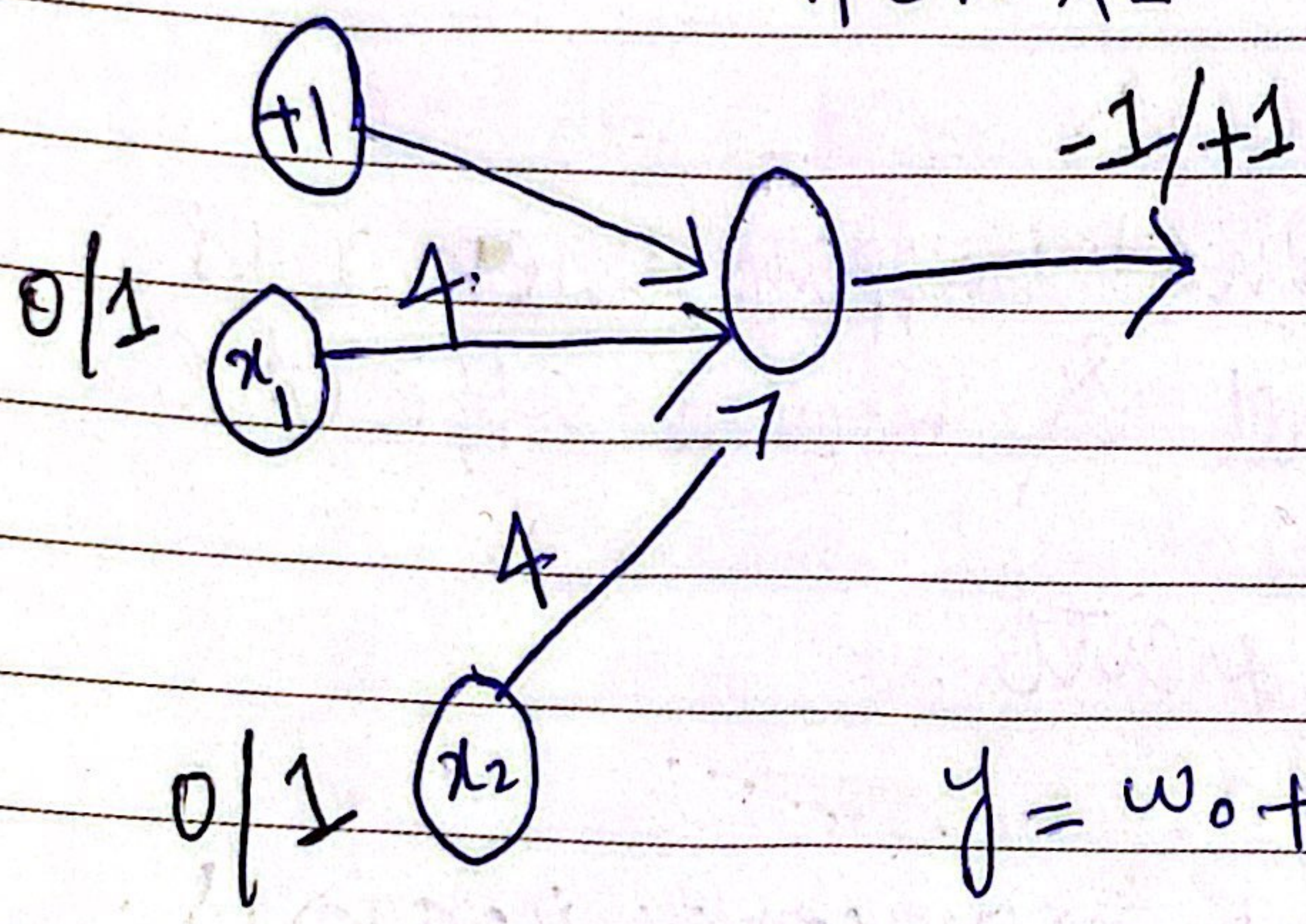
Amina Tabassum
NUID: 002190127
HW #03

Question NO: 01

x_1 OR x_2



$$y = w_0 + w_1 x_1 + w_2 x_2$$

for OR gate

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

AND gate

x_1	x_2	y	$\bar{y}$
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

for tanh function:

output = -1 for input $x \leq -2$
output = +1 for input $x \geq +2$
output = 0 for $x = 0$

So, selecting $w_0 = -2$ gives:

$$y = w_0 + w_1 x_1 + w_2 x_2 = -2 + 4x_1 + 4x_2$$

$x_1 = 0, x_2 = 0$ gives $y = -2 \Rightarrow \tanh(y) = -1$

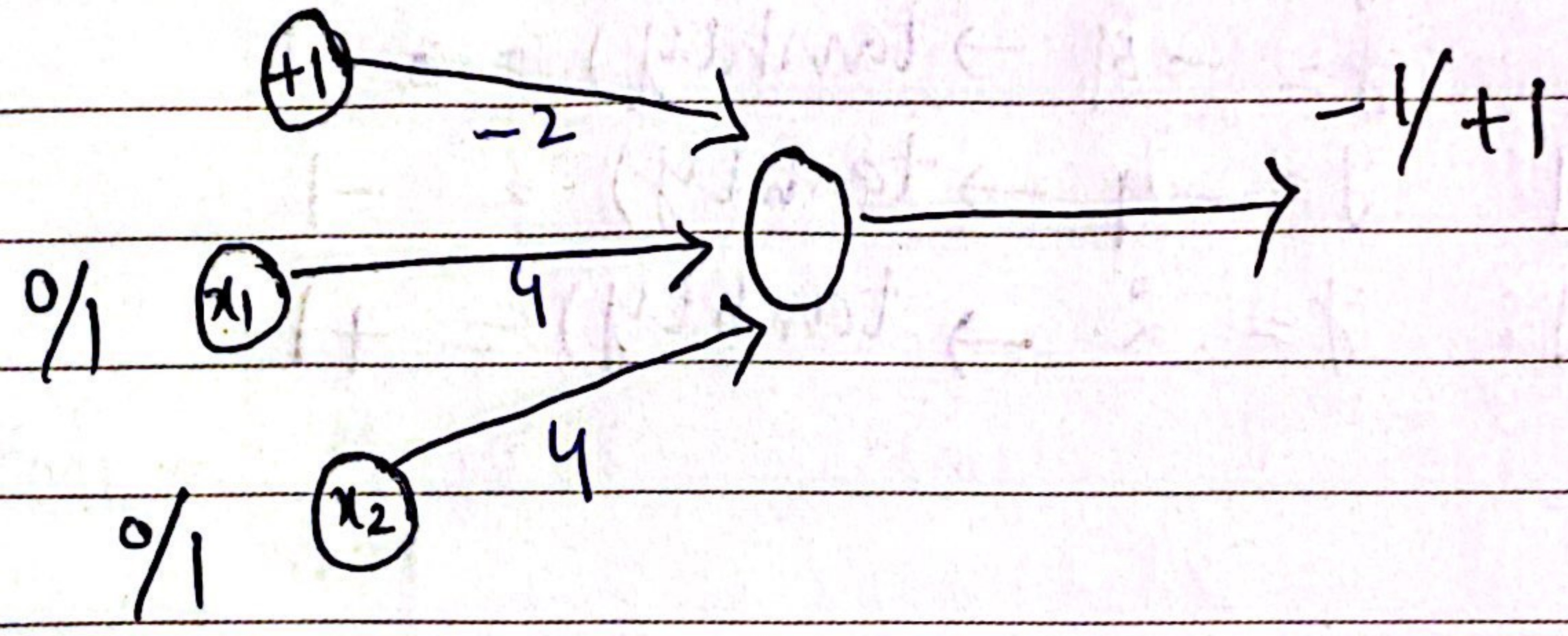
$x_1 = 0, x_2 = 1$ gives $y = 2 \Rightarrow \tanh(y) = 1$

$x_1 = 1, x_2 = 0$ gives $y = 2 \Rightarrow \tanh(y) = 1$

$x_1 = 1, x_2 = 1$ gives $y = 6 \Rightarrow \tanh(y) = 1$

So,

x_1 OR x_2.



Similarly for NOT(x_1 AND x_2):

$$y = w_0 + w_1 x_1 + w_2 x_2 = 6 - 4x_1 + w_2 x_2$$
$$= 6 - 4x_1 - 4x_2$$

Selecting $w_2 = -4$ gives:

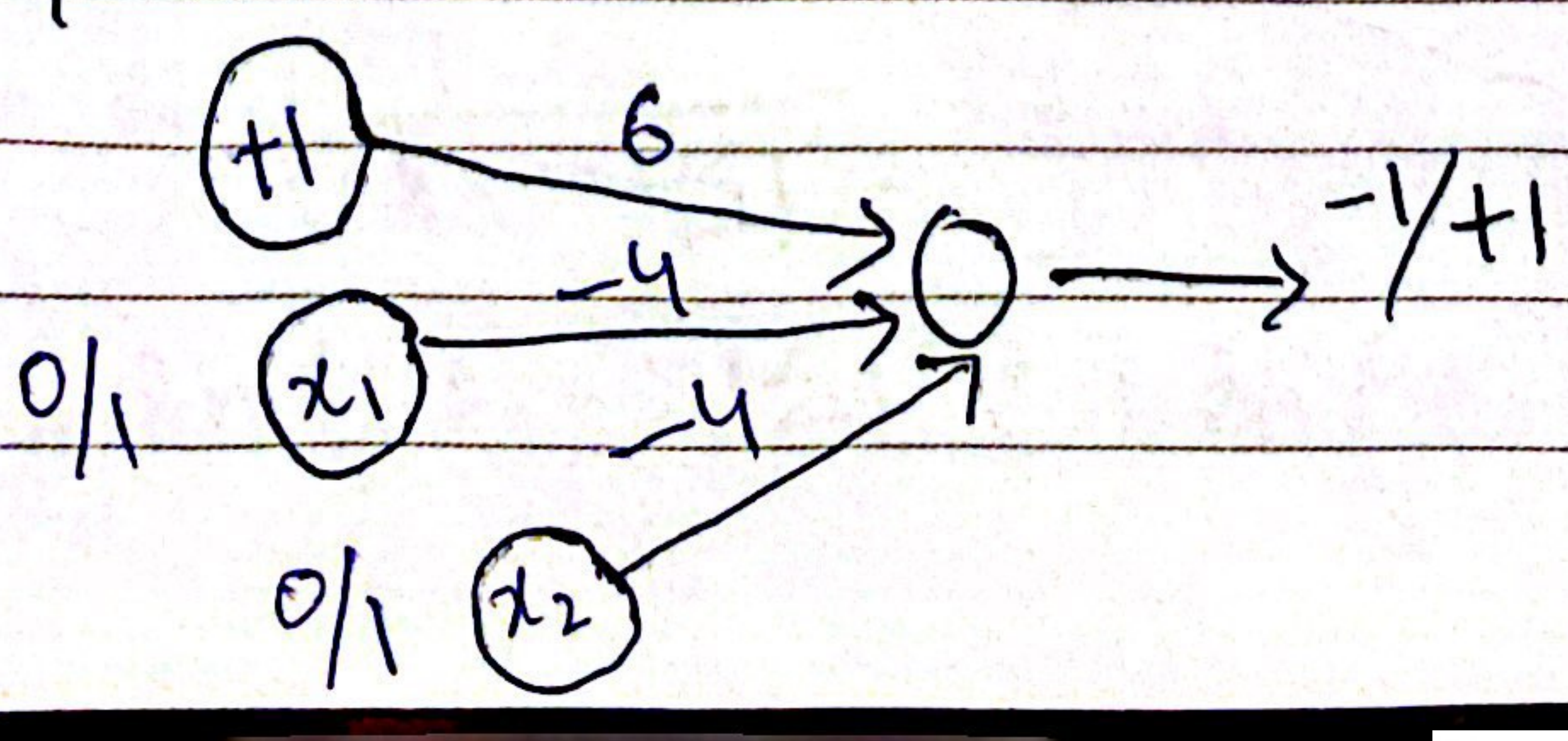
when $x_1 = 0, x_2 = 0$: $y = 6 \rightarrow \tanh(y) = 1$

$x_1 = 0, x_2 = 1$: $y = 2 \rightarrow \tanh(y) = 1$

$x_1 = 1, x_2 = 0$: $y = 2 \rightarrow \tanh(y) = 1$

$x_1 = 1, x_2 = 1$: $y = -2 \rightarrow \tanh(y) = -1$

So,



Designing x_1 AND x_2

$y = w_0 + w_1 x_1 + w_2$

Selecting $w_2 = 2$ gives:

$$y = -2 + 2x_1 + 2x_2$$

Verifying these values for AND gate:

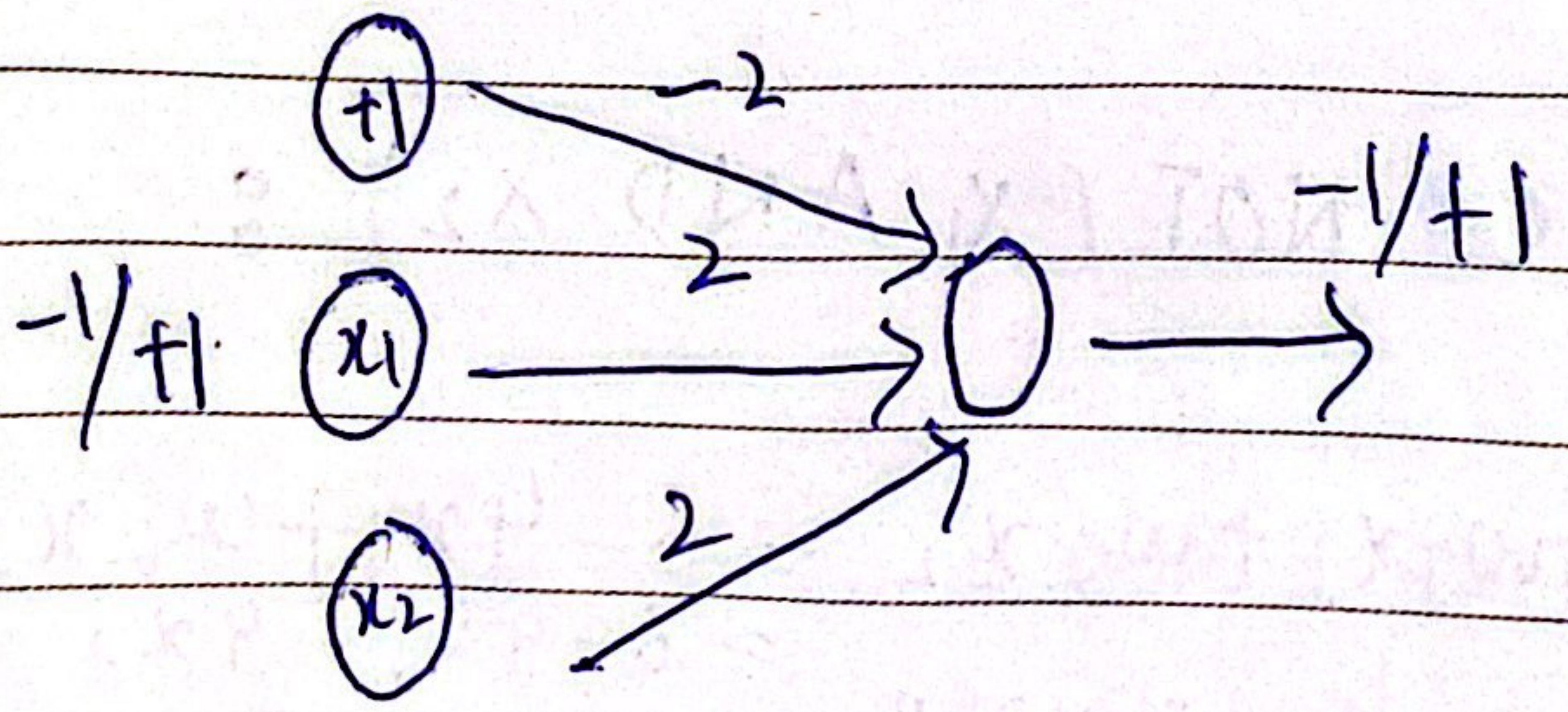
$x_1 = -1, x_2 = -1$: $y = -2 \rightarrow \tanh(y) = -1$

$x_1 = -1, x_2 = 1$: $y = -4 \rightarrow \tanh(y) = -1$

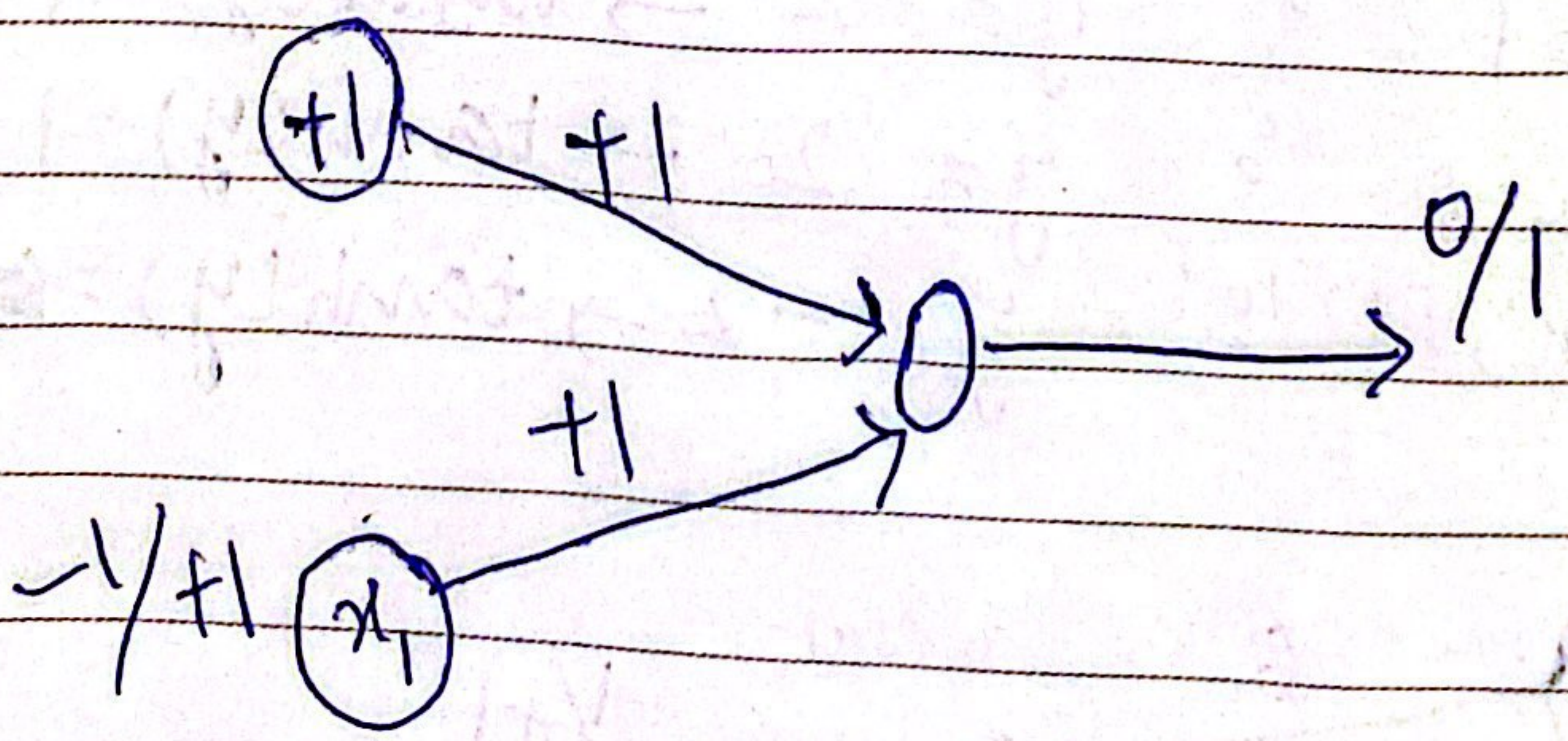
$x_1 = 1, x_2 = -1$: $y = -4 \rightarrow \tanh(y) = -1$

$x_1 = 1, x_2 = 1$: $y = 2 \rightarrow \tanh(y) = +1$

So,



Designing or converting to 0/1



Question NO: 02

likelihood function under conditional distribution for i.i.d dataset is given by:

$$x \in \{(x_1, t_1), \dots, (x_N, t_N)\}$$

$$P(t|x, w) = \mathcal{N}(t|y(x, w), \beta^{-1}I)$$

$$\theta_{ML} = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, w), \beta^{-1}I)$$

Since, gaussian distribution is given by:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}[(x-\mu)^T\Sigma^{-1}(x-\mu)]}$$

so,

$$\ln(\theta_{ML}) = \ln\left[\prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, w), \beta^{-1}I)\right]$$

$$= \sum_{n=1}^{N} \ln \mathcal{N}(t_n|y(x_n, w), \beta^{-1}I)$$

$$= \ln\left(\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\right) + \ln\left(e^{-\frac{1}{2}[(x-\mu)^T\Sigma^{-1}(x-\mu)]}\right)$$

$$= \ln\left((2\pi)^{D/2}|\Sigma|^{1/2}\right)^{-1} + \ln\left(e^{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)}\right)$$

$$= \ln (2\pi)^{-D/2} + \ln (\Sigma^{-1})^{-1} + \ln \left(e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} \right)$$

$$= \underbrace{-\frac{D}{2} \ln (2\pi) + \ln |\Sigma|}_{\text{constant}} - \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \ln$$

$\Rightarrow$

$$\theta_{ML} = \text{constant} - \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)$$

Now, putting values of μ & Σ as per given distribution:

$$\mu = y(x, w), \quad \Sigma = \beta^{-1} I$$

So, $x = t_n$.

$$\underset{x}{\text{argmax}}\ \theta_{ML} = -\frac{1}{2} \left(\sum_{n=1}^{N} (t - y(x_n, w))^T (\beta^{-1} I)^{-1} (t_n - y(x_n, w) \right) + \text{const.}$$

$$= -\frac{\beta I}{2} \sum_{n=1}^{N} (t_n - y(x_n, w))^T (t_n - y(x_n, w))$$

$$= -\frac{\beta}{2} \sum_{n=1}^{N} \| t_n - y(x_n, w) \|^2 + \text{const.} \qquad \text{— (1)}$$

Now, comparing eq. (1) with eq. of min loss squares:

Scanned with CamScanner

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \| y(x_n, w) - t_n \|^2 \qquad (2)$$

Hence, maximizing log-likelihood is equivalent to minimizing sum-of-squares error.

Question NO: 03

Assuming $t_{nK} = 1$ for training example n for class K_n. The likelihood eq. is:

$$P(t_n | x_n, w) = \prod_{n=1}^{N} p(t_{nk} = 1 | x_n, w)$$

$$= \prod_{n=1}^{N} y_{K_n}(x_n, w)$$

$$= \prod_{n=1}^{N} \prod_{K=1}^{K} y_K(x_{n}, w)^{t_{nk}}.$$

where $t_{nk} = 1$ at K_n and $t_{nk} = 0$ otherwise.

So, log-likelihood is:

$$\log p(t_n | x_n, w) = \log \prod_{n=1}^{N} \prod_{k=1}^{K} y_K(x_n, w)^{t_{nk}}$$

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

Scanned with CamScanner

Hence proved that:

$$E(w) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \cdot \ln y_k(x_n, w)$$

Question NO: 04

Since, expression for sigmoid function is:

$$y = \sigma(a) = \frac{1}{1+e^{-a}}$$

Re-scaling & shifting of binary outputs is given by:

$$y = 2\sigma(a) - 1$$

and for binary classifier

$$E(w) = -\sum_{n=1}^{N} \{t_n \ln(y_n) + (1-t_n)\ln(1-y_n)\}$$

Applying inverse transform gives:

$$E(w) = -\sum_{n=1}^{N} \frac{1+t_n}{2} \ln \frac{1+y_n}{2} + \left(1 - \frac{1+t_n}{2}\right) \times \ln\left(1 - \frac{(1+y_n)}{2}\right)$$

$$= \left(-\frac{1}{2} \sum_{n=1}^{N} \{ (1+t_n) \ln(1+y_n) + (1-t_n) \ln(1-y_n) \} \right) + N \ln 2 +$$

Ignoring $N \ln 2$ since, it is independent of w.

Applying linear transformation gives:

$$y = \sigma(a) = \frac{1}{1+e^{-a}}$$

$$= 2\sigma(a) - 1$$

$$= 2\left(\frac{1}{1+e^{-a}}\right) - 1$$

$$= \frac{1-e^{-a}}{1+e^{-a}} = \frac{e^{-a/2} - e^{a/2}}{e^{a/2} + e^{-a/2}}$$

$$= \tanh(a/2)$$

$\tanh(a/2)$ activation function.

Scanned with CamScanner