*****Turning In Late, Using 2 out of 2 Late Days*****
Previously used 0 late days

1: Logical Functions with Neural Networks

(a)

NAND of two binary inputs

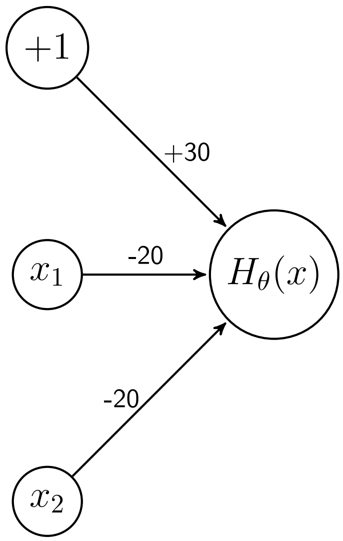


Table 1: NAND Truth Table

X_1	X_2	Z	$H_\theta(x)$
0	0	+30	1
0	1	+10	1
1	0	+10	1
1	1	-10	0

(b)

Parity of three binary inputs: found by cascading X_1 XOR X_2 into XNOR X_3.

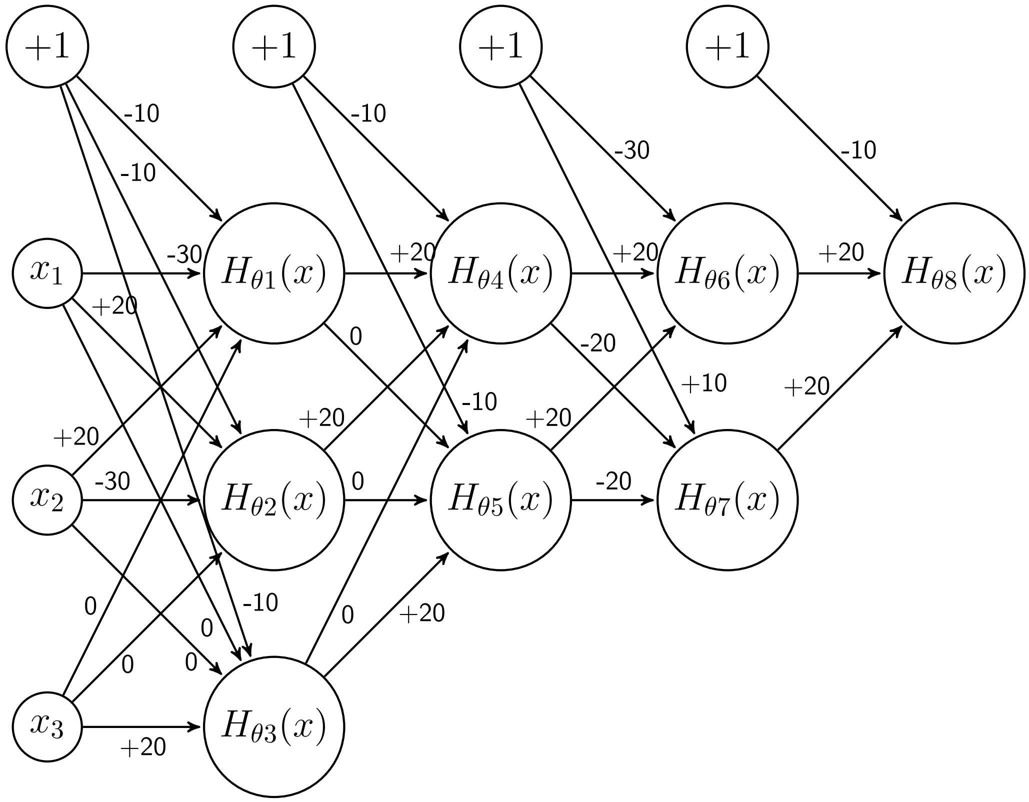
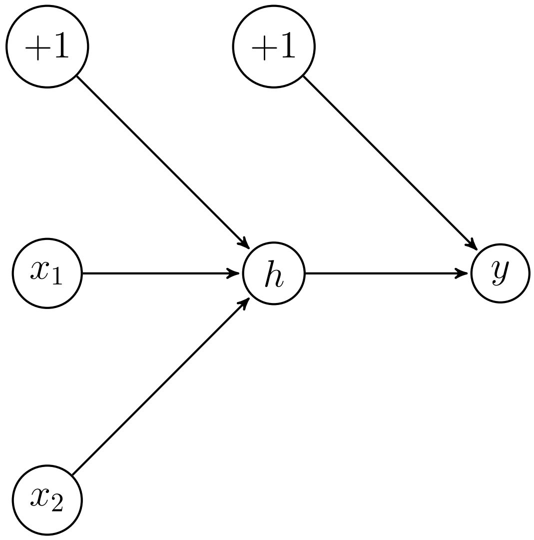


Table 2: Three Binary Input Truth Table

X_1	X_2	X_3	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	$H_{\theta 8}(x)$
0	0	0	-10	-10	-10	-10	-10	-30	+10	+10	1
1	0	0	-40	+10	-10	+10	-10	-10	-10	-10	0
0	1	0	+10	-40	-10	+10	-10	-10	-10	-10	0
0	0	1	-10	-10	+10	-10	+10	-10	-10	-10	0
1	1	0	-20	-20	-10	-10	-10	-30	+10	+10	1
1	0	1	-40	+10	+10	+10	+10	+10	-30	+10	1
0	1	1	+10	-40	+10	+10	+10	+10	-30	+10	1
1	1	1	-20	-20	+10	-10	+10	-10	-10	-10	0

2: Back Propagation with Momentum



Epoch 1, n 1

$$\Delta = [0, 0, 0, 0, 0]$$
$$a^{(1)} = [1, 1, 0]$$
$$a^{(2)} = g\left(([0.1, 0.1, 0.1]\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = 0.5498$$
$$a^{(3)} = g\left(([0.1, 0.1]\begin{bmatrix} 1 \\ 0.5498 \end{bmatrix}\right) = 0.5387$$
$$\delta^{(3)} = 0.5387 - 1 = -0.4613$$
$$\delta^{(2)} = 0.1 * -0.4613 * 0.5498 * (1 - 0.5498) = -0.0114$$
$$\Delta = [-.0114, 0.0114, 0, -.2536, -.4613]$$

Epoch 1, n 2

$$\Delta = [-.0114, 0.0114, 0, -.2536, -.4613]$$
$$a^{(1)} = [1, 0, 1]$$
$$a^{(2)} = g\left(([0.1, 0.1, 0.1]\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = 0.5498$$
$$a^{(3)} = g\left(([0.1, 0.1]\begin{bmatrix} 1 \\ 0.5498 \end{bmatrix}\right) = 0.5387$$
$$\delta^{(3)} = 0.5387 - 0 = 0.5387$$
$$\delta^{(2)} = 0.1 * -0.5387 * 0.5498 * (1 - 0.5498) = 0.0133$$
$$\Delta = [0.0019, -0.0114, 0.0133, 0.0426, 0.0774]$$

Theta update

$$\Delta = [0.0019, -0.0114, 0.0133, 0.0426, 0.0774]/2$$
$$= [0.001, -0.0057, 0.0056, 0.0213, 0.0387]$$
$$\Theta = [0.099, 0.1057, 0.0944, 0.0787, 0.0613]$$

Epoch 2, n 1

$$\Delta = [0, 0, 0, 0, 0]$$
$$a^{(1)} = [1, 1, 0]$$
$$a^{(2)} = g\left(([0.099, 0.1057, 0.0944] \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right) = 0.551$$
$$a^{(3)} = g\left(([0.0787, 0.0613] \begin{bmatrix} 1 \\ 0.551 \end{bmatrix} \right) = 0.5281$$
$$\delta^{(3)} = 0.5281 - 1 = -0.4719$$
$$\delta^{(2)} = 0.0613 * -0.4719 * 0.551 * (1 - 0.551) = -0.0072$$
$$\Delta = [-0.0072, -0.0072, 0, -0.4719, -0.26]$$

Epoch 2, n 2

$$\Delta = [-0.0072, -0.0072, 0, -0.4719, -0.26]$$
$$a^{(1)} = [1, 0, 1]$$
$$a^{(2)} = g\left(([0.099, 0.1057, 0.0944] \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = 0.5482$$
$$a^{(3)} = g\left(([0.0787, 0.0613] \begin{bmatrix} 1 \\ 0.5482 \end{bmatrix} \right) = 0.528$$
$$\delta^{(3)} = 0.528 - 0 = 0.528$$
$$\delta^{(2)} = 0.0613 * -0.528 * 0.5482 * (1 - 0.5482) = 0.008$$
$$\Delta = [0.0008, -0.0072, 0.008, 0.0561, 0.0294]$$

Theta update

$$\Delta = [0.0008, -0.0072, 0.008, 0.0561, 0.0294]/2$$
$$= [0.0004, -0.0036, 0.004, 0.028, 0.0147]$$
$$\Theta = [0.0987, 0.1083, 0.0916, 0.0645, 0.0464]$$

3: TANH Neural Networks

(a)

The advantage of using a tanh instead of a sigmoid function in a neural network is that we have steeper gradients. Thus when we initialize with small weights, or near the origin, with a sigmoid function it will be much slower to converge to a solution. Also, when we approach relatively flat spots in the gradient, the tanh function will continue to perform where the sigmoid function may slow or struggle to converge at a reasonable rate.

(b)

$$y_k(x,\theta) = \sigma\left(\sum_{j=1}^{M}\theta_{jk}^{(2)}\sigma\left(\sum_{i=1}^{d}\theta_{ij}^{(1)}x_i + \theta_{0j}^{(1)}\right) + \theta_{0k}^{(32)}\right) \tag{1}$$

$$y_k(x,\theta) = \sigma\left(\alpha\sigma\left(z\right)\right) \tag{2}$$

$$y_k(x,\theta) = tanh\left(\alpha tanh\left(z\right)\right) \tag{3}$$

$$\sigma = \frac{1}{1+e^{-z}} \tag{4}$$

$$tanh = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{5}$$

$$(2) = \frac{1}{1+e^{-\alpha\frac{1}{1+e^{-z}}}}$$

$$(3) = \frac{e^{\alpha\frac{e^z-e^{-z}}{e^z+e^{-z}}} - e^{-\alpha\frac{e^z-e^{-z}}{e^z+e^{-z}}}}{e^{\alpha\frac{e^z-e^{-z}}{e^z+e^{-z}}} + e^{-\alpha\frac{e^z-e^{-z}}{e^z+e^{-z}}}}$$

$$tanh = \sigma * 2 - 1$$

$$equation(3) = equation(2) * 2 - 1$$