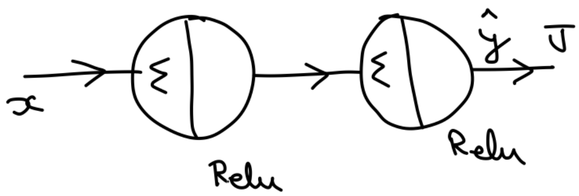
Backpropagation

Problem 1: What is being back propagated?

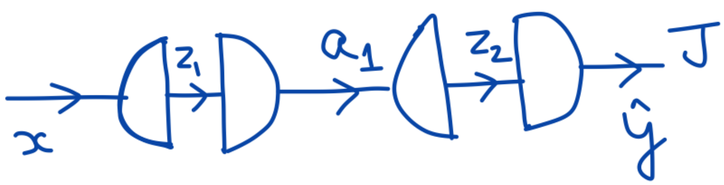
Problem 2: In the following setup




```
 - How many functions are involved?
 - How many weights are involved?
 - With only one training sample x=3, y=2 and weights w1=0.5, w2=0.8 Whats y
_hat?
 - Whats the MSE Cost?
```

Solution 2:

```
- 5 Functions: z1, a1, z2, y_hat, J
- 2 Weights: w1, w2
-
```

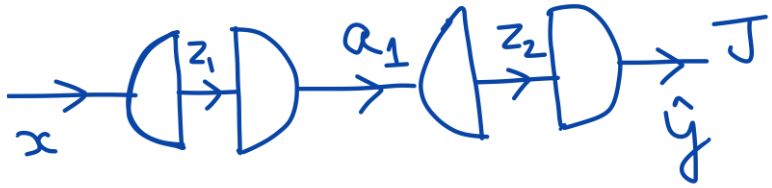


X	Z_1	a_1	Z_2	$\hat{y}$	y	$J(y,\hat{y})$
a_0	$w_1 x$	$Relu(z_1)$	$w_2 a_1$	a_2 $Relu(z_2)$		$\frac{(y-\hat{y})^2}{2N}$
3	$0.5x$		$0.8\,a_1$		2	
3	1.5	1.5	1.2	1.2	2	0.32

Step Forward Layer 1 **Step Forward Layer 2**

Problem 3:

Just expanded the above neural network to show weighted sum and activation components



To go backward, compute the following for previously given weights and training sample

$$\frac{dJ}{d\hat{y}} \quad \frac{d\hat{y}}{dz2} \quad \frac{dz2}{dw2} \quad \frac{dz2}{da1} \quad \frac{da1}{dz1} \quad \frac{dz1}{dw1}$$

Solution 3:

Functions	Derivatives	value of derivatives
$J(\hat{y}) = \frac{(y-\hat{y})^2}{2N}$	$\frac{dJ}{d\hat{y}} = -(y-\hat{y})$	-0.8
$\hat{y}(z_2) = Relu(z_2)$	$\frac{d\hat{y}}{dz_2} = \begin{cases} 1 \text{ if } z_2>0 \\ 0 \text{ if } z_2<0 \end{cases}$	1
$z_2(a_1) = w_2 a_1$	$\frac{dz_2}{da_1} = w_2$	0.8
$z_2(w_2) = w_2 a_1$	$\frac{dz_2}{dw_2} = a_1$	1.5
$a_1(z_1) = Relu(z_1)$	$\frac{da_1}{dz_1} = \begin{cases} 1 \text{ if } z_1>0 \\ 0 \text{ if } z_1<0 \end{cases}$	1
$z_1(w_1) = w_1 x$	$\frac{dz_1}{dw_1} = x$	3

Problem 4:

Whole purpose of back propagation or this notebook is to compute ***how cost changes with weights*** i.e

$$\frac{dJ}{dweights}$$

What are the values of:

a) $\frac{dJ}{dw1}$

b) $\frac{dJ}{dw2}$

(Hint: Use Chain Rule & Problem 3 results)

Solution 4:

Propagating cost to w2

$$\frac{dJ}{d\hat{y}} \qquad -0.8$$

$$\frac{dJ}{dz_2} = \frac{dJ}{d\hat{y}} \cdot \frac{d\hat{y}}{dz_2} \qquad -0.8$$

$$\frac{dJ}{dw_2} = \frac{dJ}{dz_2} \cdot \frac{dz_2}{dw_2} \qquad -1.2$$

Propagating cost to w1

$$\frac{dJ}{d\hat{y}} \qquad -0.8$$

$$\frac{dJ}{dz_2} = \frac{dJ}{d\hat{y}} \cdot \frac{d\hat{y}}{dz_2} \qquad -0.8$$

$$\frac{dJ}{da_1} = \frac{dJ}{dz_2} \cdot \frac{dz_2}{da_1} \qquad -0.64$$

$$\frac{dJ}{dz_1} = \frac{dJ}{da_1} \frac{da_1}{dz_1} \qquad -0.64$$

$$\frac{dJ}{dw_1} = \frac{dJ}{dz_1} \frac{dz_1}{dw_1} \qquad -1.92$$

5. Update Weights

Problem 1: In the above problem after you have computed $\frac{dJ}{dw1}$ and $\frac{dJ}{dw2}$

what are new weights if the learning rate is 0.1

Solution 1:

Weight Update
(Gradient Descent)

$$w_1 := w_1 - \alpha \frac{dJ}{dw_1}$$

α: learning rate

$$w_2 := w_2 - \alpha \frac{dJ}{dw_2}$$

$$w_1 := 0.5 - 0.1(-1.92)$$

$$w_2 := 0.8 - 0.1(-1.2)$$

w_1
0.692

$w2$
0.92

Forward Again

Problem 1 : Plug in those new weights and check if the cost is reduced.

Solution 1 :

Forward Epoch 2

X	Z_1	a_1	Z_2	$\hat{y}$	y	$J(y,\hat{y})$
a_0	$w_1 x$	$Relu(z_1)$	$w_2 a_1$	a_2 $Relu(z_2)$		$\frac{(y-\hat{y})^2}{2N}$
3	$0.692x$		$0.92\, a_1$		2	
3	2.076	2.076	1.91	1.91	2	0.004

Step Forward Layer 1 **Step Forward Layer 2**