Increase stability

② Calc → db_2



$\dfrac{\partial A2}{\partial Z2}$ $\dfrac{\partial L}{\partial A2}$

$$\frac{\partial L}{\partial b_2} = \underbrace{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}}_{\frac{\partial L}{\partial Z_2}} \times \underbrace{\frac{\partial Z_2}{\partial b_2}}_{1}$$

$$Z_2 = \frac{\partial (w^2 A^1 + b^2)}{\partial w_2}$$

$$dw_2 = \frac{\partial L}{\partial w_2}$$

$$\frac{\partial L}{\partial b2} = dz_2 \quad \} \; (300, 3)$$

Shape of b → (1, 3)

$$b = b_2 - \eta \times \frac{\partial L}{\partial b2}$$

(300) → summing → (1, 3)

> $db_2 = np.sum(dz^2, axis=0, keepdims=True)_m$

$\#\ (1, 3)$

→ Moving to Second layer.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = [5\ \ 7\ \ 9]/2$$

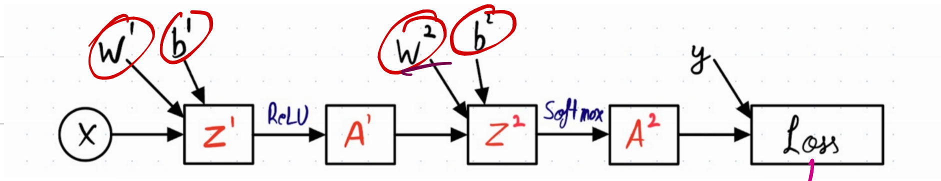
$$\Rightarrow [5/2\ \ 7/2\ \ 9/2]$$

y.shape = (no o rows, no of cloum)

x.shape[1]

$$= [0\ \ 1\ \ 0]$$
$$= [1\ \ 0\ \ 0]$$
$$= [0\ \ 0\ \ 1]$$

→ Moving to 1st layer



$$\frac{\partial Z2}{A1} \qquad \frac{\partial A2}{\partial Z2} \qquad \frac{\partial L}{\partial A2}$$

$Z_2 = \underline{W_2 A_1 + b_1}$

$$\frac{\partial Z2}{\partial A_1} = W_2$$

$$\boxed{\frac{\partial Z2}{\partial A_1} = W_2}$$

$$\frac{\partial L}{A_1} = \underbrace{\frac{\partial L}{A_2} \times \frac{\partial A_2}{\partial Z2}}_{dZ2} \times \left.\frac{\partial Z_2}{\partial A_1}\right\} \quad (4,3)$$

Some shape of A_1

(300,4)

$dZ2$

300, 3

$$\frac{\partial L}{\partial A_1} = \underbrace{\frac{\partial L}{\partial Z2}}_{300,3} \times \underbrace{W^{2.T}}_{3,4}\Big\} \quad 300,4$$

$> da_1 = $ np.dot(dZ2, w2.T)

(2) $\frac{\partial L}{\partial Z_1} = \frac{\partial L}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1}$ $A_1 =$

$$\frac{\partial A_1}{\partial Z_1} \begin{cases} \partial A \times 0 \;; & \text{if } Z_1 \le 0 \\ \partial A \times 1 \;; & \text{if } Z_1 > 0 \end{cases}$$

$A_1 = \max(0, Z_1)$

$$\frac{\partial L}{\partial Z_1} = \frac{\partial A_1}{\partial Z_1} \left[Z_1 \geq 0 \right]$$

$> dA_1 [Z_1 \le 0] = 0$

$> dZ_1 = dA_1$

$\rightarrow \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial w_1} = \frac{\partial (w'x + b')}{\partial w_1}$

$= X$

$= \frac{\partial L}{\partial Z_1} \times X$

$\rightarrow \quad \frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial z_1}$

> $dw_1 =$ np.dot$(x.T, dz_1)/m$

> $db_1 =$ np.sum(dz, axis=0, keepdims=True)/m

log(1) = 0

$-(x y_2)$ — log(0.1) (x) (~) (-) ~0

log(0.5)

low loss

Fish → Salmon, Tuna, Hammer sharks

Width	Height	Class
5cm	20cm	Salmon
6cm	21cm	Salmon
30cm	200cm	H. shark
40cm	300 cm	H. shark
10cm	50cm	Tuna
15cm	60 cm	Tuna