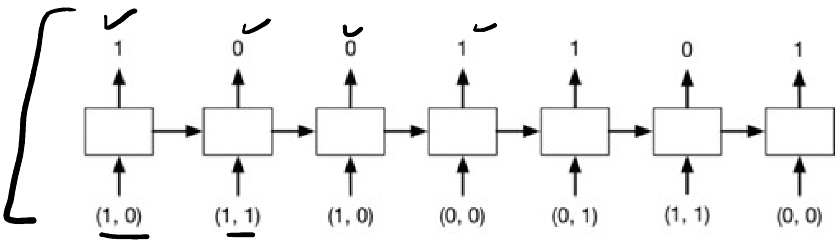
1. **Binary Addition** [4pts] In this problem, you will implement a recurrent neural network which implements binary addition. The inputs are given as binary sequences, starting with the *least* significant binary digit. (It is easier to start from the least significant bit, just like how you did addition in grade school.) The sequences will be padded with at least one zero on the end. For instance, the problem

$$100111 + 110010 = 1011001$$

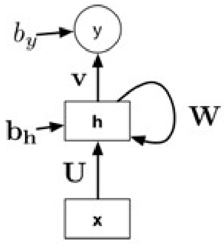
would be represented as:

- **Input 1:** 1, 1, 1, 0, 0, 1, 0
- **Input 2:** 0, 1, 0, 0, 1, 1, 0
- **Correct output:** 1, 0, 0, 1, 1, 0, 1

There are two input units corresponding to the two inputs, and one output unit. Therefore, the pattern of inputs and outputs for this example would be:




Design the weights and biases for an RNN which has two input units, three hidden units, and one output unit, which implements binary addition. All of the units use the hard threshold activation function. In particular, specify weight matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, bias vector $\mathbf{b_h}$, and scalar bias $b_y$ for the following architecture:




*Hint:* In the grade school algorithm, you add up the values in each column, including the carry. Have one of your hidden units activate if the sum is at least 1, the second one if it is at least 2, and the third one if it is 3.

We have

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Input 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Input 2 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |

| Input 2 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
|---|---|---|---|---|---|---|---|
| output | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| Carry | 0 | 1 | 1 | 0 | 0 | 1 | 0 |

We can model all possible combinations below.

| Carry | $x_1(t)$ | $x_2(t)$ | $y_t$ |
|---|---|---|---|
| 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 1 |
| 0 | 1 | 1 | 0 + 1 [Carry] = 1 |
| 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 + 1 [Carry] = 1 |
| 1 | 1 | 0 | 0 + 1 [Carry] = 1 |
| 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 + 1 [Carry] = 0 |

Let NC denote No Carry.
C denote Carry

We get 4 combinations -
1) NC & print 1
2) NC & print 0
3) C & print 1
4) C & print 0

---

We can draw possible combinations that will allow us to move from one state to another

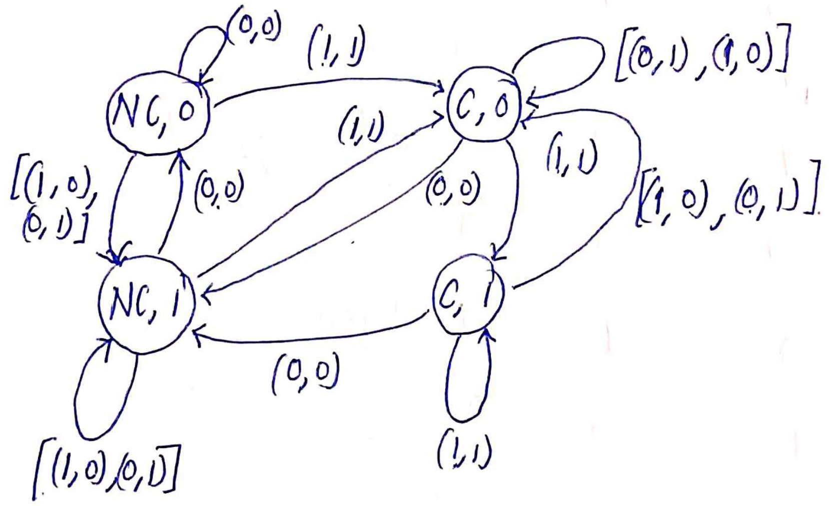
$(0,0) \to 0$

We can draw possible combinations ...
allow us to move from one state to another

$(0,0) \rightarrow 0$
$(0,0) \rightarrow$
$(0,1) \rightarrow 1$
$(1,0) \rightarrow 1$
$(1,1) \rightarrow 0, 1$
$(1)$ $((1, 0))$
$(0,1)$

| From \ To | (NC, 0) | (NC, 1) | (C, 0) | (C, 1) |
|---|---|---|---|---|
| (NC, 0) | [0, 0] | [1, 0], [0, 1] | [1, 1] | – |
| (NC, 1) | [0, 0] | [1, 0], [0, 1] | [1, 1] | – |
| (C, 0) | – | [0, 0] | [0, 1], [1, 0] | [1, 1] |
| (C, 1) | – | [0, 0] | [1, 0], [0, 1] | [1, 1] |

We can draw the table in our usual state transition as below.




Let us assign weights to our states as below.

$(NC, 0) = (0, 0, 0)$
$(NC, 1) = (0, 1, 0)$
$(C, 0) = (1, 1, 1)$
$(C, 1) = (0, 1, 1)$

We will complete the iteration table as below –

| $h_1(t-1)$ | $h_2(t-1)$ | $h_3(t-1)$ | $x_1(t)$ | $x_2(t)$ | $y_t$ | $h_1(t)$ | $h_2(t)$ | $h_3(t)$ |
|---|---|---|---|---|---|---|---|---|

| $h_1(t-1)$ | $h_2(t-1)$ | $h_3(t-1)$ | $x_1(t)$ | $x_2(t)$ | $y_t$ | $h_1(t)$ | $h_2(t)$ | $h_3(t)$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |

Comments: At time $t = 1$, NC1 → NC1

$t = 2$, NC1 → CO

$t = 3$, CO → CO

$t = 4$, CO → NC1

$t = 5$, NC1 → NC1

$t = 6$, NC1 → CO

$t = 7$, CO → NC1

We will express $y_t$ in form of $h_1(t)$, $h_2(t)$ & $h_3(t)$.

Let us consider equation.

$$a_t = 1 - h_1(t) + h_2(t) + (1 - h_3(t)) + b$$

We have

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $a_t$ | $3+b$ | $1+b$ | $1+b$ | $3+b$ | $3+b$ | $1+b$ | $3+b$ |
| $y_t$ | 1 | 0 | 0 | 1 | 1 | 0 | 1 |

We can have a rule where

$$y_t = \begin{cases} 1 & \text{if } a_t \geq 0 \\ 0 & \text{else.} \end{cases}$$

For $a_t > 0$ we must have $3+b > 0$.

$$\Rightarrow b > -3$$

$$\Rightarrow \boxed{b = -2.5}$$

On doing so

$$y_t = \text{step}(a_t)$$

$$= \text{step}\left(1 - h_1 + h_2 + 1 - h_3 + (-2.5)\right)$$

$$= \text{step}\left(-h_1 + h_2 - h_3 - \underline{0.5}\right)$$

$$= \text{step}\left(\underline{V}\, h_t - b_y\right)$$

where $V = (-1 \quad 1 \quad -1)$

$$h_t = \begin{pmatrix} h_1(t) \\ h_2(t) \\ h_3(t) \end{pmatrix}$$

$$b_y = 0.5$$

We will consider $h_1(t-1)$, $h_2(t-1)$, $h_3(t-1)$, $x_1(t)$ & $x_2(t)$ & represent $h_1(t)$ as a step function

We see $h_1(t)$ is dependent on $h_3(t-1)$, $x_1(t)$, $x_2(t)$

We write $\cancel{h_1}^{d}(t) = 0 \cdot h_1(t-1) + 0 \cdot h_2(t-1) + h_3(t-1) + x_1(t) + x_2(t) + b$

$$\Rightarrow \boxed{d(t) = h_3(t-1) + x_1(t) + x_2(t) + b}$$

$$\Rightarrow \boxed{d(t) = h_3(t-1) + x_1(t) + x_2(t) + b.}$$

We can get

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $d(t)$ | $1+b$ | $2+b$ | $2+b$ | $1+b$ | $1+b$ | $2+b$ | $1+b$ |
| $h_1(t)$ | 0 | 1 | 1 | 0 | 0 | 1 | 0 |

We say $h_1(t) = \begin{cases} 1 & \text{if } d(t) > 0 \\ 0 & \text{else.} \end{cases}$

For $d(t) > 0$ such that $h_1(t) = 1$, we must have

$$2+b > 0$$
$$\Rightarrow b > -2$$

We get $\boxed{b = -1.5}$

Hence $h_1(t) = \text{step}\left(h_3(t-1) + x_1(t) + x_2(t) - 1.5\right)$

We will get exactly the same step function for $h_3(t)$

$\underline{W} h_{t-1} + U x_t + b$

Also $h_2(t) = 1$ always

We can say $h_2(t) = \text{step}\left(h_2(t-1) - 0.5 + x_1(t) + x_2(t)\right)$

Overall

$$h(t) = \text{step}\left(\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} h_1(t-1) \\ h_2(t-1) \\ h_3(t-1) \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} -1.5 \\ -0.5 \\ -1.5 \end{pmatrix}\right)$$

$$\boxed{h_t = \text{step}\left(W h_{t-1} + U x_t - b_h\right)}$$

$$\boxed{h_t = \text{step}\left(\underline{W}\, \underbrace{h_{t-1}} + \underline{U}\, \underbrace{x_t} - \underbrace{b_h}\right)}$$

Here; $W = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$U = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$ & $b_h = \begin{pmatrix} 1.5 \\ 0.5 \\ 1.5 \end{pmatrix}$