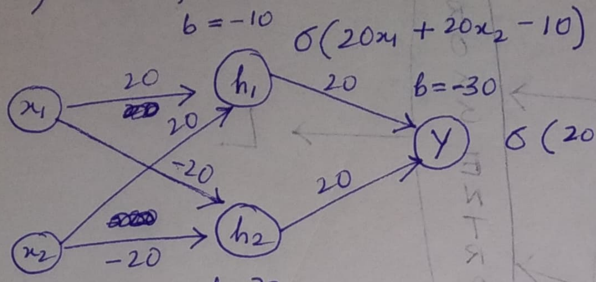
AI21MTECH02002 Assignment - 3 AI5000 FoML

1 a) XOR function perception with one hidden layer



$\sigma(-20x_1-20x_2+30)$

h_1

$\sigma(20*0+20*0-10) \approx 0$

$\sigma(20*1+20*1-10) \approx 1$

$\sigma(20*0+20*1-10) \approx 1$

$\sigma(20*1+20*0-10) \approx 1$

$x_1 \quad x_2$

h_2

$\sigma(-20*0-20*0+30) \approx 1$

$\sigma(-20*1-20*1+30) \approx 0$

$\sigma(-20*0-20*1+30) \approx 1$

$\sigma(-20*1-20*0+30) \approx 1$

$x_1 \quad x_2$

y

$\sigma(20*0+20*1-30) \approx 0$

$\sigma(20*1+20*0-30) \approx 0$

$\sigma(20*1+20*1-30) \approx 1$

$\sigma(20*1+20*1-30) \approx 1$

$h_1 \quad h_2$

XOR

x_1	x_2	h_1	h_2	y
0	0	0	1	0
1	1	1	0	0
0	1	1	1	1
1	0	1	1	1

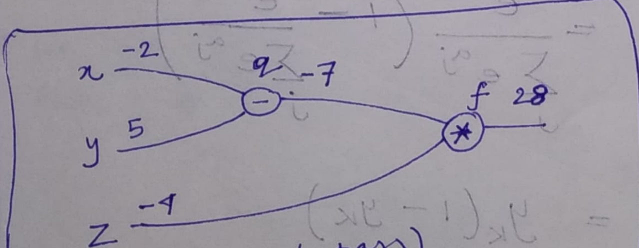
b) $q = x - y$

$f = q * z$

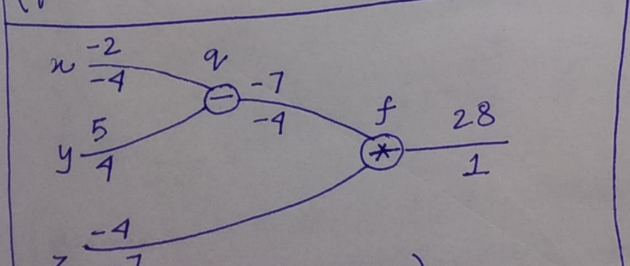
$x = -2$

$y = 5$

$z = -4$



(After Forward pass)



(After Backprop)

$\frac{df}{dz} = q$ (gradient on $z = -7$)

$\frac{df}{dq} = z$ (gradient on $q = -4$)

First we backprop through $f = q*z$

Also,

$\frac{dq}{dx} = 1$

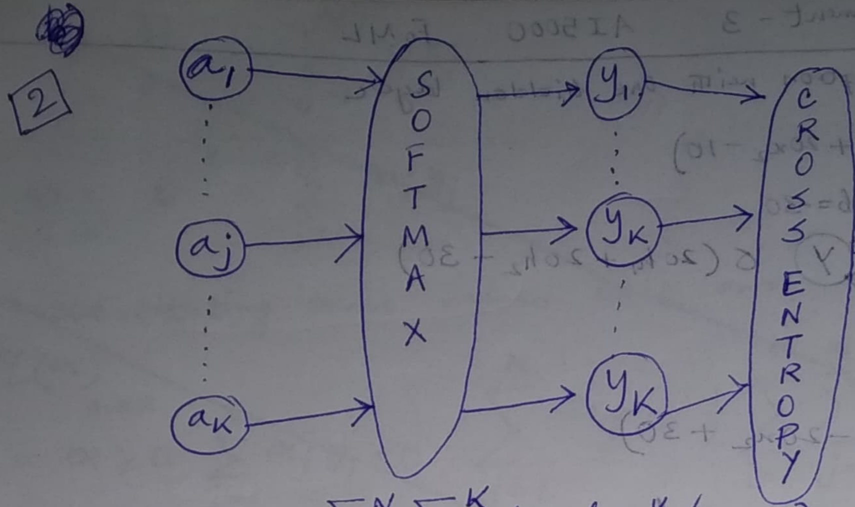
$\frac{dq}{dy} = -1$

Now, let's backprop through $q = x - y$

$\frac{df}{dx} = \frac{df}{dq} * \frac{dq}{dx} = z*1 = z = -4$

$\frac{df}{dy} = \frac{df}{dq} * \frac{dq}{dy} = z * -1 = -z = 4$

2



$$E(w) = -\sum_{n=1}^{N}\sum_{K=1}^{K} t_{Kn} \ln y_K(x_n, w)$$

let us write the above cross-entropy in simplified terms omitting the parameters and the number of data samples.

$$E = -\sum_{K=1}^{K} t_K \ln y_K$$

Also,

$$y_K(x, w) = \frac{e^{a_K(x,w)}}{\sum_j e^{a_j(x,w)}}$$ can be written as $$y_K = \frac{e^{a_K}}{\sum_j e^{a_j}}$$

Derivation of softmax:

$$\frac{\partial y_K}{\partial a_{j=K}} = \frac{e^{a_K}\sum_j e^{a_j} - e^{a_K}e^{a_K}}{\left(\sum_j e^{a_j}\right)^2} = \frac{e^{a_K}\left(\sum_j e^{a_j} - e^{a_K}\right)}{\sum_j e^{a_j}\cdot\sum_j e^{a_j}}$$

$$\left[\because y_K = \frac{e^{a_K}}{\sum_j e^{a_j}}\right]$$

$$= \frac{e^{a_K}}{\sum_j e^{a_j}}\left(1 - \frac{e^{a_K}}{\sum_j e^{a_j}}\right)$$

$$= y_K(1 - y_K)$$

$$\frac{\partial y_K}{\partial a_j} = \frac{0 - e^{a_K}e^{a_j}}{\left(\sum_j e^{a_j}\right)^2}$$

$$= -y_K y_j$$

$$\frac{\partial E}{\partial a_j} = -\sum_{K \neq j} \left(t_K \cdot \frac{1}{y_K} \cdot \frac{\partial y_K}{\partial a_j} \right) - t_j \cdot \frac{1}{y_j} \cdot \frac{\partial y_j}{\partial a_j}$$

$$= -\sum_{K \neq j} \left(t_K \cdot \frac{1}{y_K} \cdot (-y_K y_j) \right) - t_j \cdot \frac{1}{y_j} \cdot y_j (1 - y_j)$$

$$= \sum_{K \neq j} t_K y_j + t_j y_j - t_j$$

Considering all K classes (Both $K = j$ and $K \neq j$)

$$= \sum_K t_K y_K - t_K$$

$$= y_K \sum_K t_K - t_K$$

[$\because \sum_K t_K = 1$ as it is one-hot vector]

(Hence proved).

$$\boxed{\frac{\partial E}{\partial a_j} = y_K - t_K}$$

---

3 Jensen's Inequality: If $g(x)$ is a convex function, and $\mathbb{E}[g(X)]$ and $g(\mathbb{E}[X])$ are finite,

then

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

We have $f(x) = x^2$ a convex function.

Also, $E_{AV} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}_x \left[(y_m(x) - f(x))^2 \right]$

$$E_{ENS} = \mathbb{E}_x \left[\left(\frac{1}{M} \sum_{m=1}^{M} y_m(x) - f(x) \right)^2 \right]$$

By applying Jensen's Inequality to the uniform measure on $\{1, 2, \dots, M\}$ we can write —

$$\left(\frac{1}{M} \sum_{m=1}^{M} (y_m(x) - f(x)) \right)^2 \leq \frac{1}{M} \sum_{m=1}^{M} (y_m(x) - f(x))^2$$

[Assume $X = y(x) - f(x)$ ~~which is a convex function~~ $f(x) = x^2$

$g(X) = X^2$ which is a convex function]

Take Expectation on both sides, we get

$$\mathbb{E}_x \left[\left(\frac{1}{M} \sum_{m=1}^{M} \left(y_m(x) - f(x) \right) \right)^2 \right] \leq \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}_x \left[\left(y_m(x) - f(x) \right)^2 \right]$$

or, $\boxed{E_{ENS} \leq E_{AV}}$ (Hence proved)

As, we can see that Jensen's Inequality is True whenever the function is convex. So, irrespective of the ~~functions~~ function, if it is convex Jensen's Inequality can be applied. We just need to prove the convexity of the underlying function. Jensen's Inequality applies to every convex function.