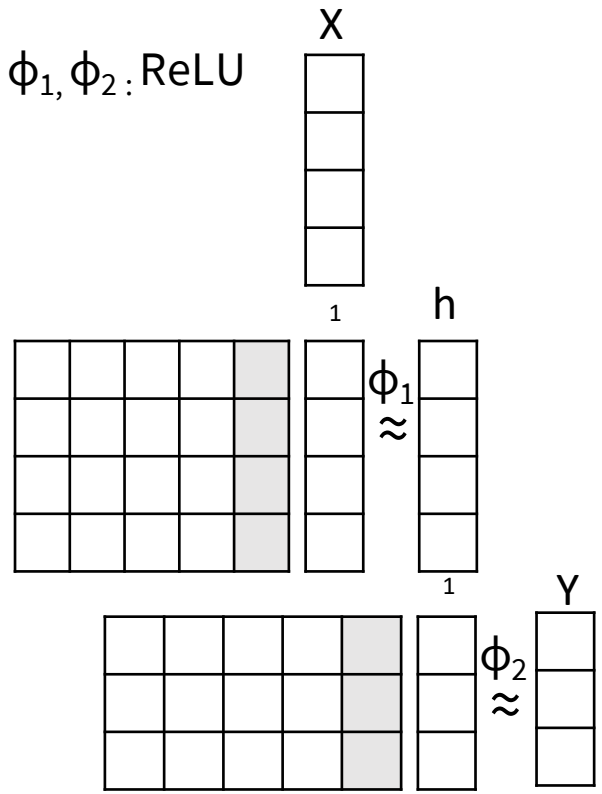
RNN

Name	Dharini Baskaran
Identity Key	dhba5060

	Level	Completed
	Beginner	16
	Intermediate	4
	Advanced	
	Expert	

Goal	
4722	16
5722	18
Total Completed	
20	

MLP Parameter Sizes



size(W_1) = 4×4

size(b_1) = 4×1

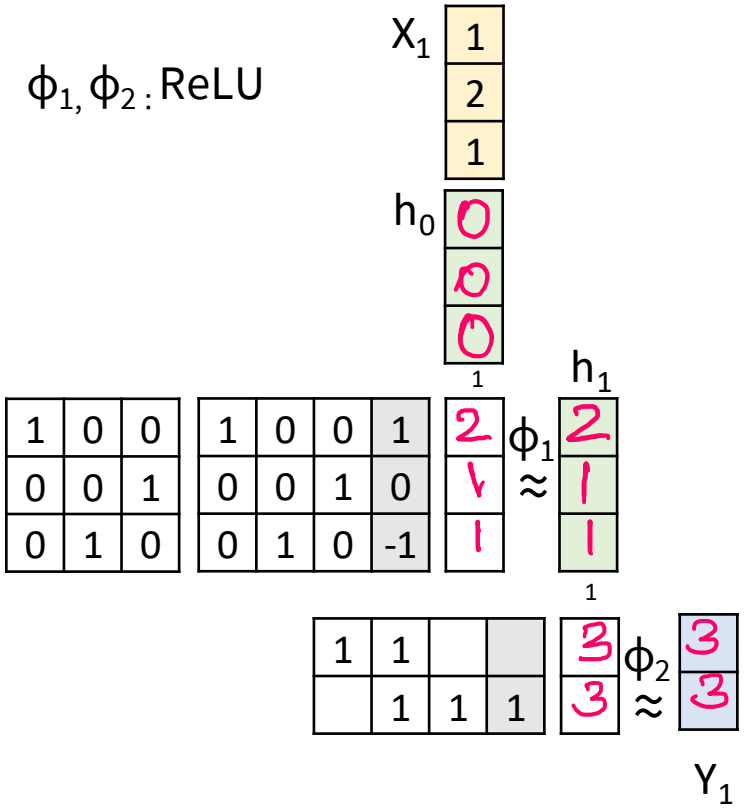
size(W_2) = 3×4

size(b_2) = 3×1

$$Y = \phi_2(\mathrm{W}_2 \cdot \phi_1(W_1 \cdot X + b_1) + b_2)$$

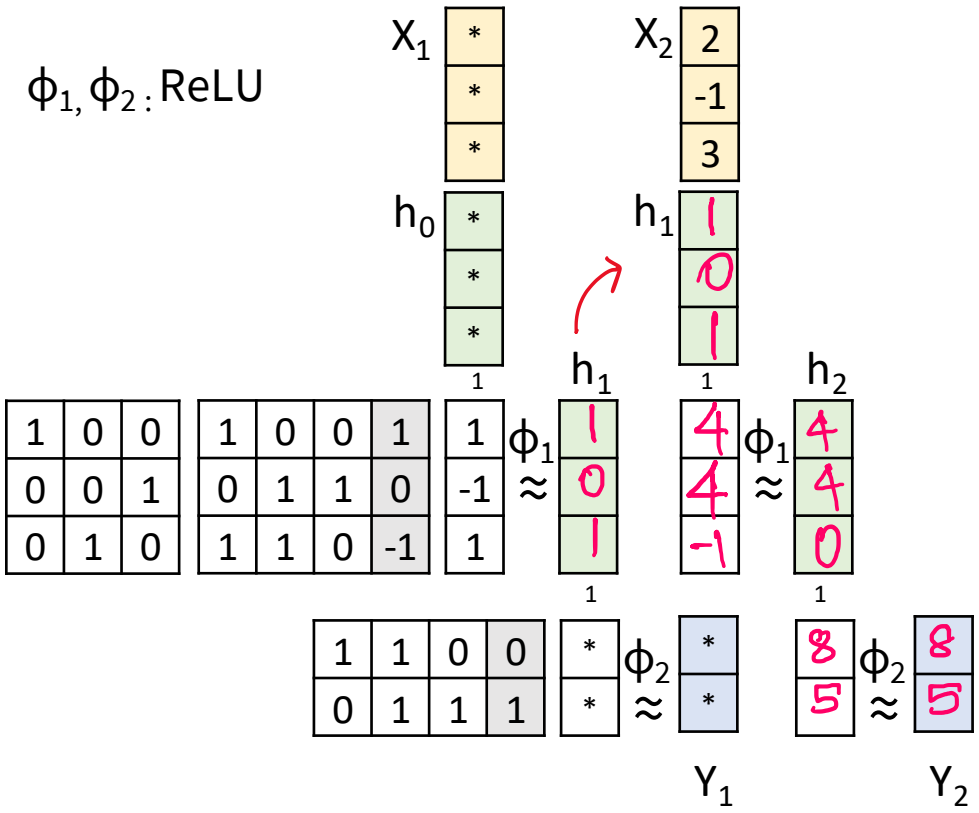
2 Calculate an RNN (t = 1)

This activity is standalone, not dependent on other activities.



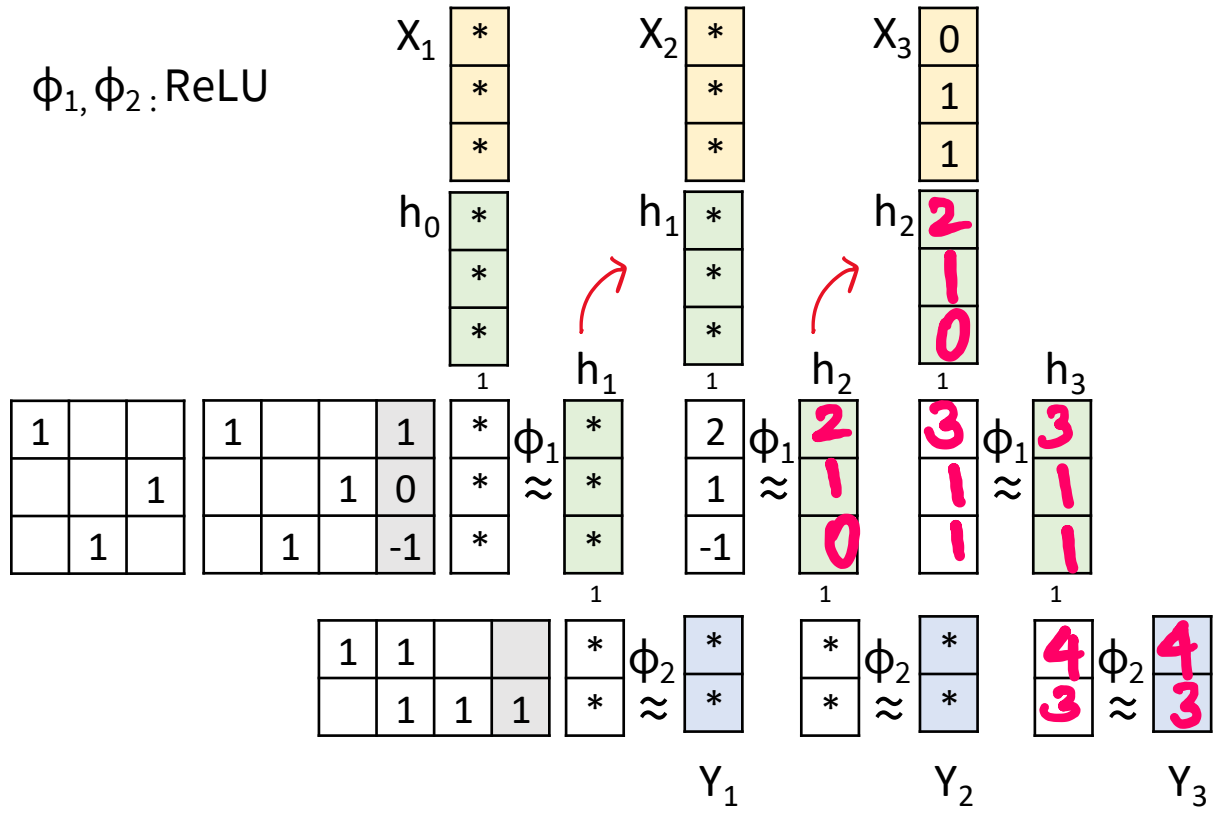
3 Calculate an RNN (t = 2)

This activity is standalone, not dependent on other activities.

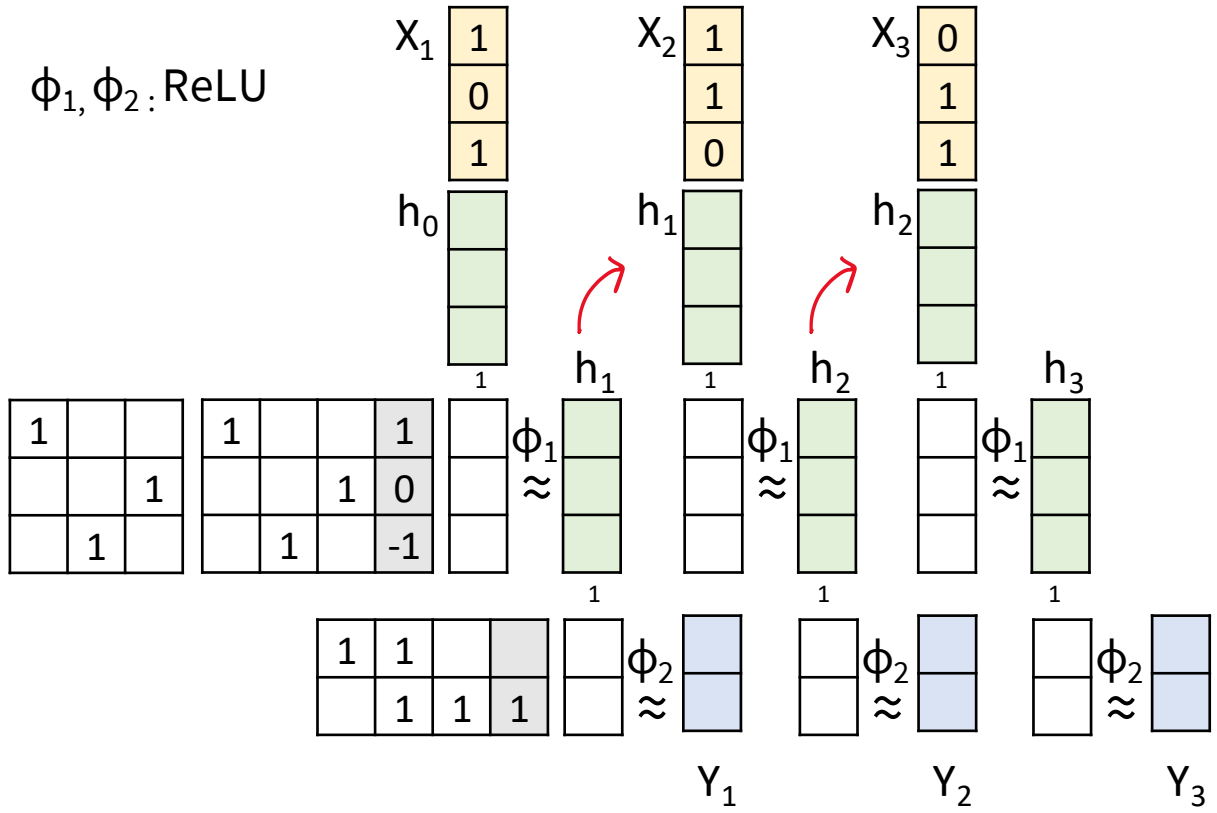


4 Calculate an RNN (t = 3)

This activity is standalone, not dependent on the previous one.



5 Counting Parameters (small)



size(X_t) = 3

size(h_t) = 3

size(y_t) = 2

size(W_{hx}) = 3x3

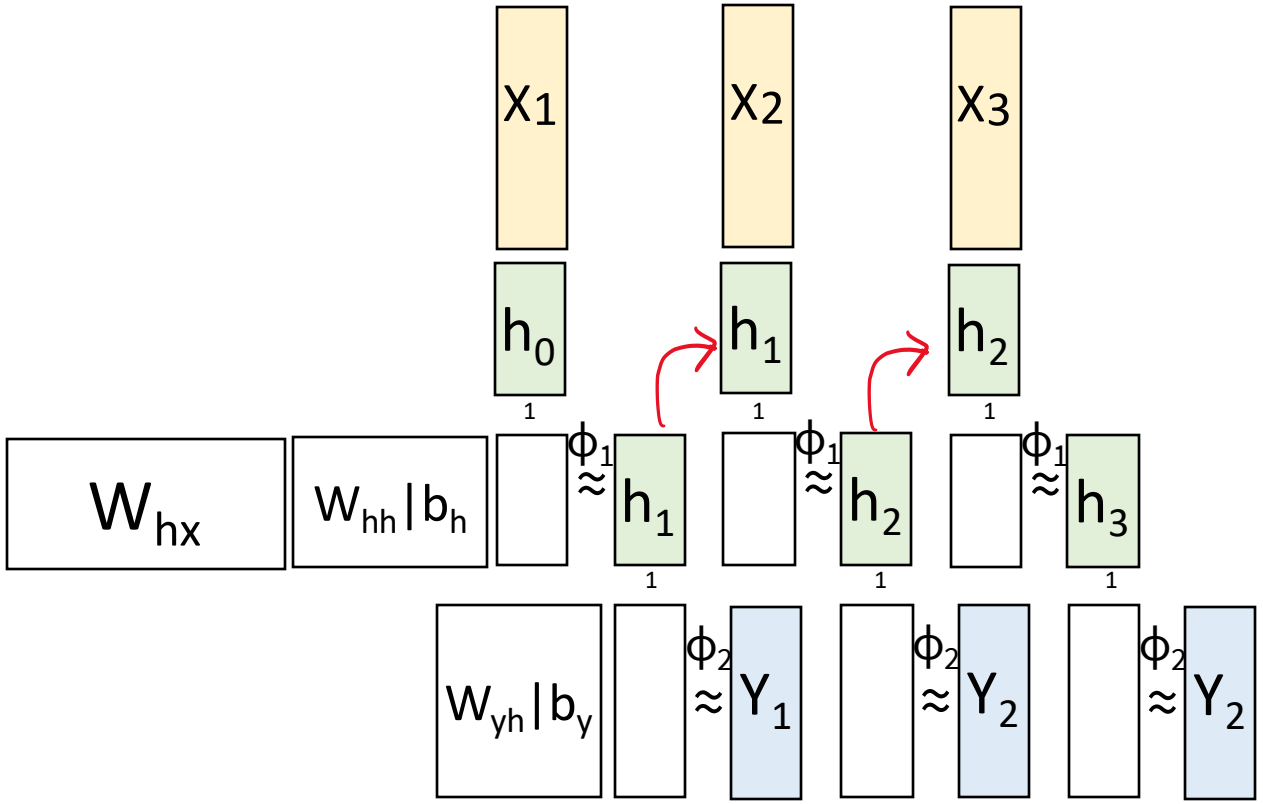
size(b_h) = 3x1

size(W_{hh}) = 3x3

size(W_{yh}) = 2x3

size(b_y) = 2x1

6 Counting Parameters (large)



size(X_t) = 512 x 1

size(h_t) = 128 x 1

size(Y_t) = 256 x 1

size(W_{hx}) = 128 x 512

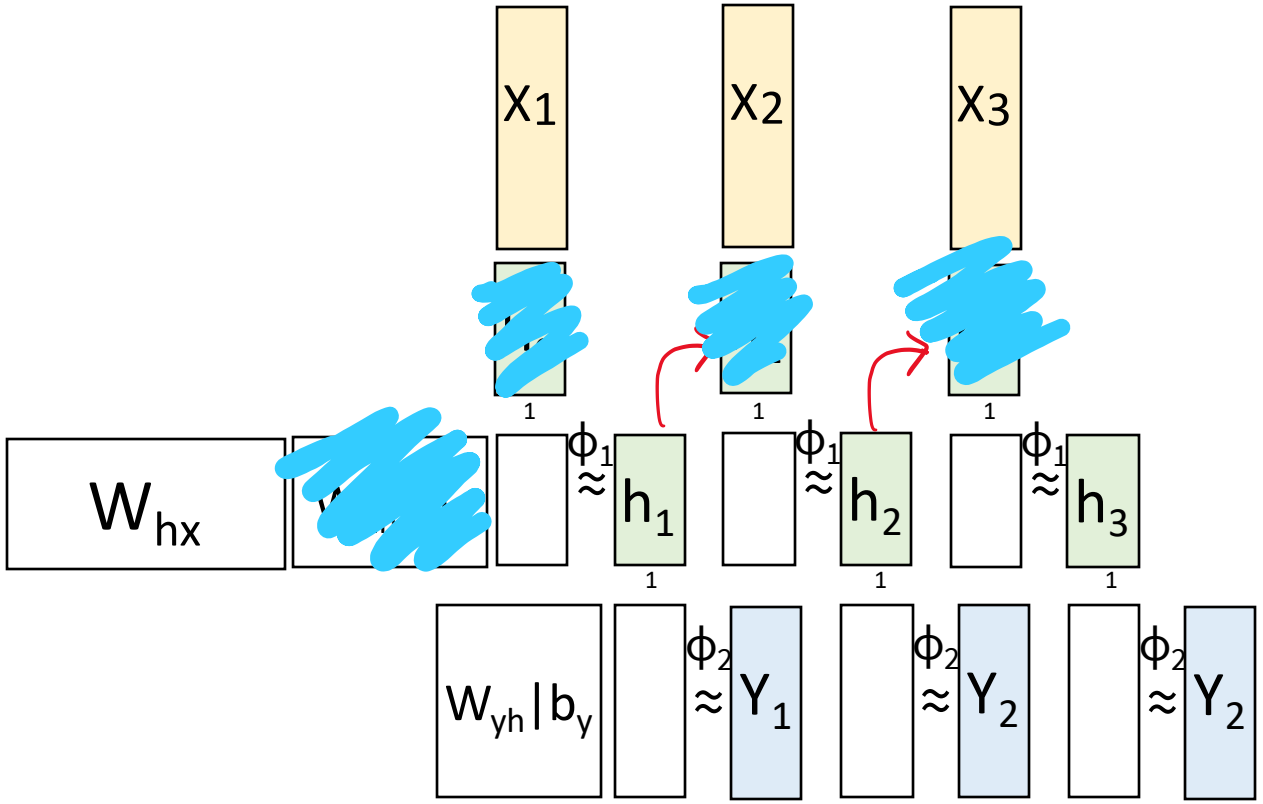
size(b_h) = 128 x 1

size(W_{hh}) = 128 x 128

size(W_{yh}) = 256 x 128

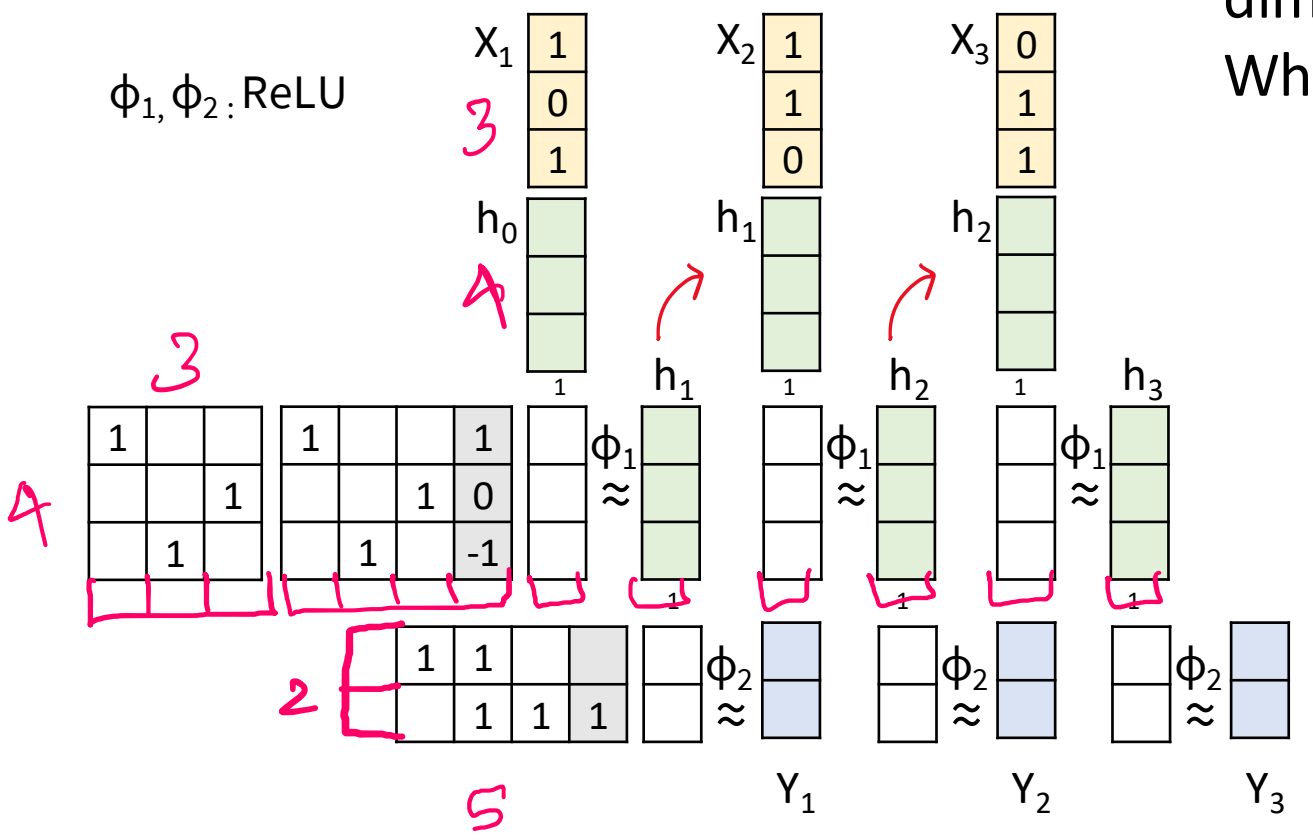
size(b_y) = 256 x 1

7 RNN → MLP



Delete the components to turn this architecture into a regular MLP without recurring hidden states.

Adding Parameters



Suppose we increase the hidden state's dimension by 1.
What would be the new parameter sizes?

size(W_{hx}) = 4x3

size(b_h) = 4x1

size(W_{hh}) = 4x4

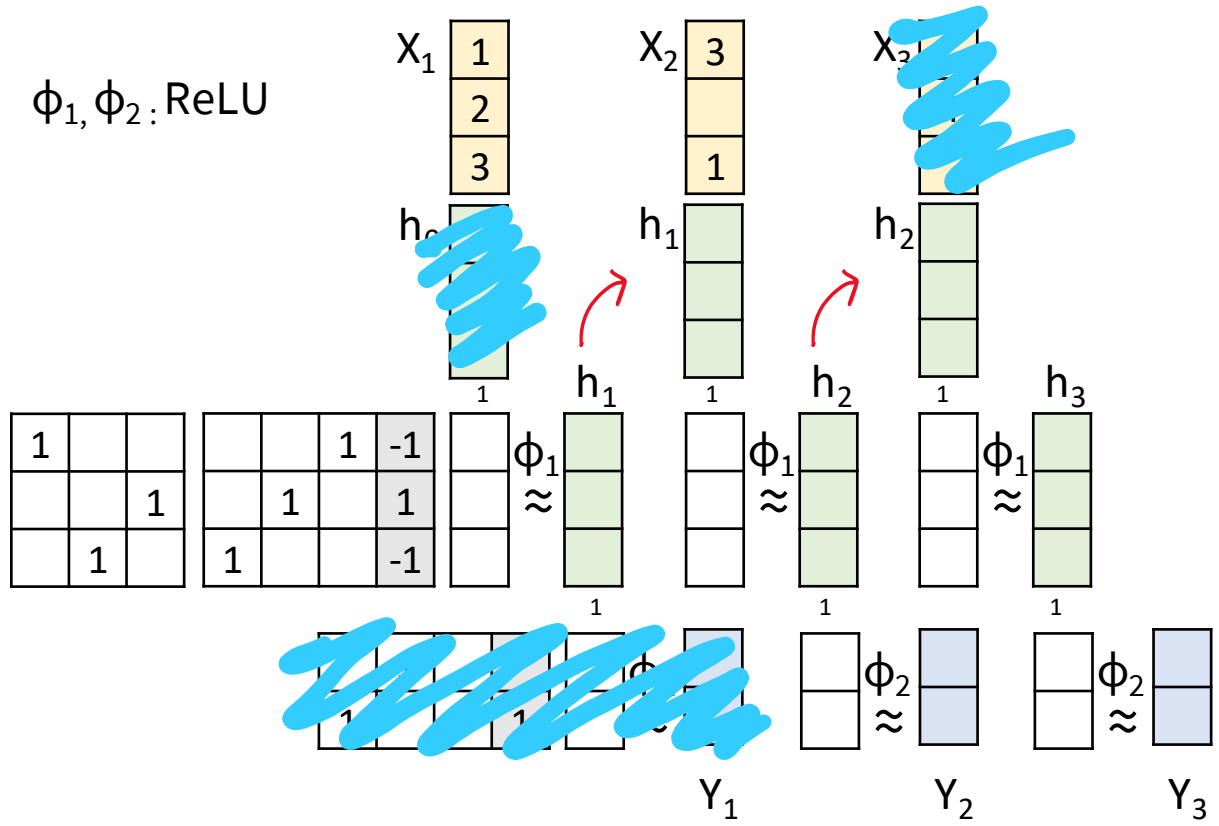
size(W_{yh}) = 2x4

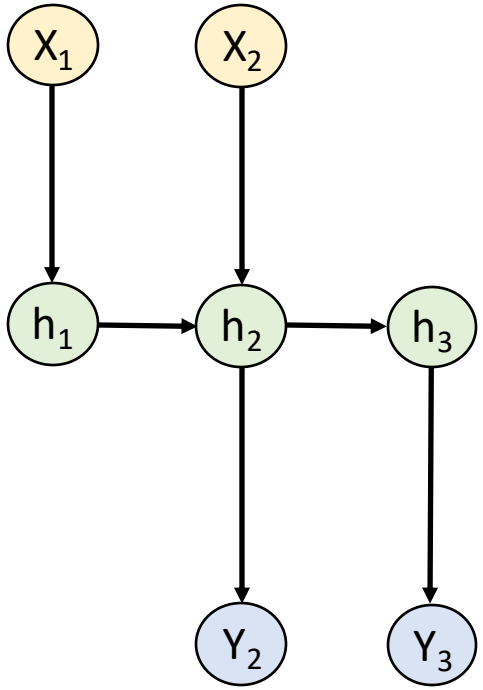
size(b_y) = 2x1

# of New Parameters = 13

(Hint: You can try to draw the extra cells as visual aid)

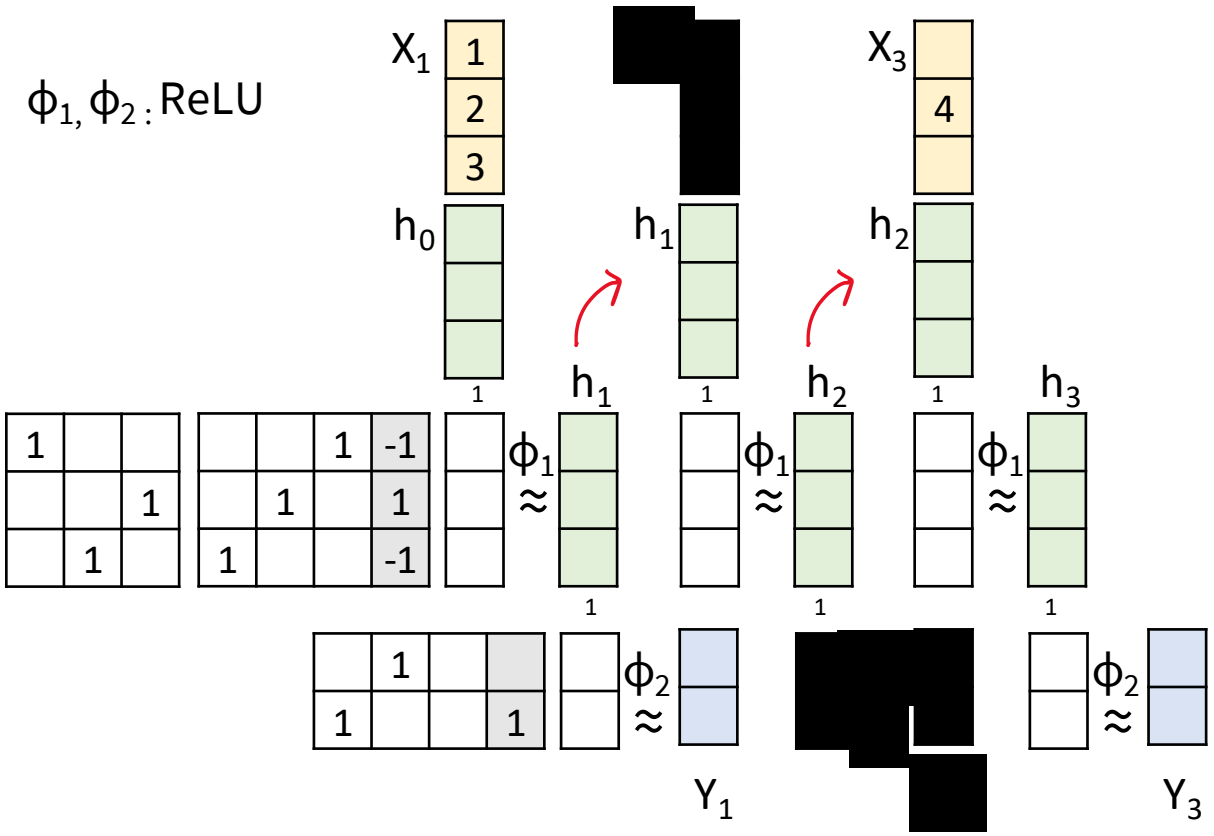
Graph → Matrix



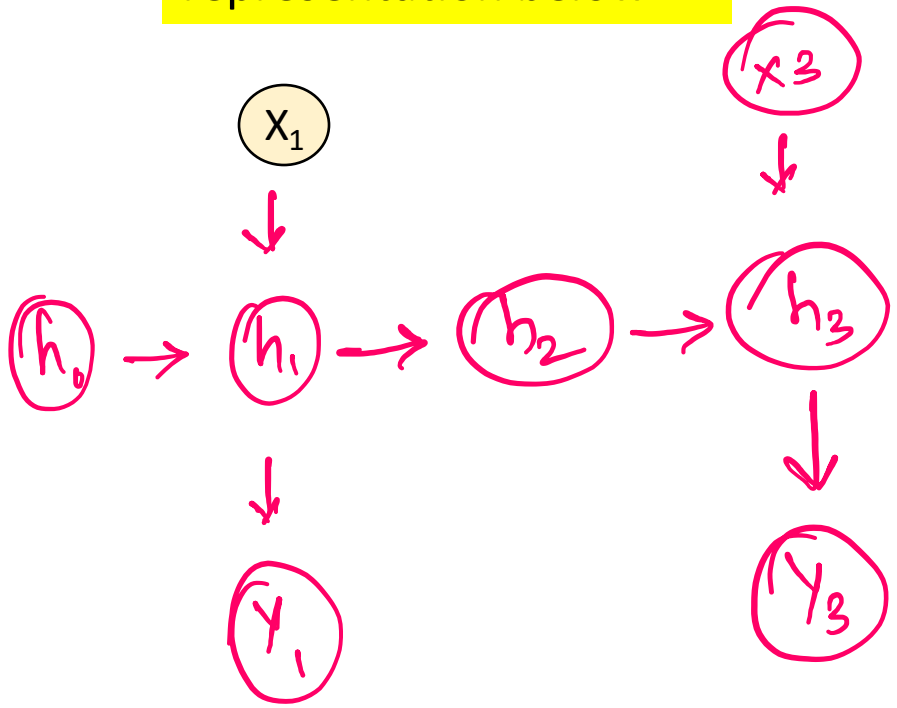


Delete the unwanted parts to match the graph to the right.

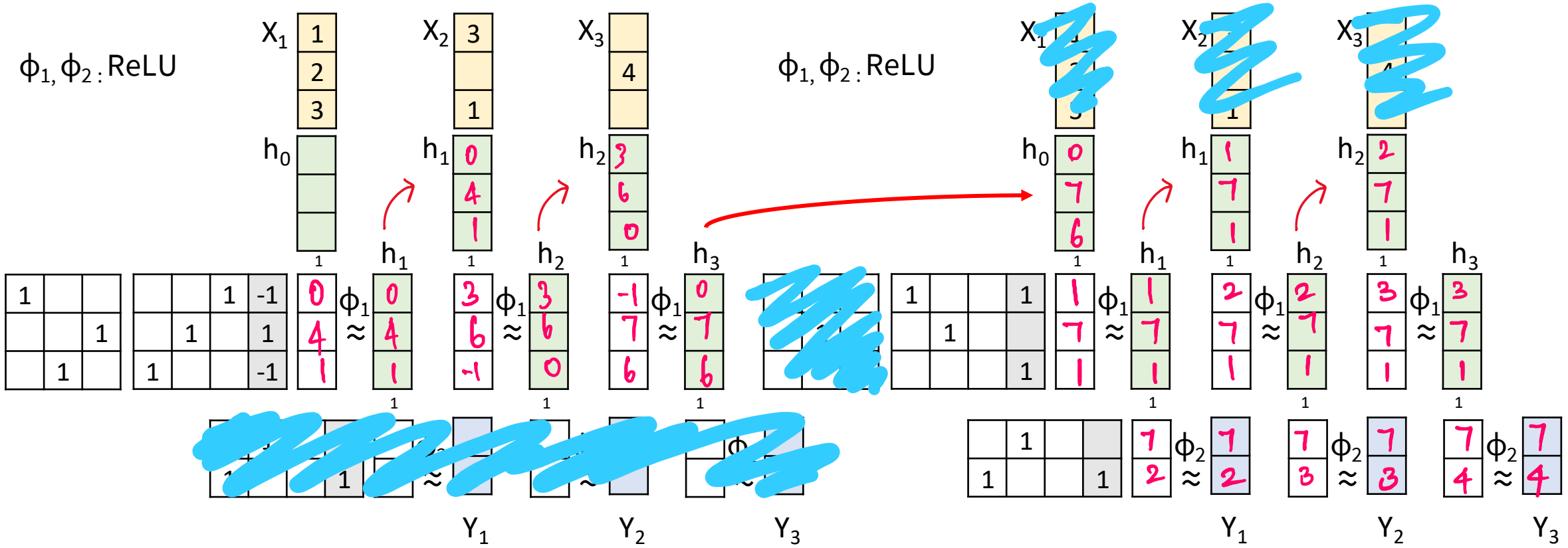
RNN Matrix → Graph



Complete the graphical representation below

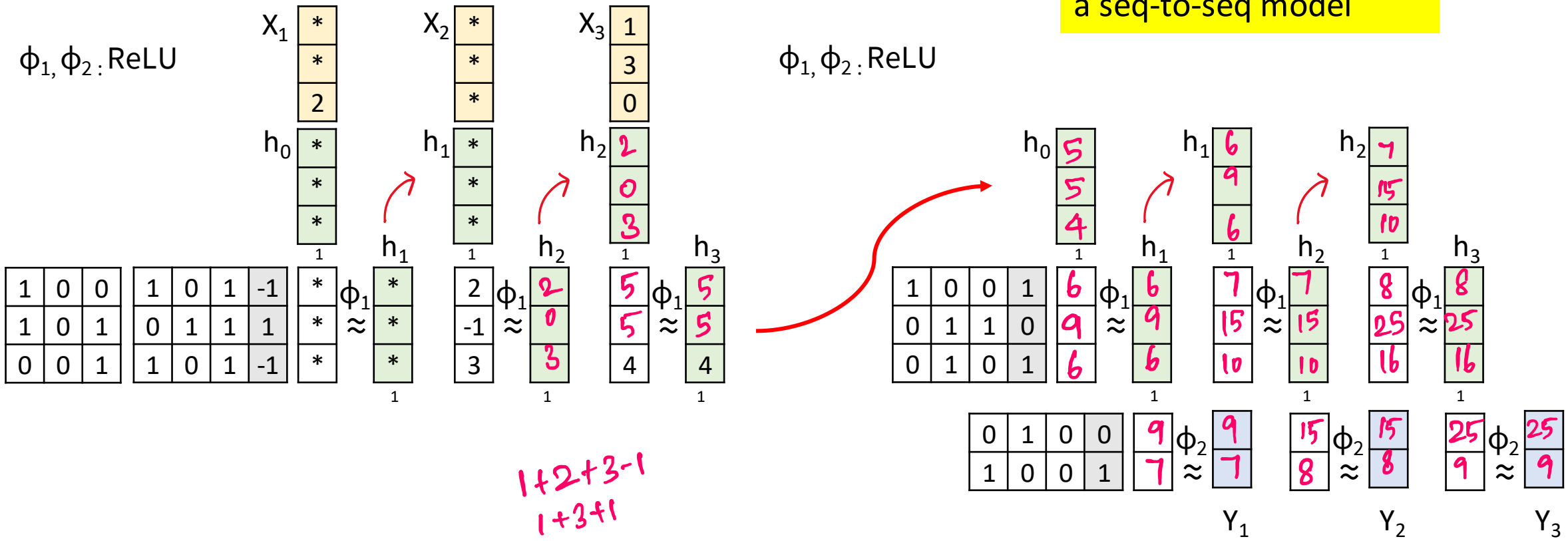


Sequence to Sequence



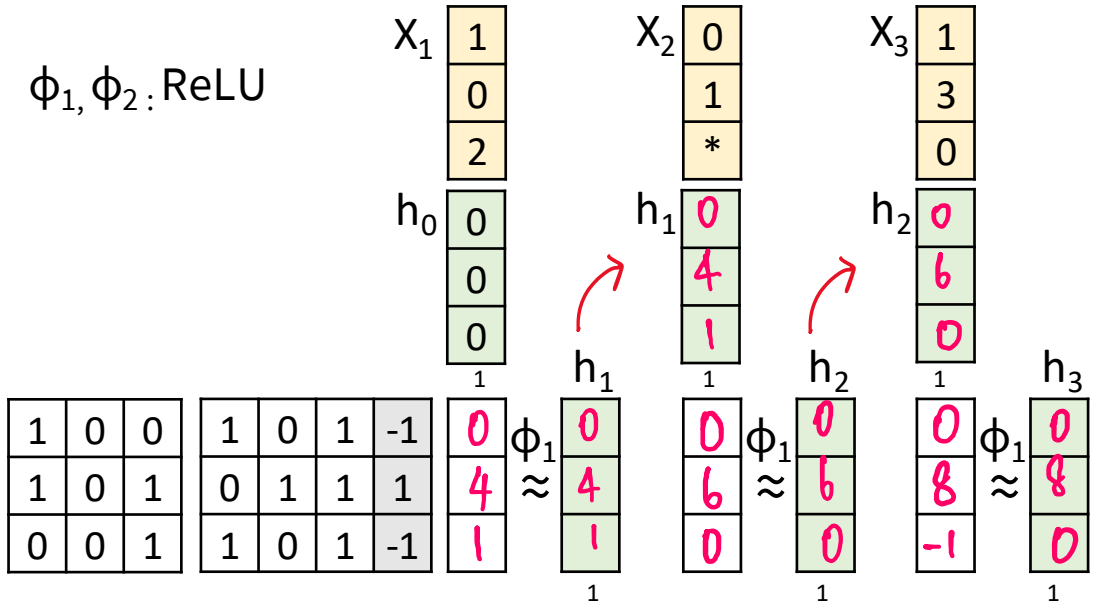
Sequence to Sequence

Remove unwanted components to make it a seq-to-seq model



Calculate an Encoder

ϕ_1, ϕ_2 : ReLU



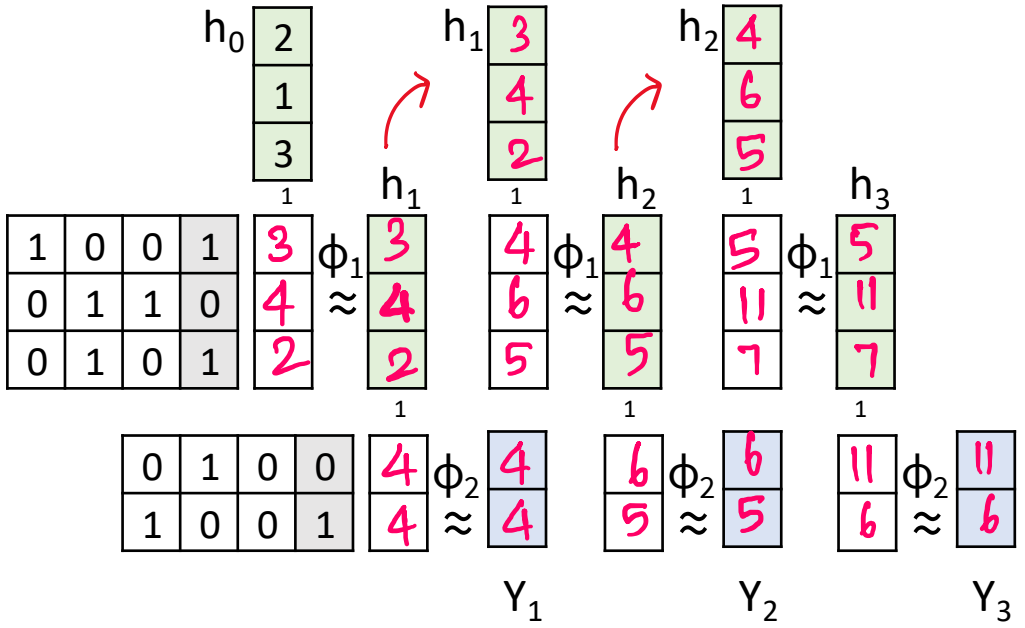
1 -1
1+2+1
2 -1

4+1+1
1-1

1 -1
1+6+1

Calculate a Decoder

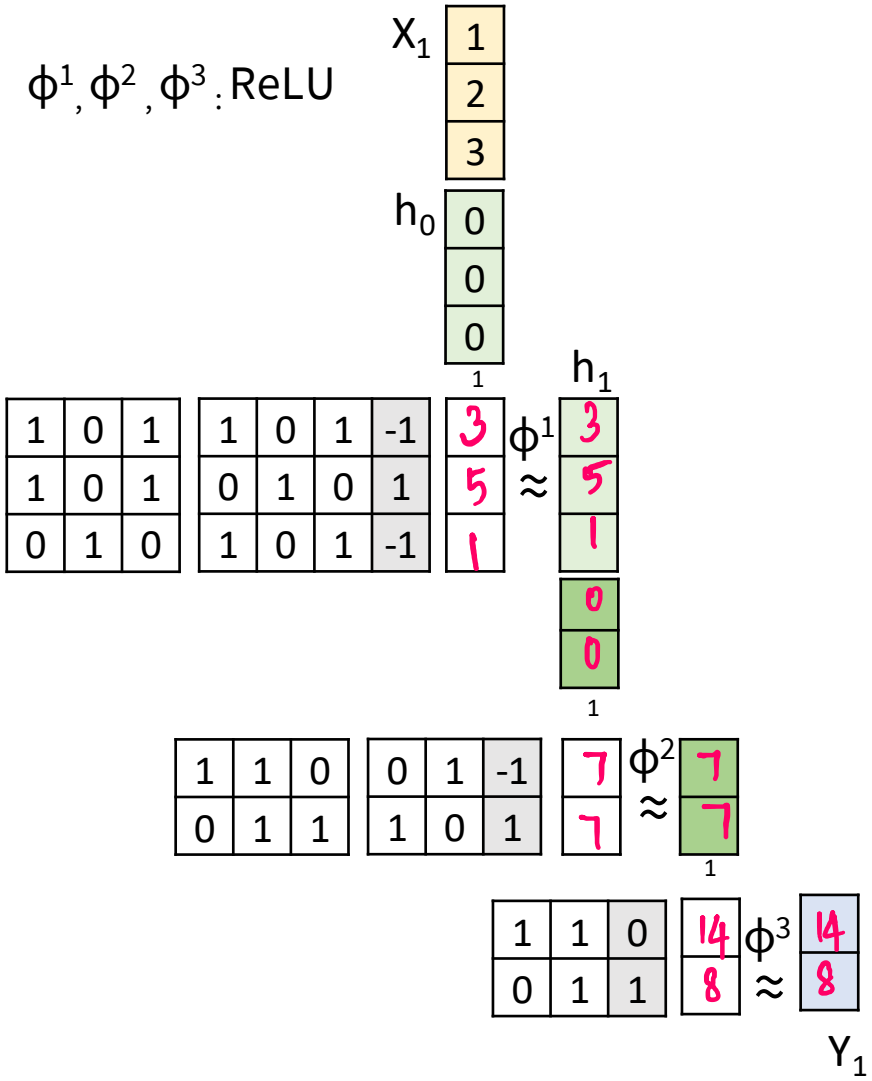
ϕ_1, ϕ_2 : ReLU



15

Multilayer RNN (t = 1)

This activity is standalone, not dependent on other activities.



$1+3-1$

$1+3+1$

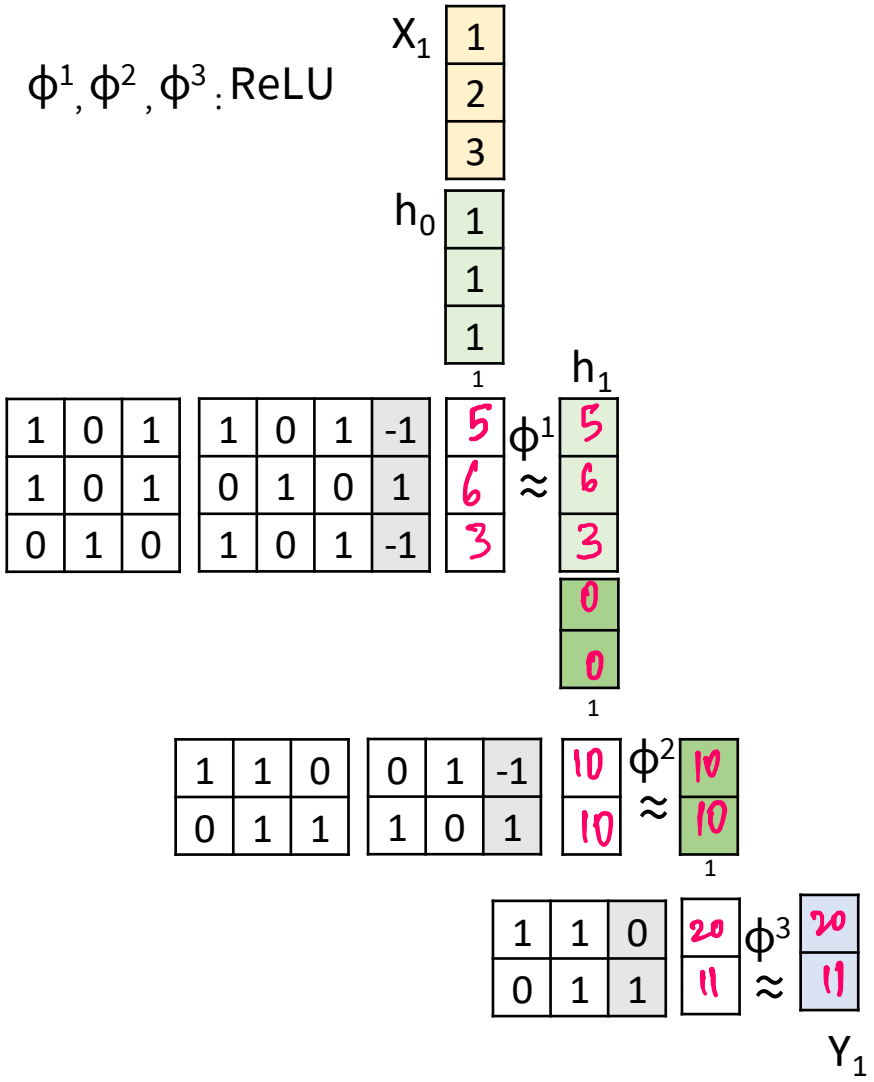
$2-1$

$3+5-1$

$5+1+1$

16

Multilayer RNN (t = 1)



This activity is standalone, not dependent on other activities.

1+3+1+1-1

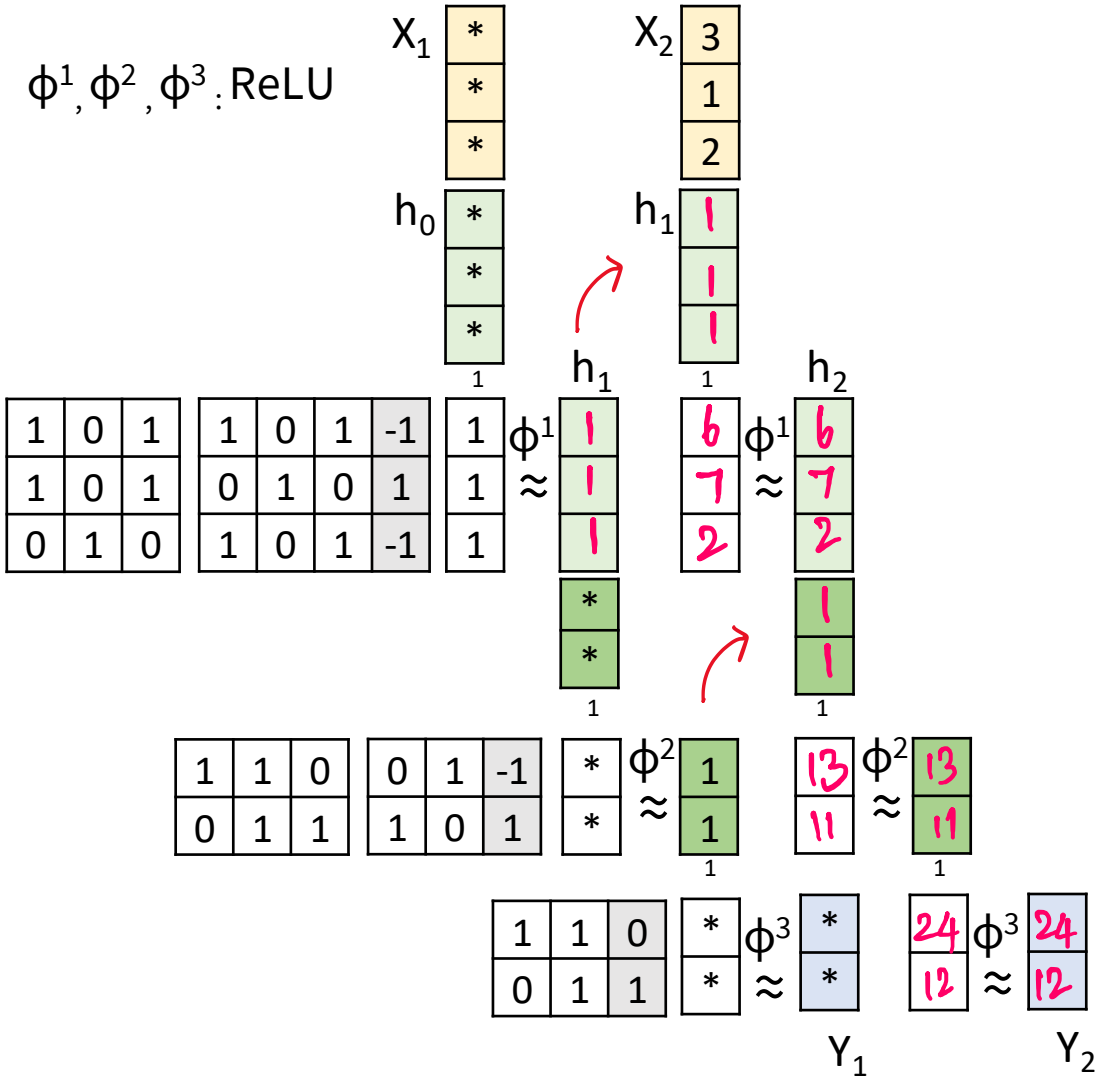
1+3+1+1

2+1+1-1

5+6-1

6+3+1

Multilayer RNN (t = 2)



This activity is standalone, not dependent on other activities.

$3+2+1+1-1$

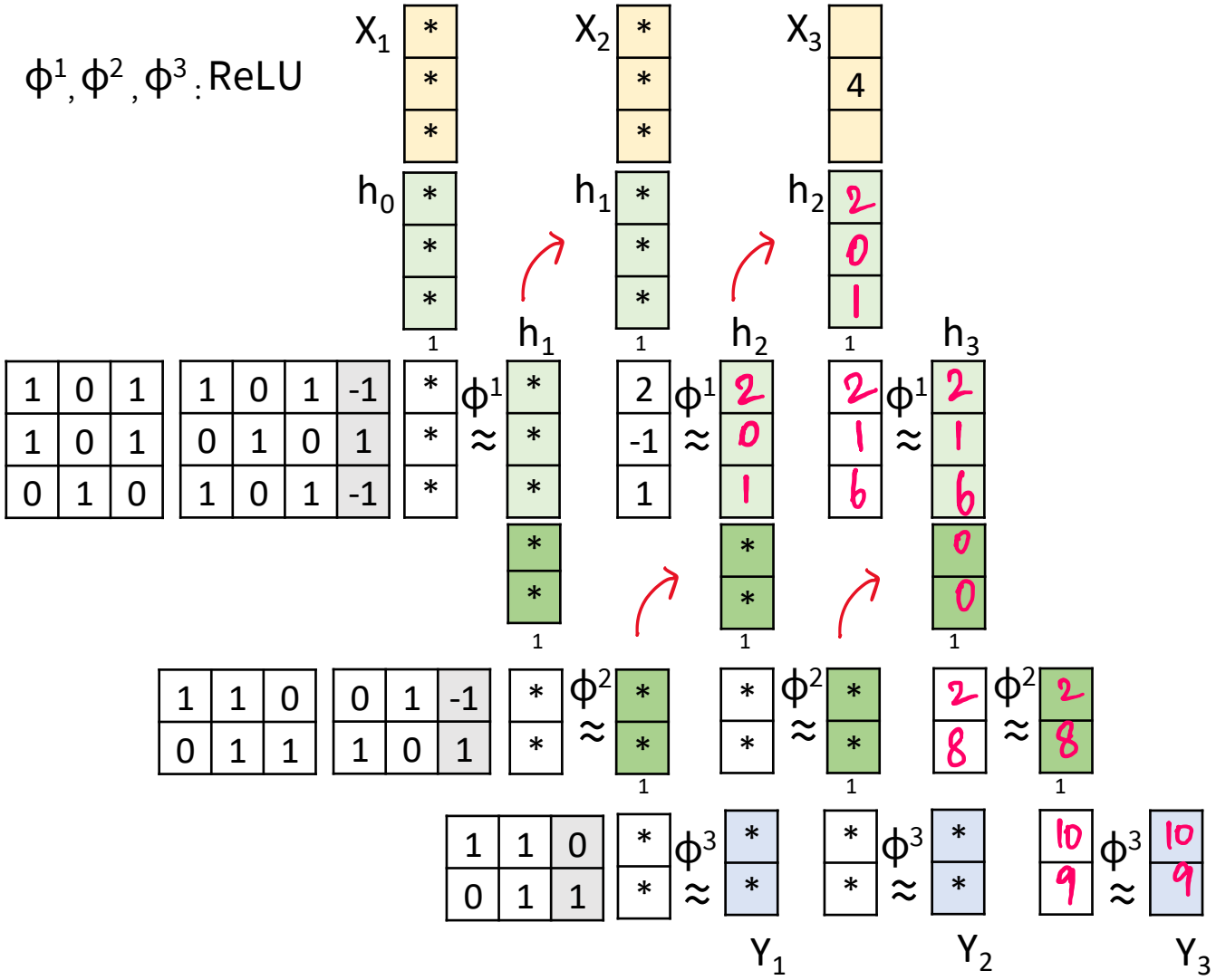
$3+2+1+1$

$1+1+1-1$

$6+7+1-1$

$7+2+1+1$

Multilayer RNN (t = 3)



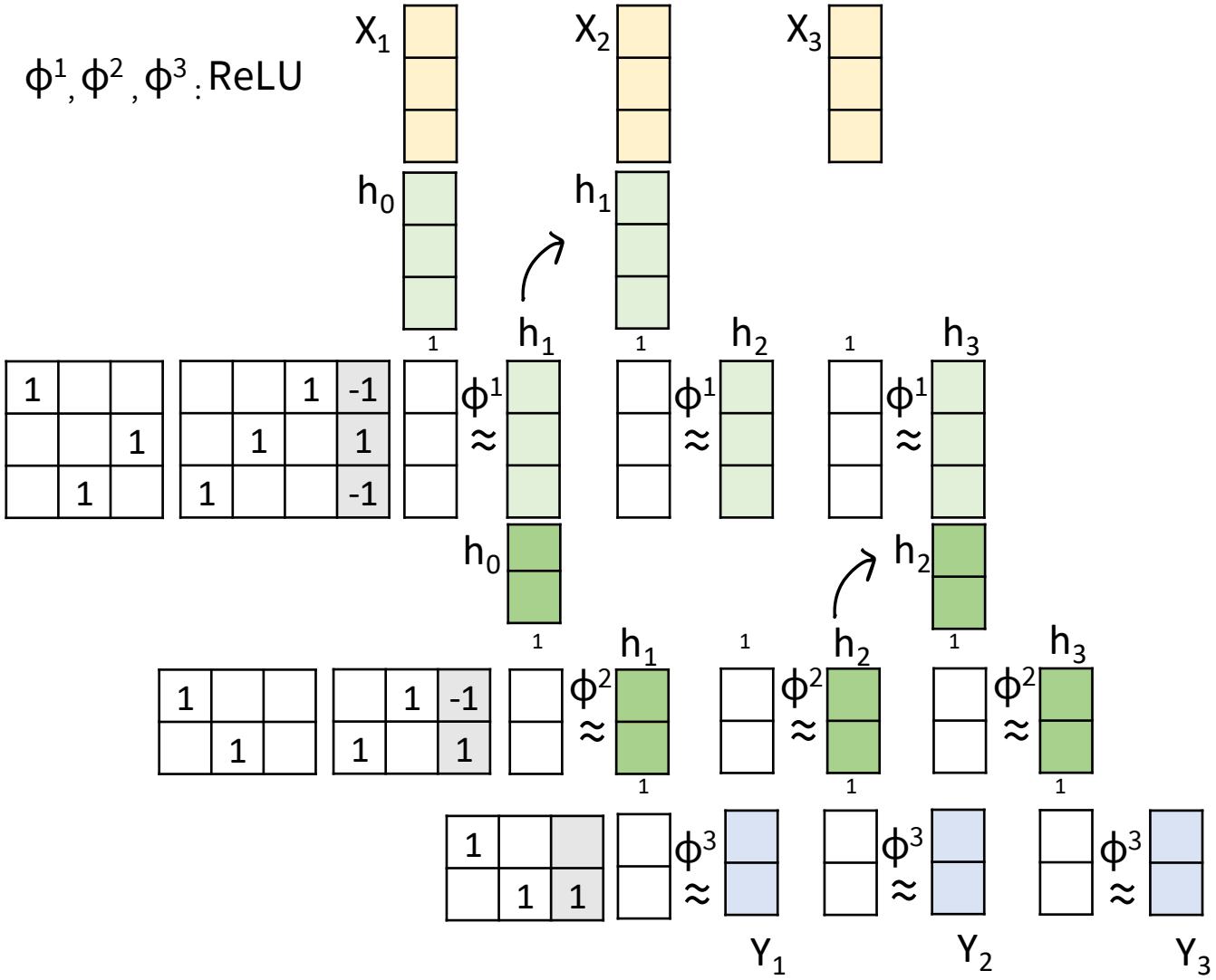
This activity is standalone, not dependent on other activities.

2+1−1
1
4+2+1−1

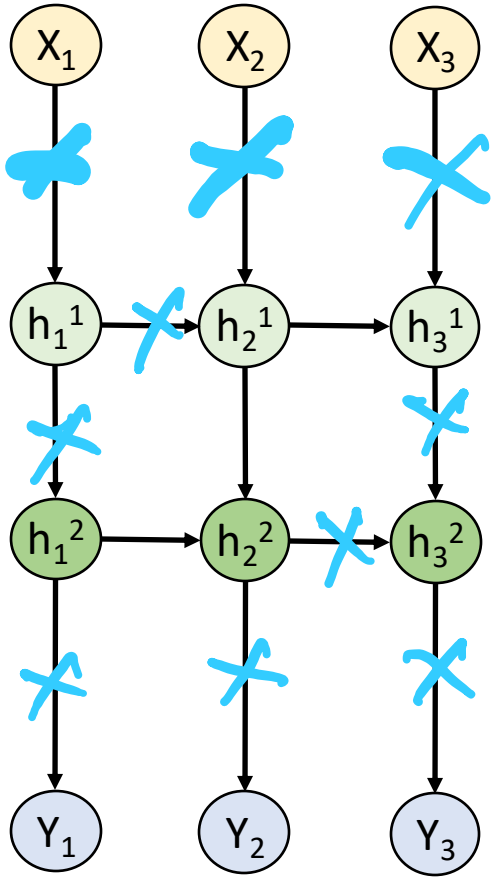
2+1−1
1+6+1

Identify "No Dependency" Links

ϕ^1, ϕ^2, ϕ^3 : ReLU



Cross out the dependency links to match the matrix form.



20 Calculate the Gradient of Softmax + CE Loss

x	e^x	round
0	1	1
1	2.71828	3
2	7.38906	7
3	20.08554	20
4	54.58715	55
5	148.41316	148

Z		Y^{Pred}	Y^{Target}	$\frac{\partial L}{\partial Z}$
0	1	.01	0	.01
1	3	.03	0	.03
4	55	0.55	0	0.55
1	3	0.03	0	0.03
0	1	0.01	1	-0.99
3	20	0.2	0	0.2
2	7	0.07	0	0.07
2	7	0.07	0	0.07
1	3	0.03	0	0.03

softmax ≈