Name:- Bidit Sadhukhan.

Reg.No:- B2230022

Assignment-1

1) x= torch.tensor([0.1, 0, 0.2, 1], dtype=torch.float32, requires_grad=True).

z= torch.tensor([-1, 1, -1, 0], dtype= torch.float32, requires_grad=True).

a= x+ 10 % 4 +2

b= a** (4*x)

c= (b+3)*5

y= c.mean()

y.backward()

$$\frac{\delta y}{\delta z_i} = \frac{\delta y}{\delta c_i} \cdot \frac{\delta c_i}{\delta b_i} \cdot \frac{\delta b_i}{\delta a_i} \cdot \frac{\delta a_i}{\delta z_i} \qquad i = 1,2,3,4$$

$$\frac{\delta a_i}{\delta z_i} = 1$$

$$b_i = a_i^{4x_i} \Rightarrow \ln b_i = 4x_i \ln a_i$$

$$\Rightarrow \frac{1}{b_i}\frac{\delta b_i}{\delta z_i} = \frac{4x_i}{a_i} \cdot \frac{\delta a_i}{\delta z_i}$$

$$\Rightarrow \frac{\delta b_i}{\delta z_i} = 4x_i\, a_i^{4x_i - 1}$$

$$\frac{\delta c_i}{\delta b_i} = 1$$

$$\frac{\delta y}{\delta c_i} = \frac{1}{4} \times 5 \times 4x_i\, a_i^{4x_i - 1}$$

$$\therefore \frac{\delta y_i}{\delta z_i} = 5x_i\, a_i^{4x_i - 1}$$

Substituting the values of a_i, x_i we get,

$\therefore$ z.grad([0.4722, 0.0000, 0.9642, 135.0000])

$$\frac{\delta y}{\delta x_i} = \frac{\delta y}{\delta c_i} \cdot \frac{\delta c_i}{\delta b_i} \cdot \frac{\delta b_i}{\delta x_i}$$

$$y = 1/4\,(c_1 + c_2 + c_3 + c_4).$$

$$\frac{\delta y}{\delta c_i} = 1/4$$

$$\frac{\delta c_i}{\delta b_i} = 5.$$

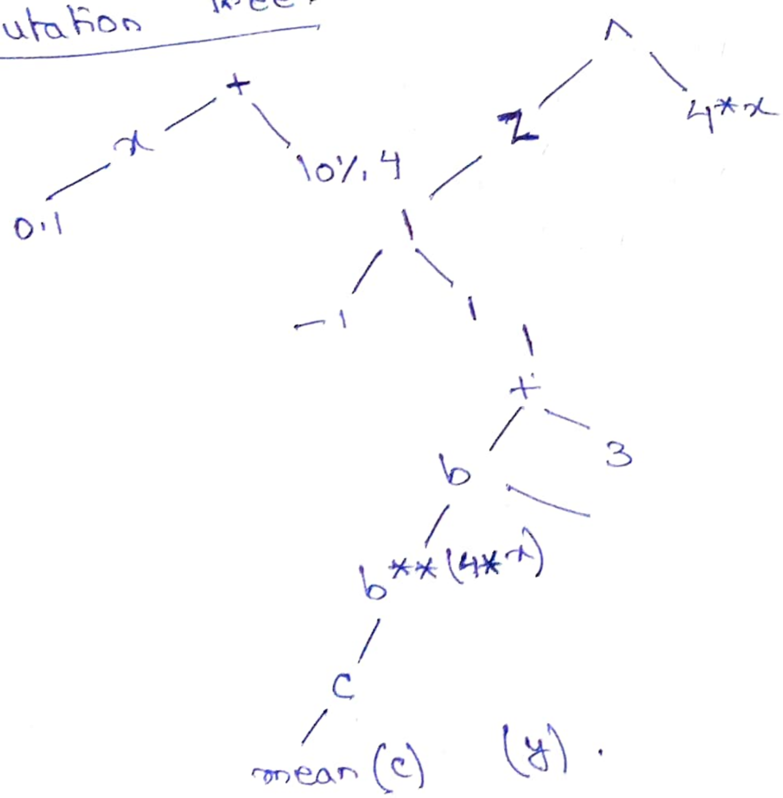
$$b = a^{4x_i}$$

$$\frac{\delta b_i}{\delta x_i} = 4x_1 a_1^{4x_1 - 1} + 4a_1^{4x_1} \ln a_1$$

$$\therefore x.\text{grad} = \Sigma\left[5\left(x_i a_i^{4x_i - 1}\right) + a_i^{4x_i} \ln a_i\right] \qquad i = 1, 2, 3, 4.$$

$$= ([0.9673,\ 5.4931,\ 2.0190,\ 579.9380])$$

Computation Tree:-



2a) $TSS = \frac{1}{2}\left[(t_1-O_1)^2+(t_2-O_2)^2\right]$

$$\Delta w_1 = -\eta \frac{\delta TSS}{\delta w_{11}}$$

$$\frac{\delta TSS}{\delta w_{11}} = \frac{\delta TSS}{\delta O_1} \cdot \frac{\delta O_1}{\delta net} \cdot \frac{\delta net}{\delta w_{11}}$$

$$\frac{\delta TSS}{\delta O_1} = -(t_1-O_1) \quad \dots \text{(i)}$$

$$O_1 = \frac{1}{1+e^{-net}}$$

$$\frac{\delta O_1}{\delta net} = O_1(1-O_1) \quad \dots \text{(ii)}$$

$$net = \sum_{i=0}^{n} w_{ij} x_i$$

$$\frac{\delta net}{\delta w_{11}} = x_1 \quad \dots \text{(iii)}$$

$$\therefore \Delta w_1 = -\eta\left[\frac{\delta TSS}{\delta O_1} \cdot \frac{\delta O_1}{\delta net} \cdot \frac{\delta net}{\delta w_{11}}\right]$$

$$= -\eta(-t_1+O_1) \cdot O_1(1-O_1) \cdot x_1$$

$$= \eta(t_1-O_1) \cdot O_1(1-O_1) \cdot x_1$$

2b) Problem of Vanishing Gradient:-

① ~~Large~~ Non-normalized input values:- If the input values are (x_i) very large or small, the weighted sum at the hidden neuron can quickly move away from 0 during the forward pass. This leads to a near zero derivative (σ') for the sigmoid activation f^n, resulting in vanishing gradients during backpropagation.

① Activation function:- If the network gets stuck in a state where the outputs are close to 0 or 1 due to bias or other factors. The error signal (t_1-O_1) will be very small. This can also lead to vanishing gradients, as the small error signal gets multiplied by the small derivative (σ') during backpropagation.