AI by Hand ✍️
25 Exercise Workbook Series

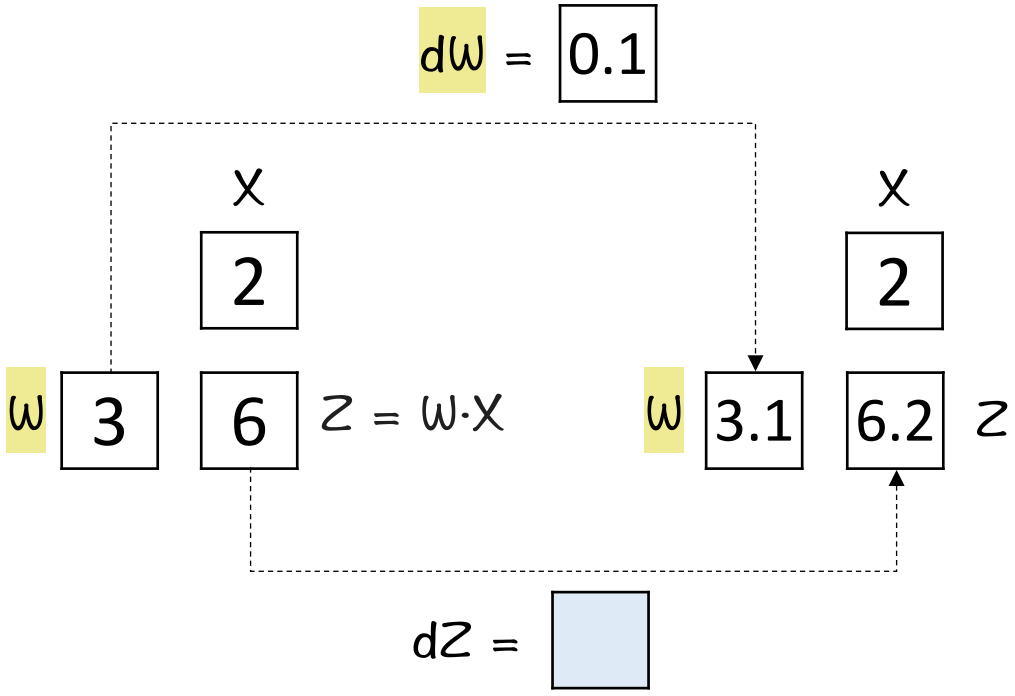
1. Dot Product
2. Matrix Multiplication
3. Linear Layer
4. Activation
5. Artificial Neuron
6. Batch
7. Connection
8. Hidden Layer
9. Deep
10. Wide
11. Softmax
12. Gradient

More to come …

http://by-hand.ai/workbook

Gradient

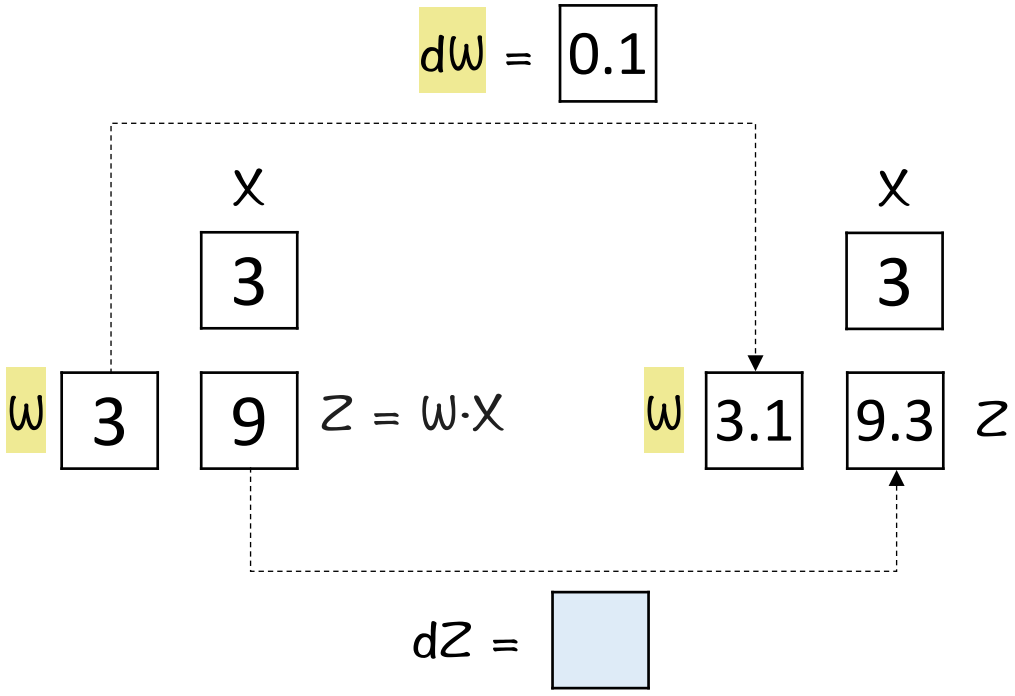
Exercise 1



$$\frac{dZ}{dW} = \frac{\square}{0.1} = \square = X$$

Gradient

Exercise 2

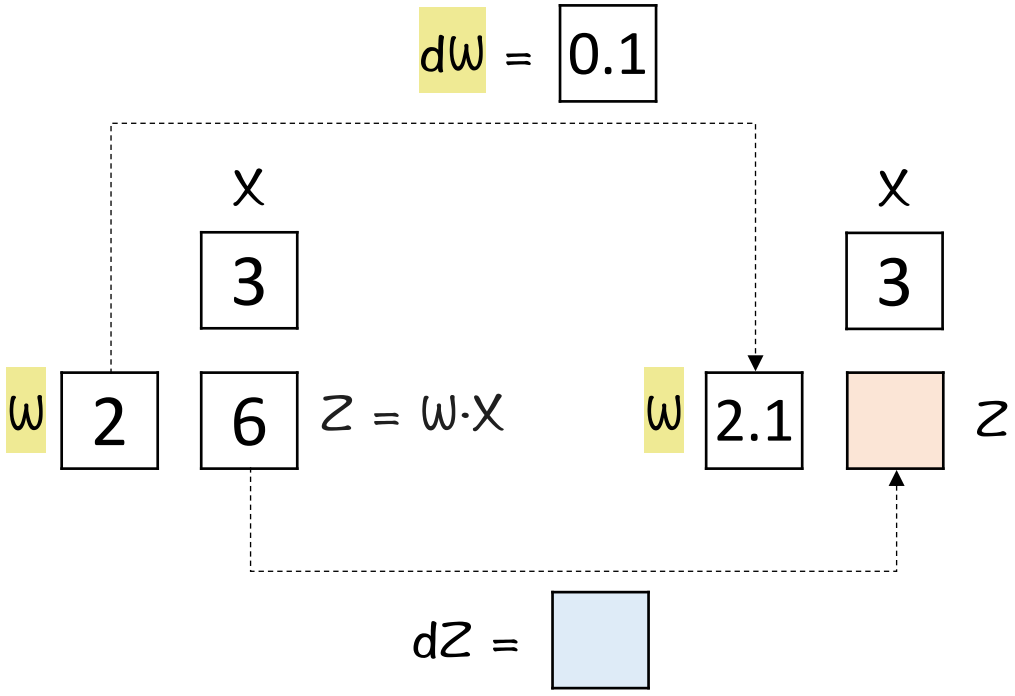


$$\frac{dZ}{dW} = \frac{\square}{0.1} = \square = X$$

0.3 3

Gradient

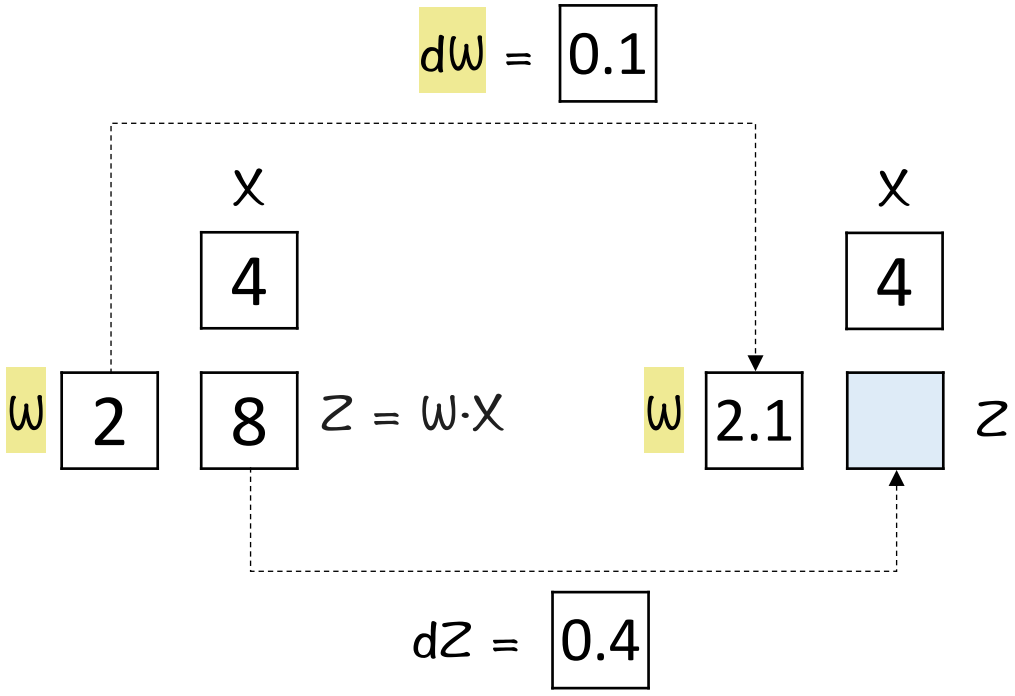
Exercise 3



$$\frac{dZ}{dW} = \frac{\square}{0.1} = 3 = X$$

Gradient

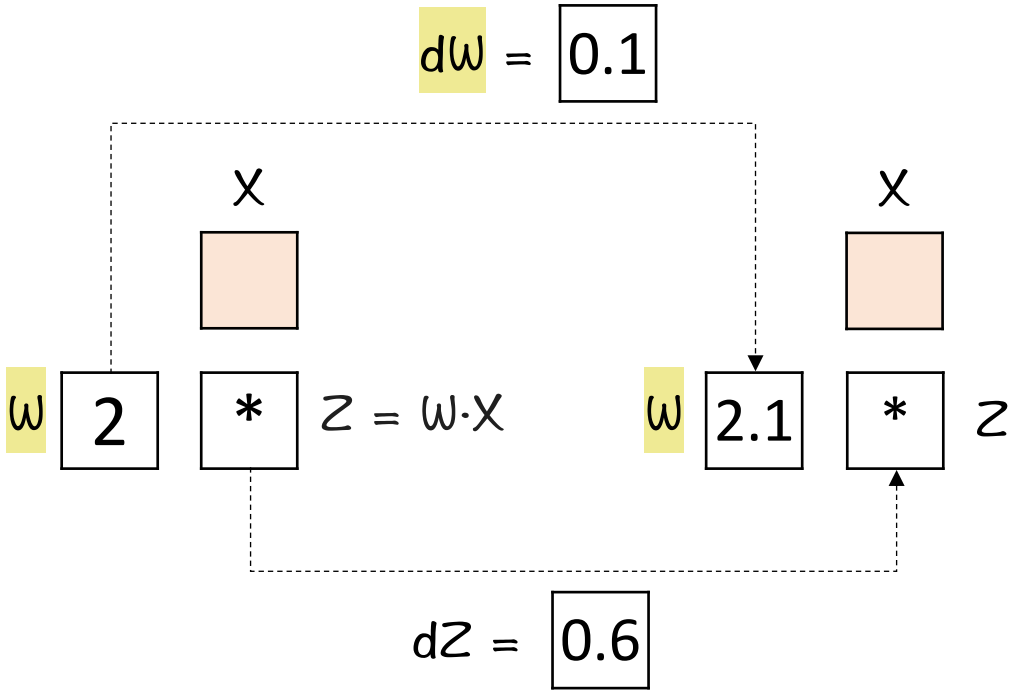
Exercise 4



$$\frac{dZ}{dW} = \frac{0.4}{0.1} = \square = X$$

Gradient

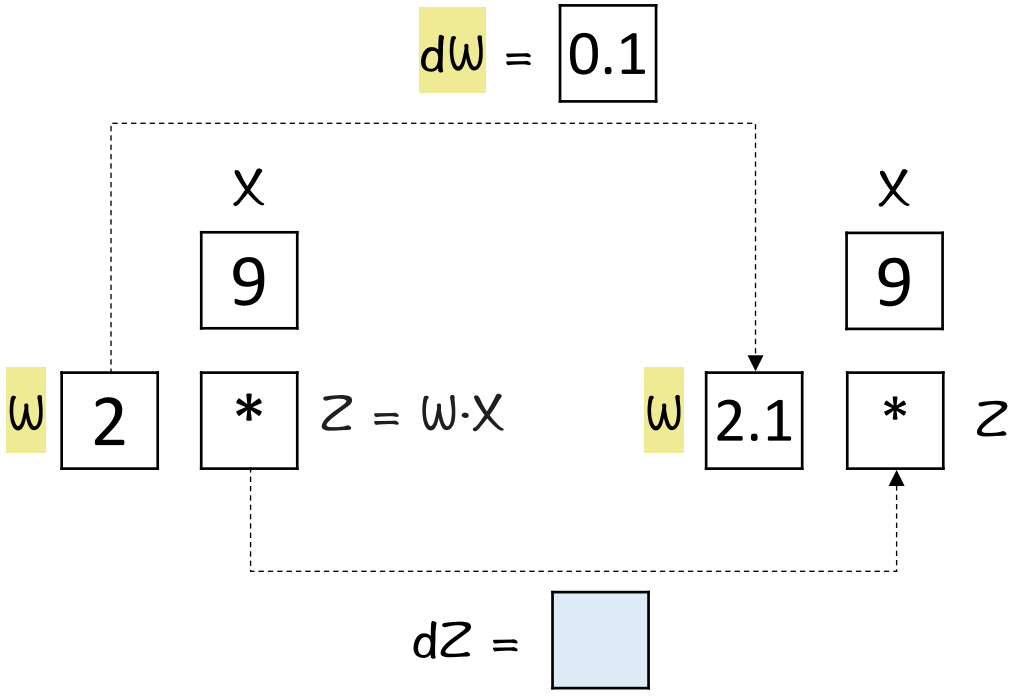
Exercise 5



$$\frac{dZ}{dW} = \frac{0.6}{0.1} = \square = X$$

Gradient

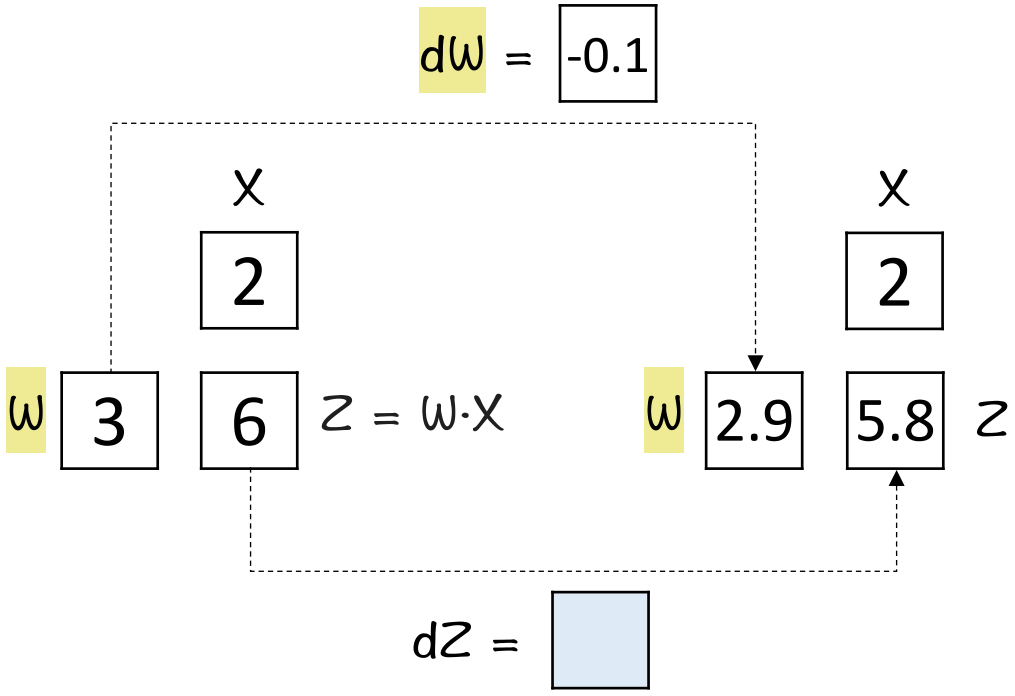
Exercise 6



$$\frac{dZ}{dW} = \frac{\square}{0.1} = \square = X$$

Gradient

Exercise 7

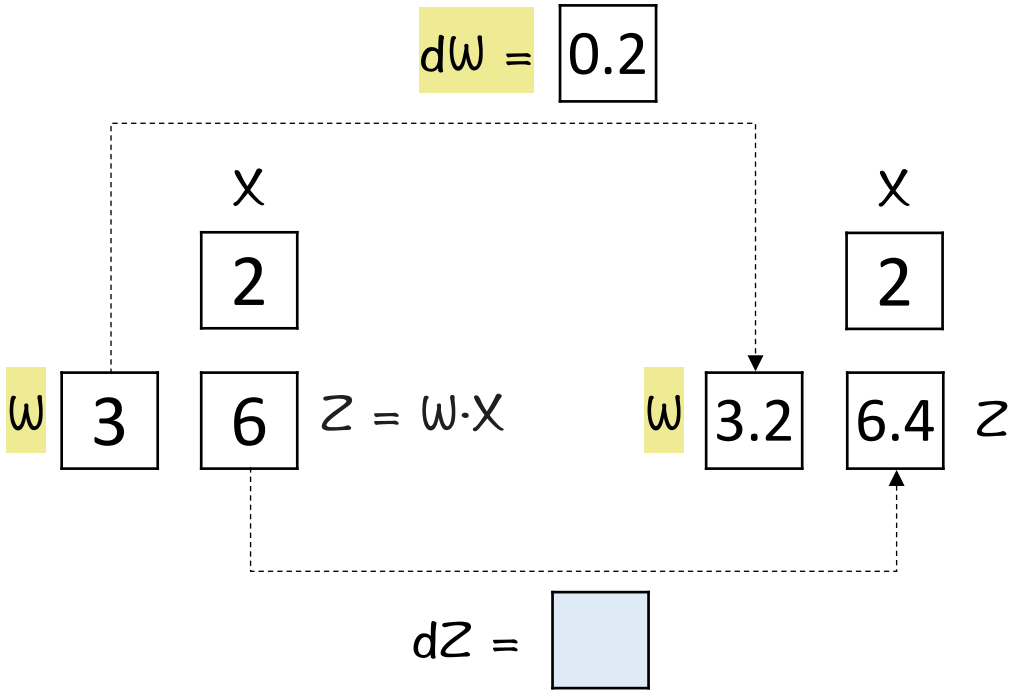


$$\frac{dZ}{dW} = \frac{\square}{-0.1} = \square = X$$

-0.2 2

Gradient

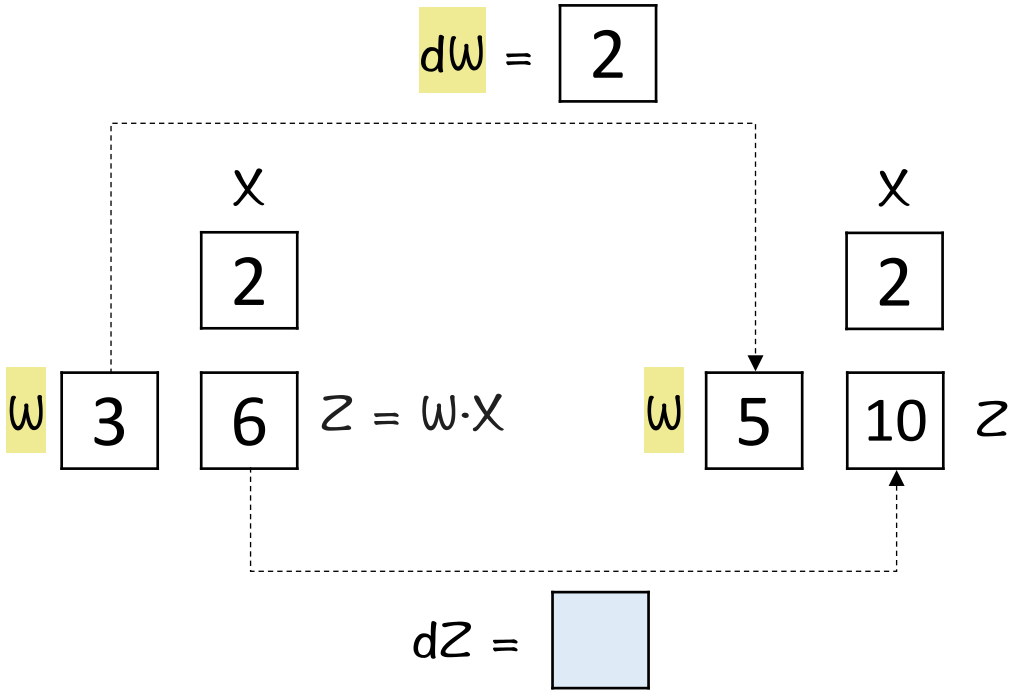
Exercise 8



$$\frac{dZ}{dW} = \frac{\square}{0.2} = \square = X$$

Gradient

Exercise 9

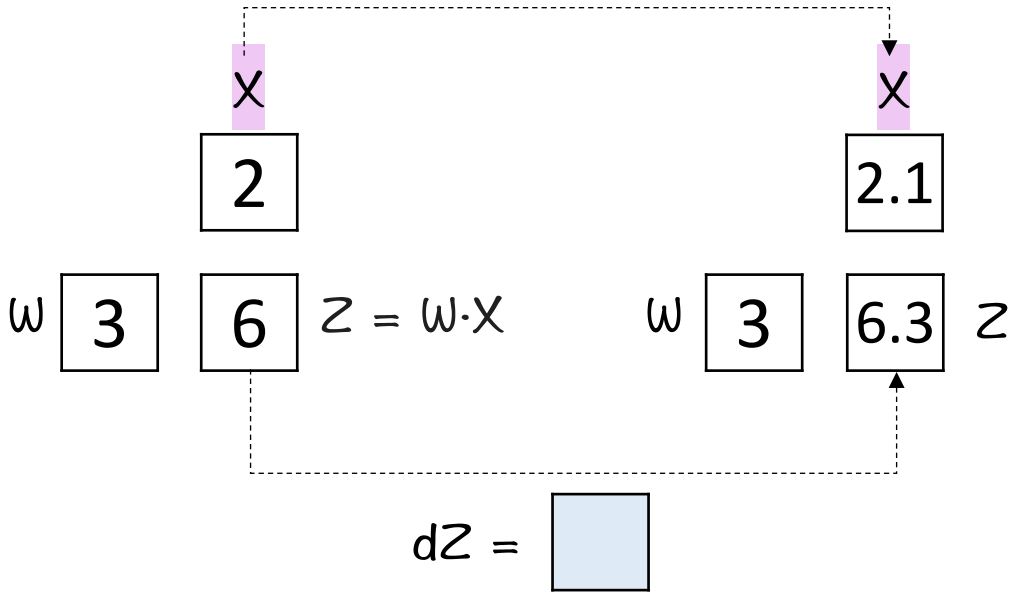


$$\frac{dZ}{dW} = \frac{\square}{2} = \square = X$$

Gradient

Exercise 10

dX = 0.1

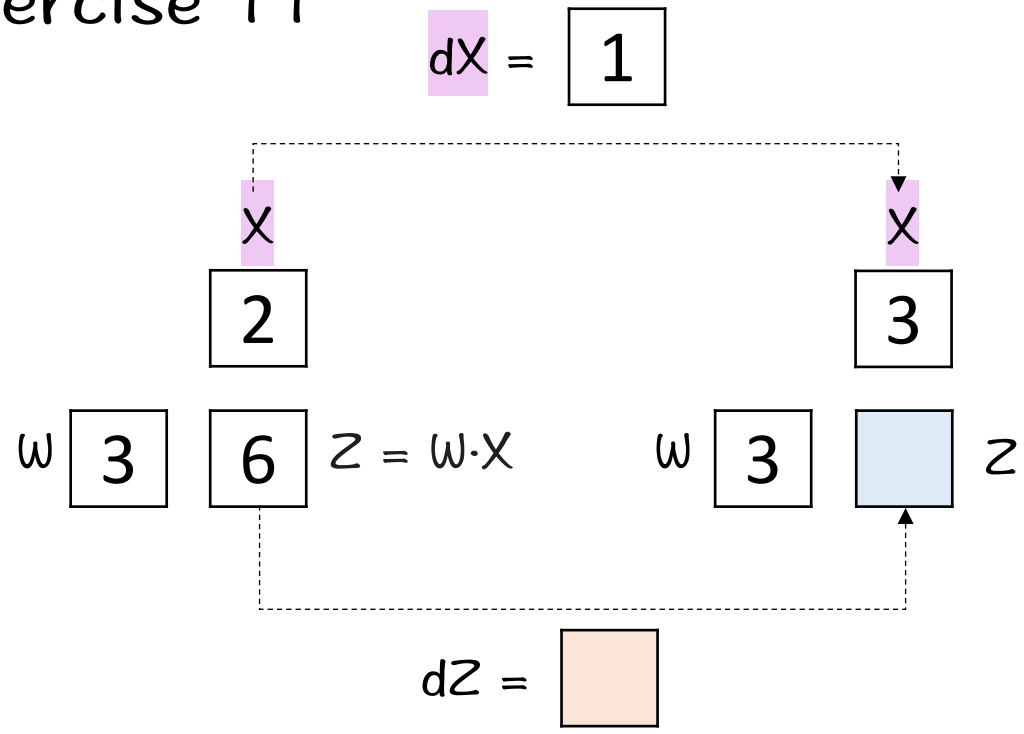


dZ = ☐

$$\frac{dZ}{dX} = \frac{\square}{0.1} = \square = W$$

Gradient

Exercise 11

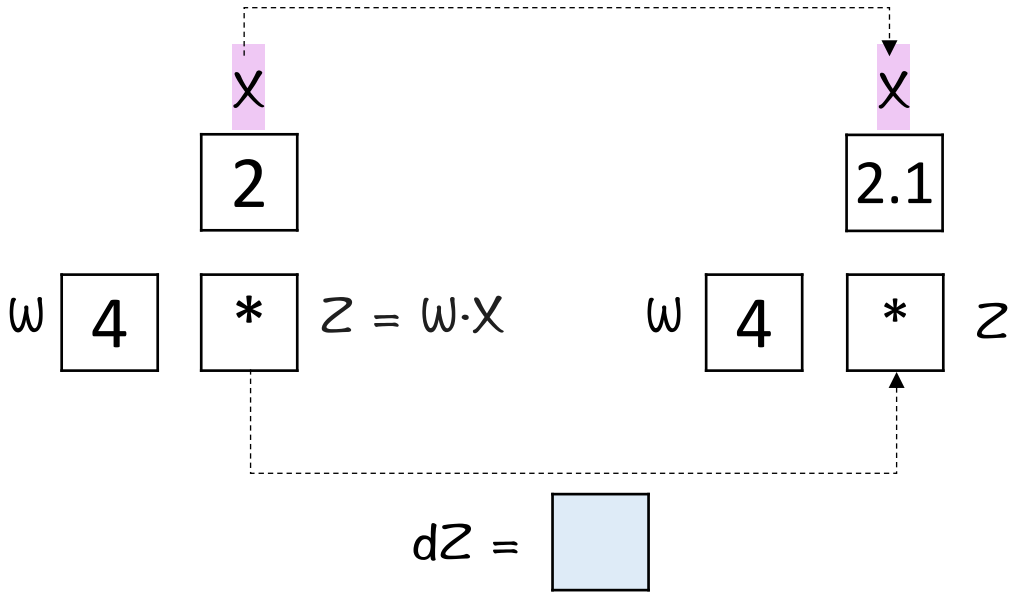


$$\frac{dZ}{dX} = \frac{\square}{1} = 3 = W$$

Gradient

Exercise 12

$dX = \boxed{0.1}$



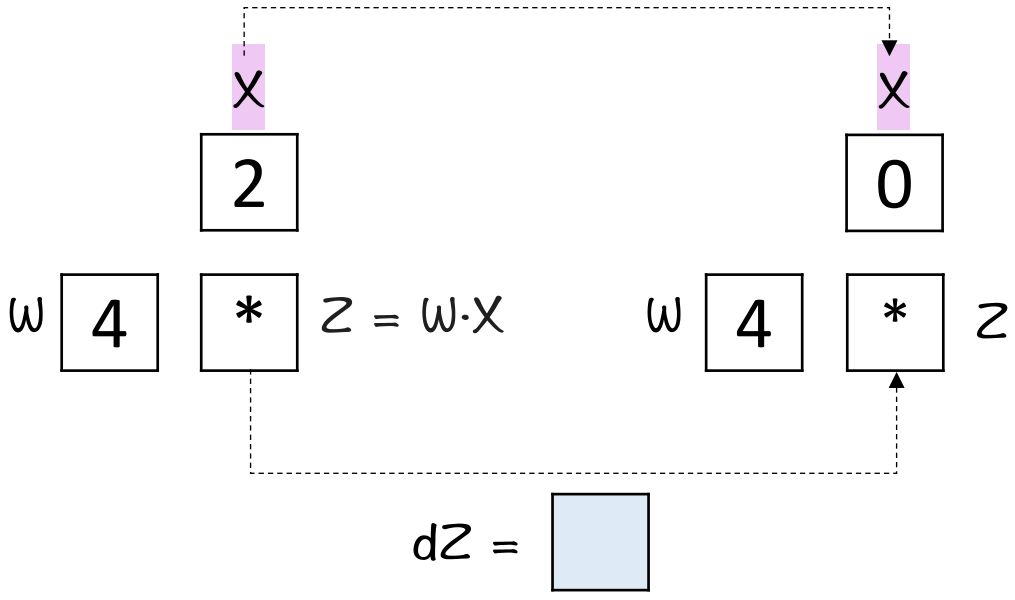
$dZ = \square$

$$\frac{dZ}{dX} = \frac{\square}{0.1} = \square = W$$

Gradient

Exercise 13

$dX =$ -2

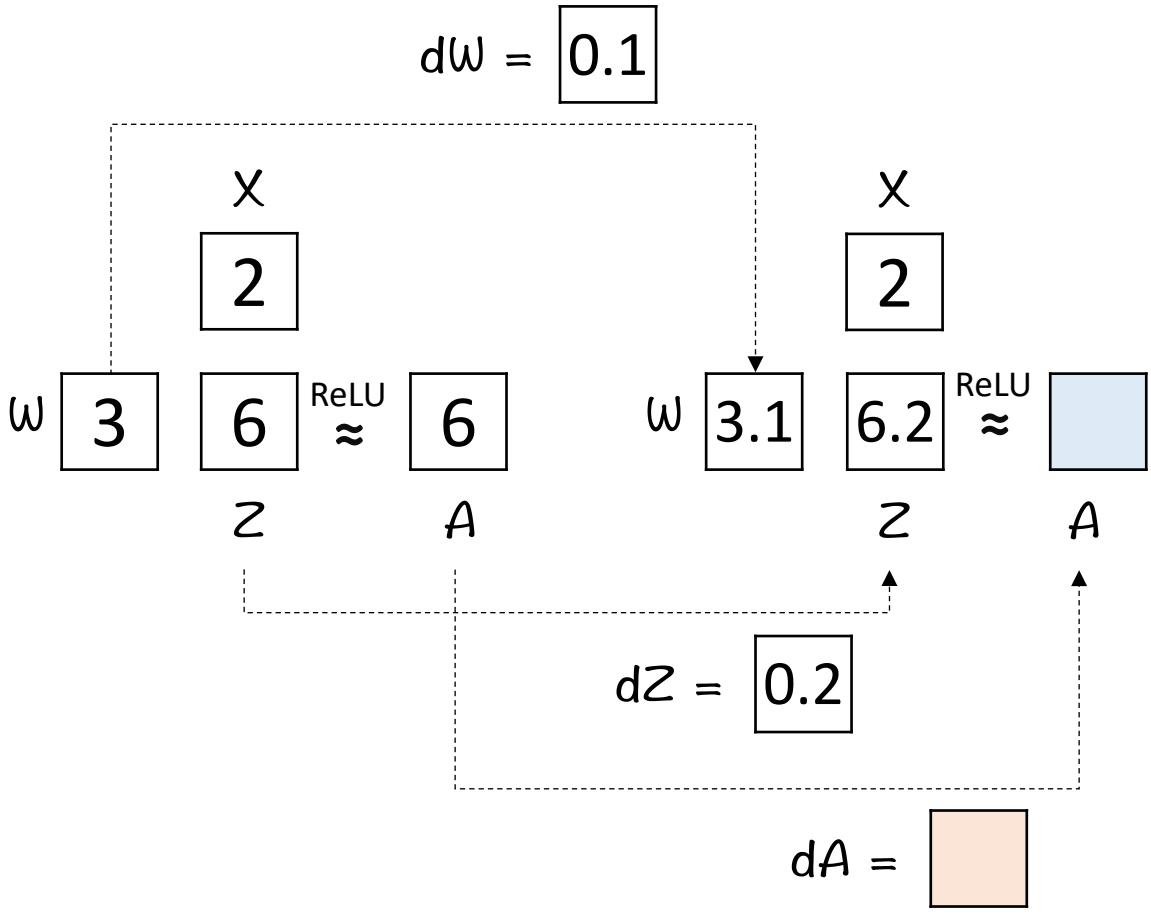


$dZ =$ ☐

$$\frac{dZ}{dX} = \frac{\square}{-2} = \square = W$$

Gradient

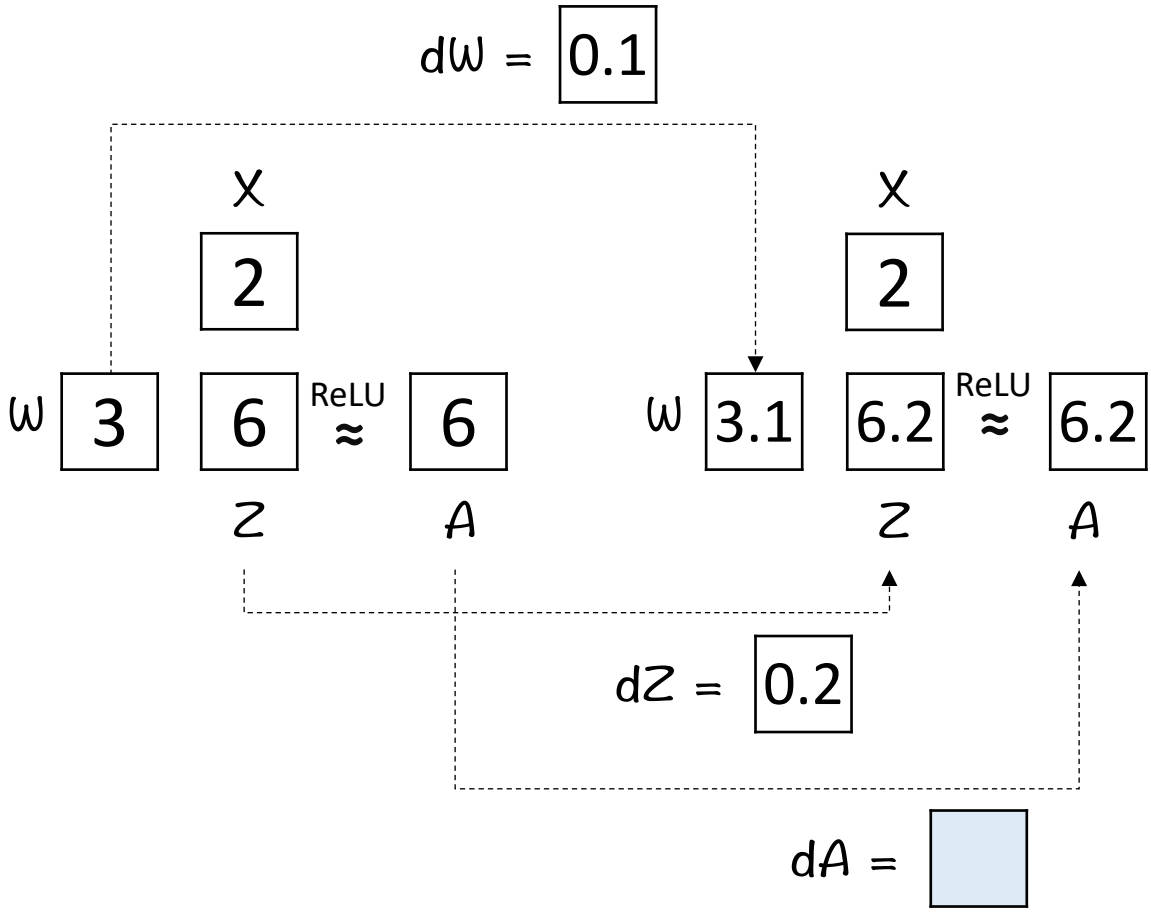
Exercise 14



6.2 0.2

Gradient

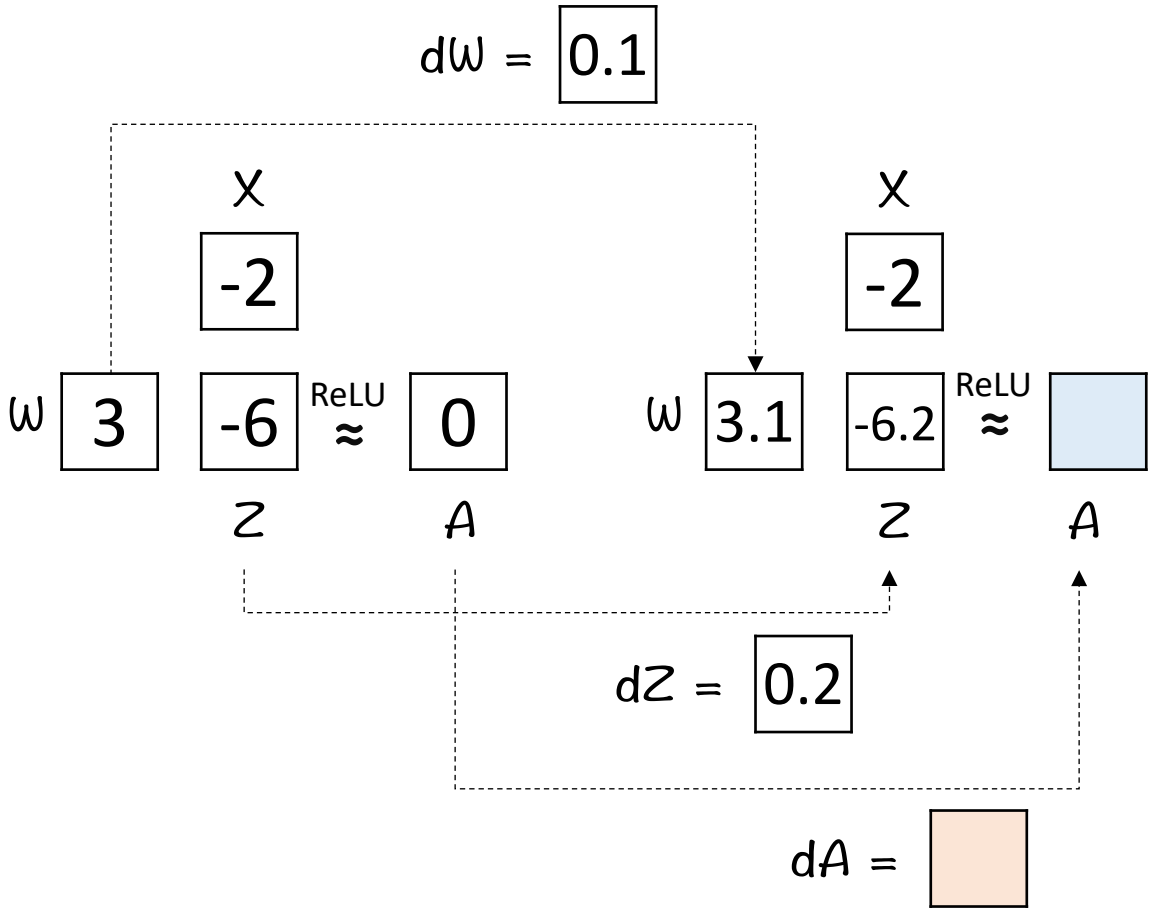
Exercise 15



$$\frac{dA}{dZ} = \frac{\square}{0.2} = \square$$

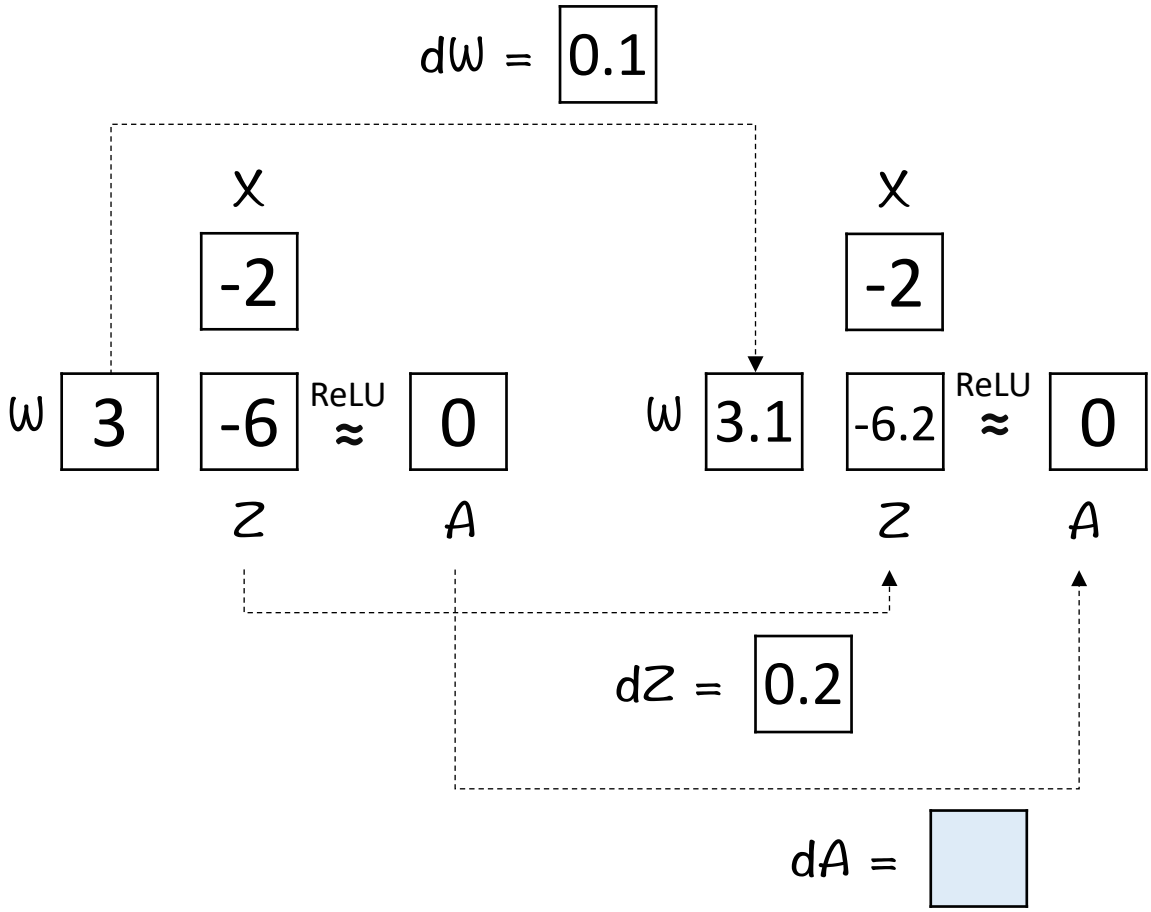
Gradient

Exercise 16



Gradient

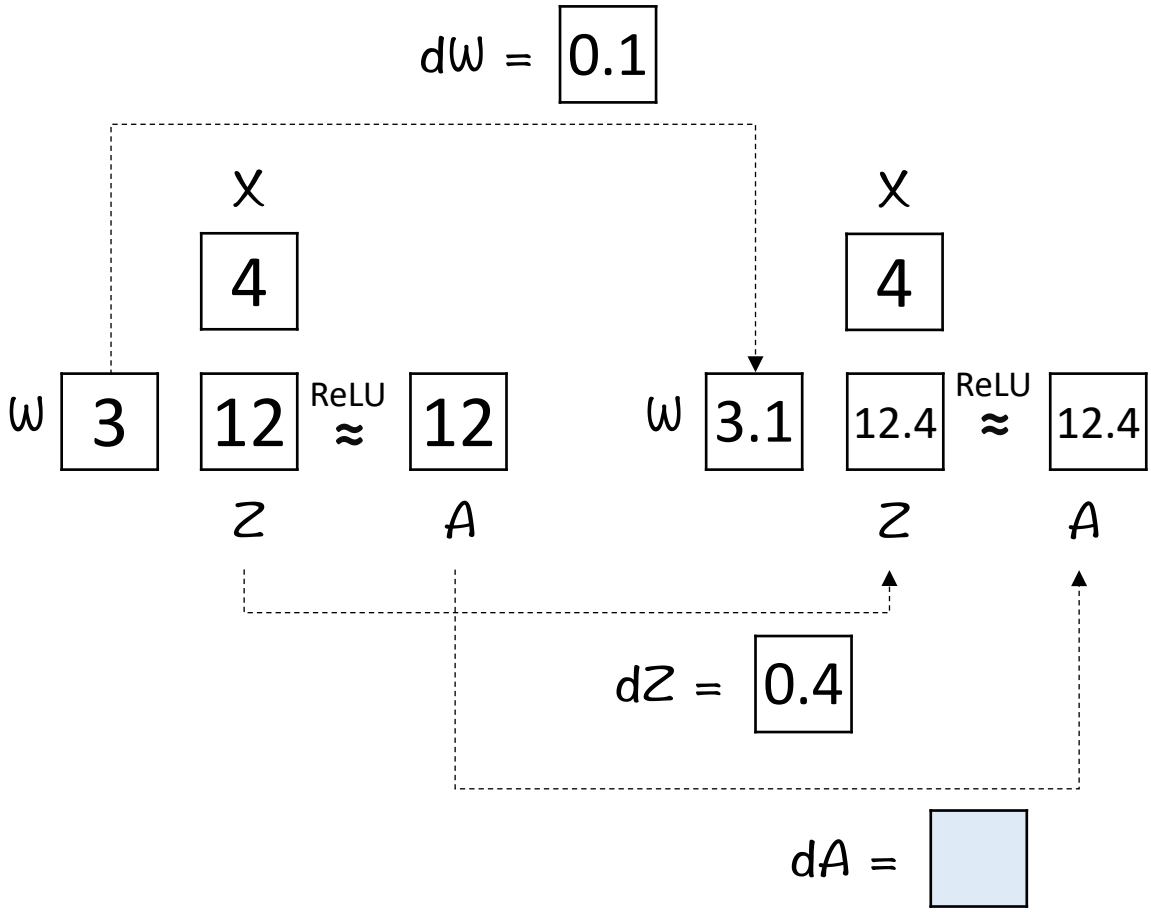
Exercise 17



$$\frac{dA}{dZ} = \frac{\square}{0.2} = \square$$

Gradient

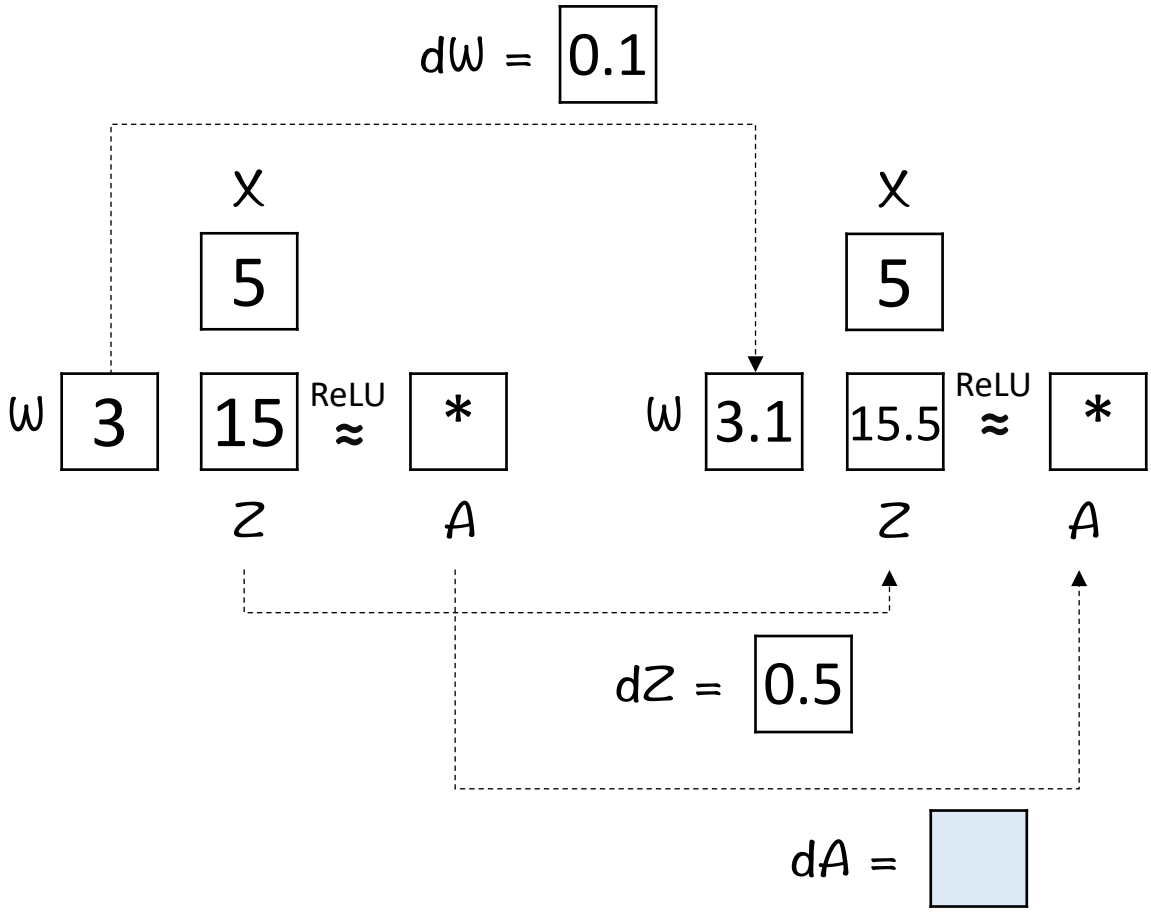
Exercise 18



$$\frac{dA}{dZ} = \frac{\square}{0.4} = \square$$

Gradient

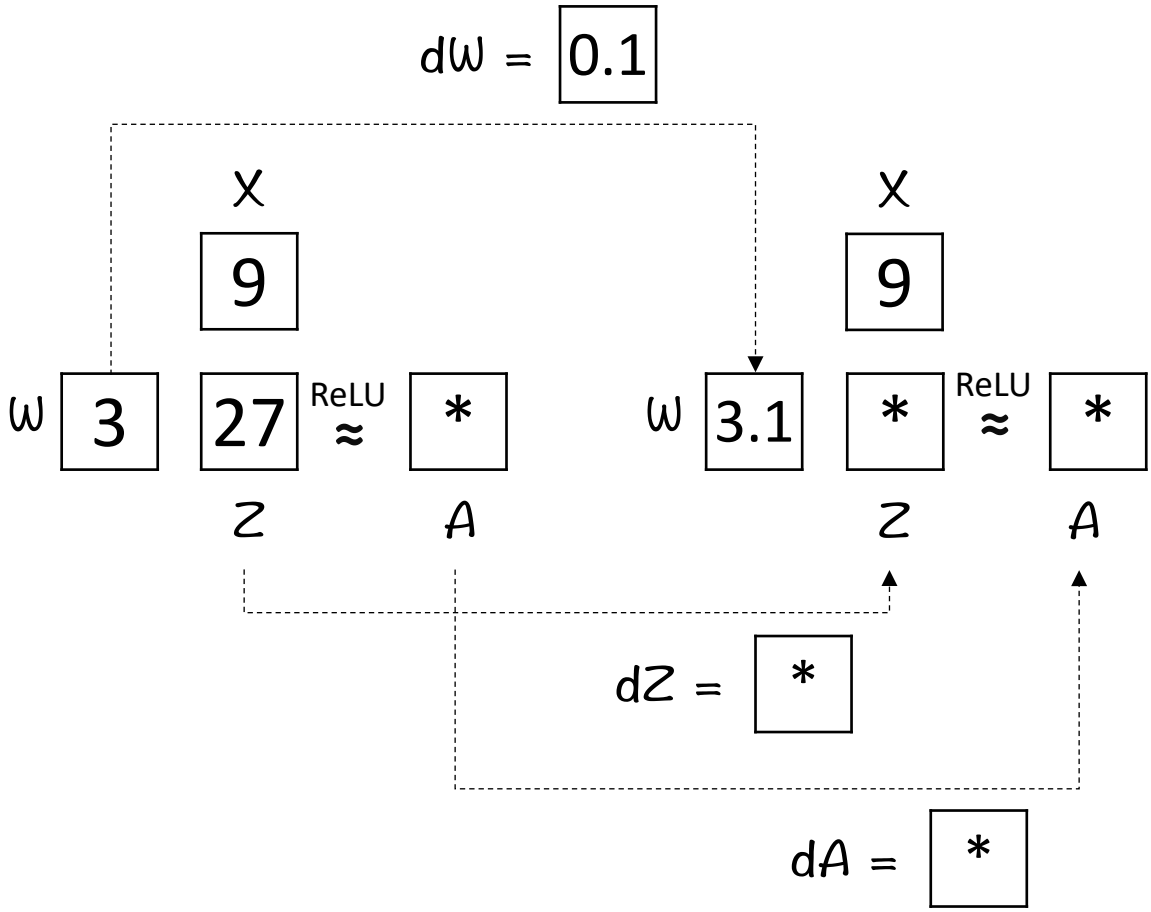
Exercise 19



$$\frac{dA}{dZ} = \frac{\square}{0.5} = \square$$

Gradient

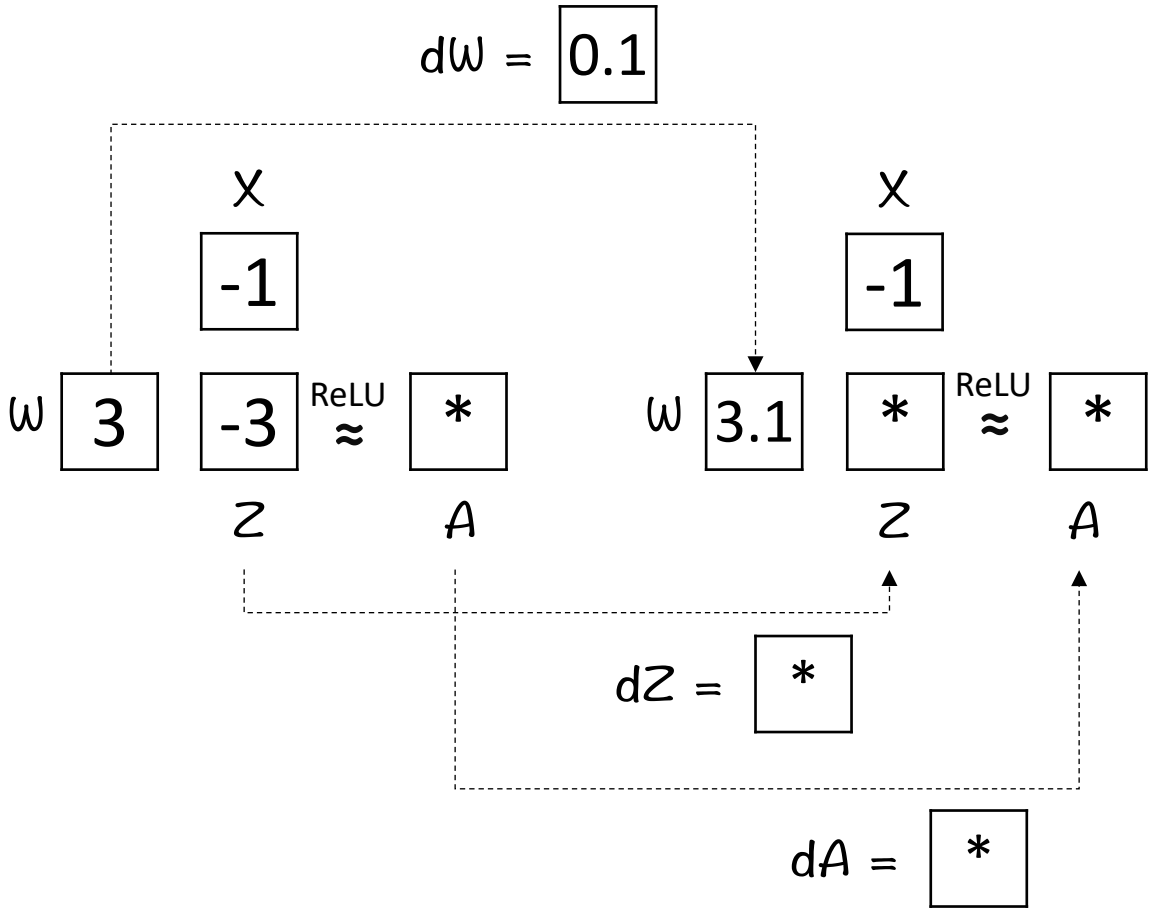
Exercise 20



$$\frac{dA}{dZ} = \frac{*}{*} = \square$$

Gradient

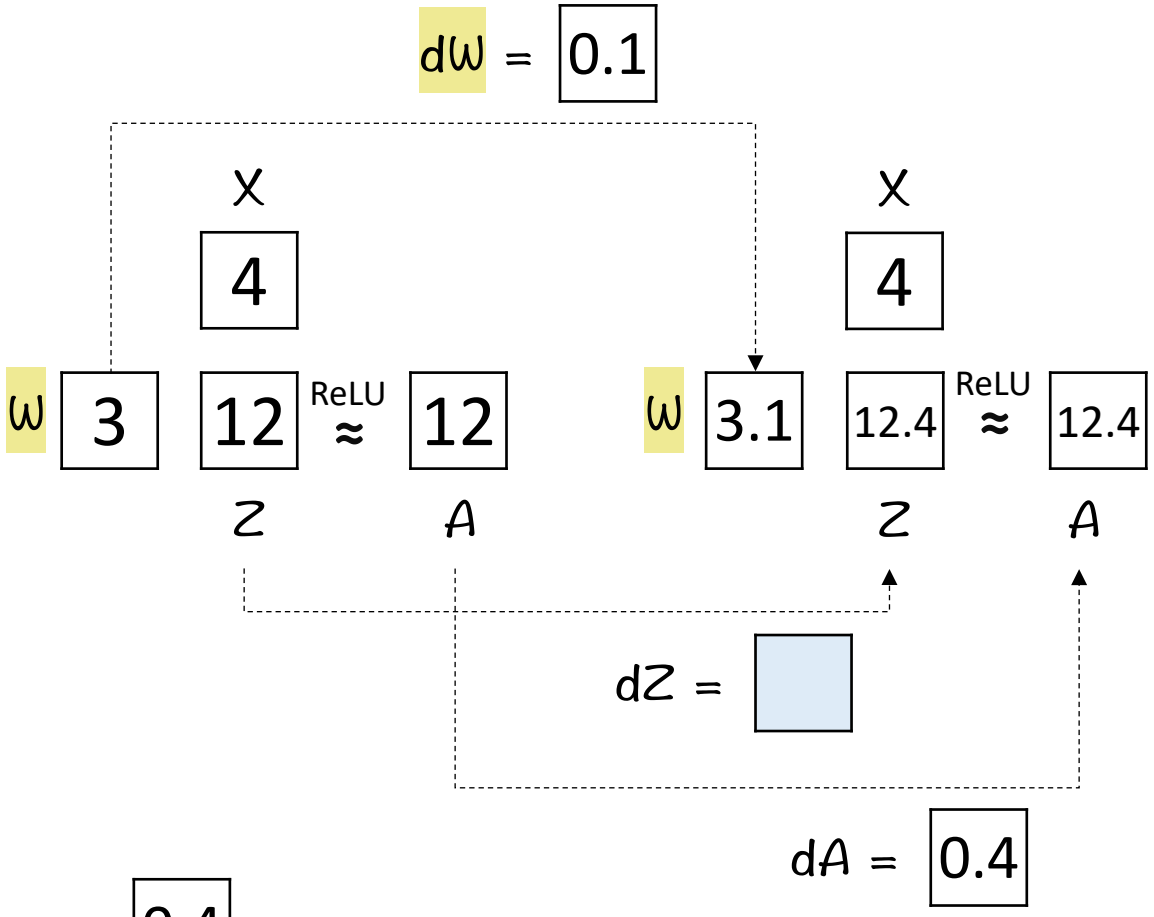
Exercise 21



$$\frac{dA}{dZ} = \frac{*}{*} = \square$$

Gradient

Exercise 22



$$\frac{dA}{dZ} = \frac{0.4}{\square} = 1$$

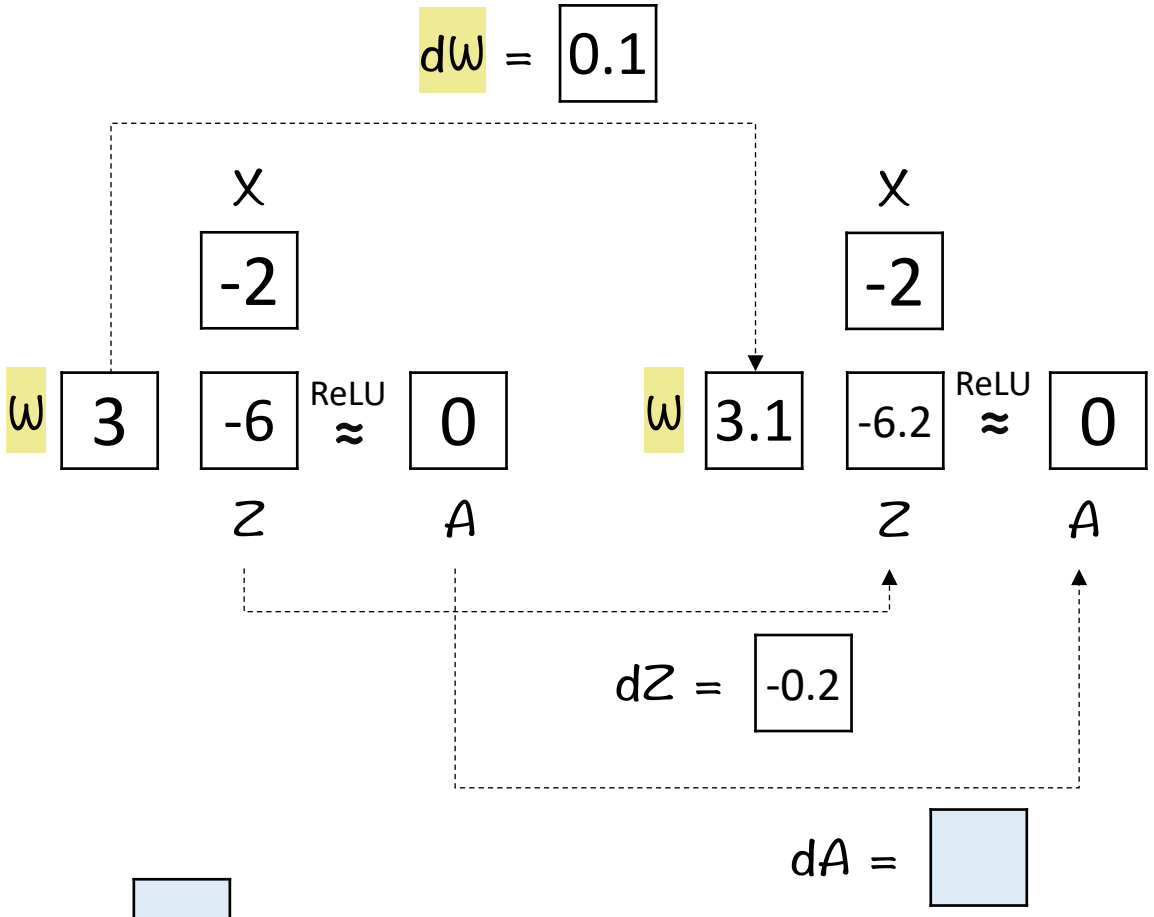
$$\frac{dZ}{dW} = \frac{\square}{0.1} = 4$$

$$\frac{dA}{dW} = \frac{0.4}{0.1} = \square$$

0.4 4

Gradient

Exercise 23



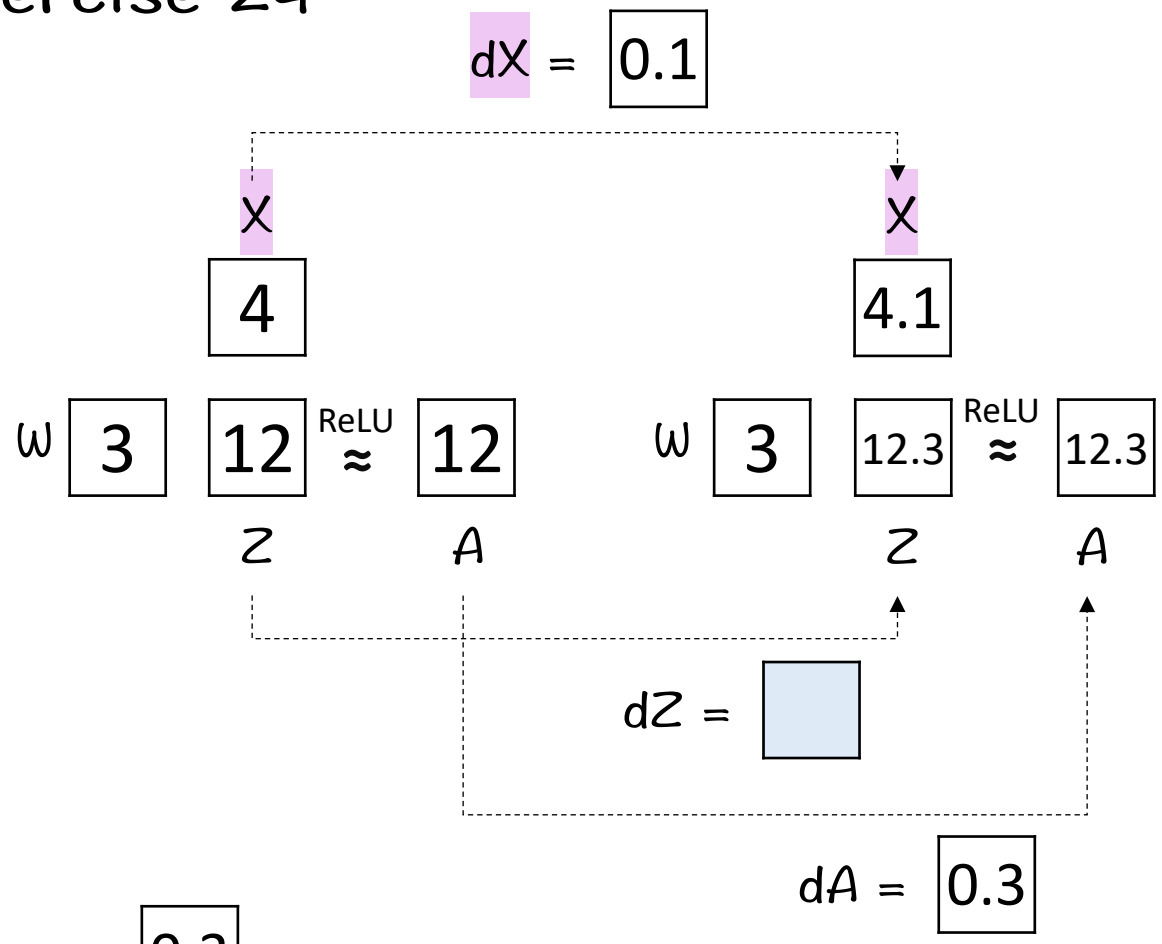
$$\frac{dA}{dZ} = \frac{\square}{-0.2} = 0$$

$$\frac{dZ}{dW} = \frac{-0.2}{0.1} = -2$$

$$\frac{dA}{dW} = \frac{\square}{0.1} = \square$$

Gradient

Exercise 24



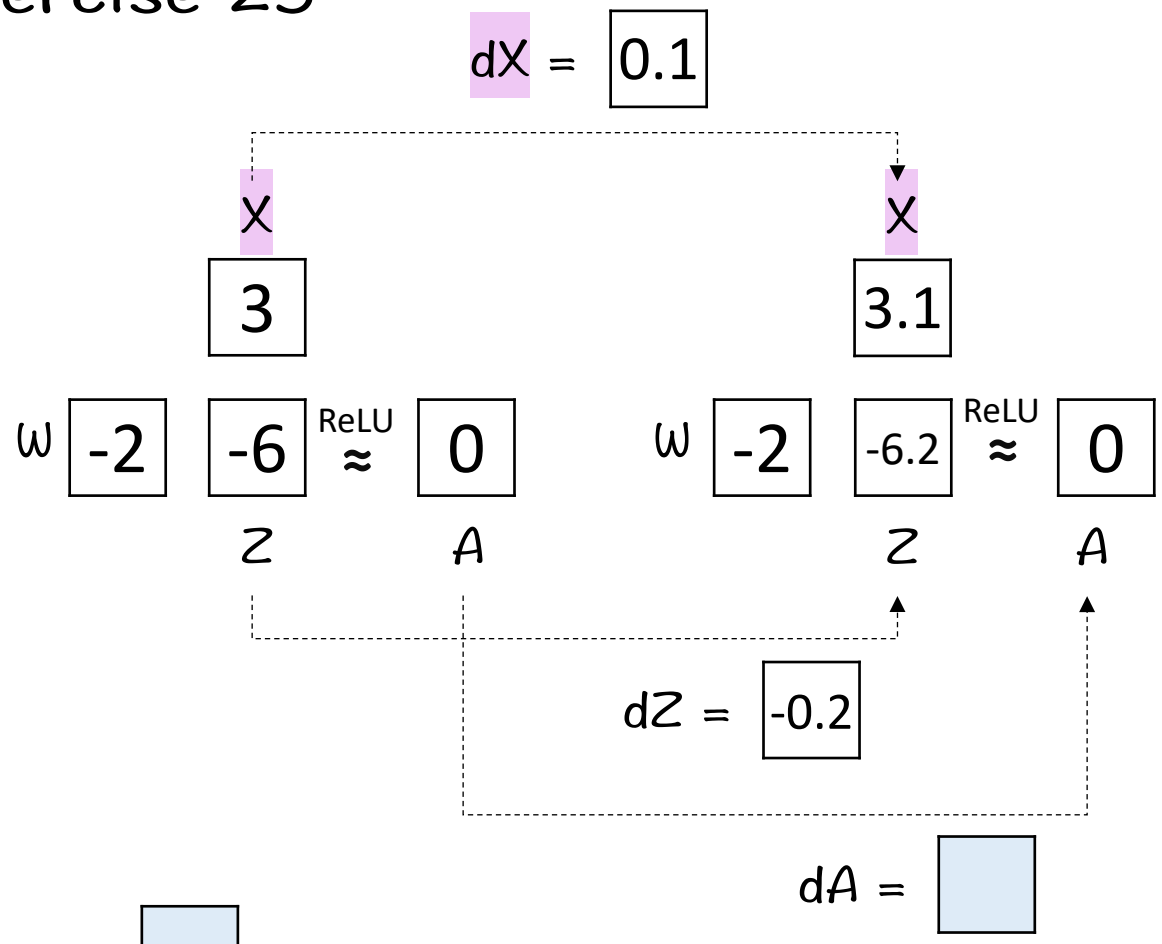
$$\frac{dA}{dZ} = \frac{0.3}{\square} = 1$$

$$\frac{dZ}{dX} = \frac{\square}{0.1} = 3$$

$$\frac{dA}{dX} = \frac{0.3}{0.1} = \square$$

Gradient

Exercise 25



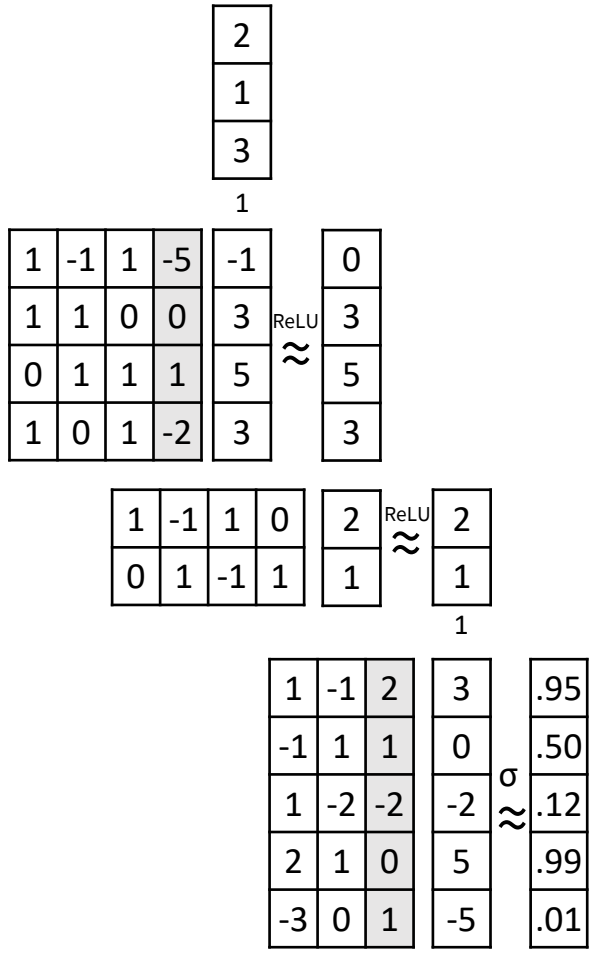
$$\frac{dA}{dZ} = \frac{\square}{-0.2} = 0$$

$$\frac{dZ}{dX} = \frac{-0.2}{0.1} = -2$$

$$\frac{dA}{dX} = \frac{\square}{0.1} = \square$$

Multi Layer Perceptron in pytorch

```
1  mlp_model = nn.Sequential(
2      nn._____( ___, ___, bias = ___ ),
3      nn._____(),
4      nn._____( ___, ___, bias = ___ ),
5      nn._____(),
6      nn._____( ___, ___, bias = ___ ),
7      nn._____()
8  )
```



Hints:

Linear Layer: { Identity | Linear | Bilinear }

Activation Function: { ReLU | Tanh | Sigmoid }

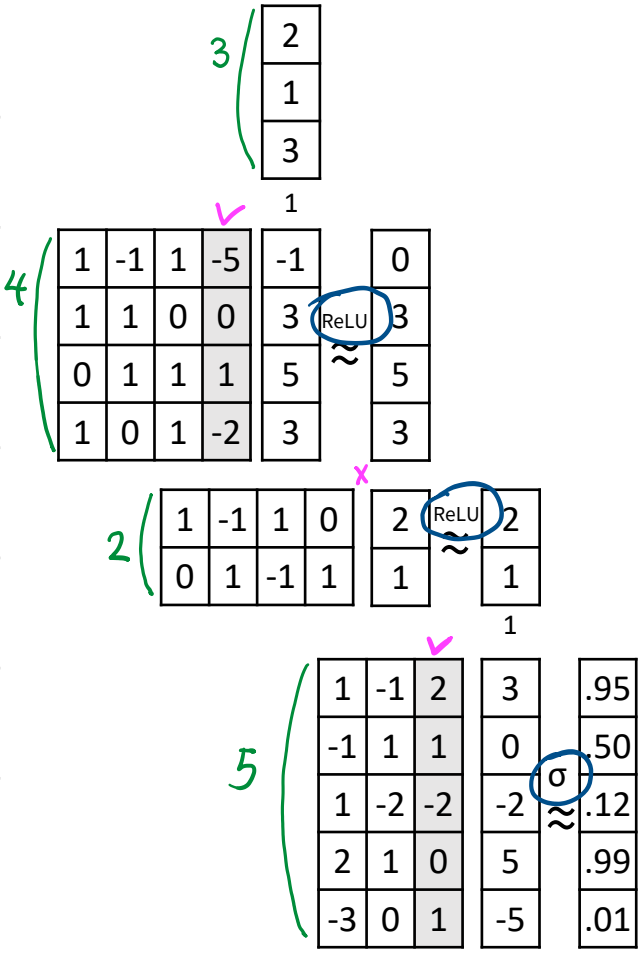
in_features: { int }

out_features: { int }

bias: { T | F }

Multi Layer Perceptron in pytorch

```
1  mlp_model = nn.Sequential(
2      nn.Linear( 3, 4, bias = T ),
3      nn.ReLU (),
4      nn.Linear( 4, 2, bias = F ),
5      nn.ReLU (),
6      nn.Linear( 2, 5, bias = T ),
7      nn.Sigmoid()
8  )
```



Hints:

Linear Layer: { Identity | Linear | Bilinear }

Activation Function: { ReLU | Tanh | Sigmoid }

in_features: { int }

out_features: { int }

bias: { T | F }

Dropout

Random Sequence

.61
.39
.75
.40
.65
.42
.23
.19
.93
.42
.87
.53
.27
.69
.50
.11
.42

Training Data:

X_1	X_2
3	5
4	1
1	1

Linear

1	0	0
1	1	0
0	1	1
1	-1	0

[≈ ReLU]

1	1

Dropout (p=0.5)

	0	0	0
0		0	0
0	0		0
0	0	0	

Linear

1	0	0	1	0
0	1	1	0	0
1	0	-1	-1	1

[≈ ReLU]

1	1

Dropout (p=0.33)

	0	0
0		0
0	0	

1	1

Linear

1	-1	0	0
0	1	-1	-2

Outputs Y

Inference

Unseen Data:

3	3
2	1
1	1

1	0	0
1	1	1
-1	1	1
1	-1	0

[≈ ReLU]

1	1

	0	0	0
0		0	0
0	0		0
0	0	0	

1	0	1	1	0
1	1	1	0	0
1	0	-1	0	1

[≈ ReLU]

1	1

	0	0
0		0
0	0	

1	1	0	0
0	1	-1	-1

Training

-4	7
10	5

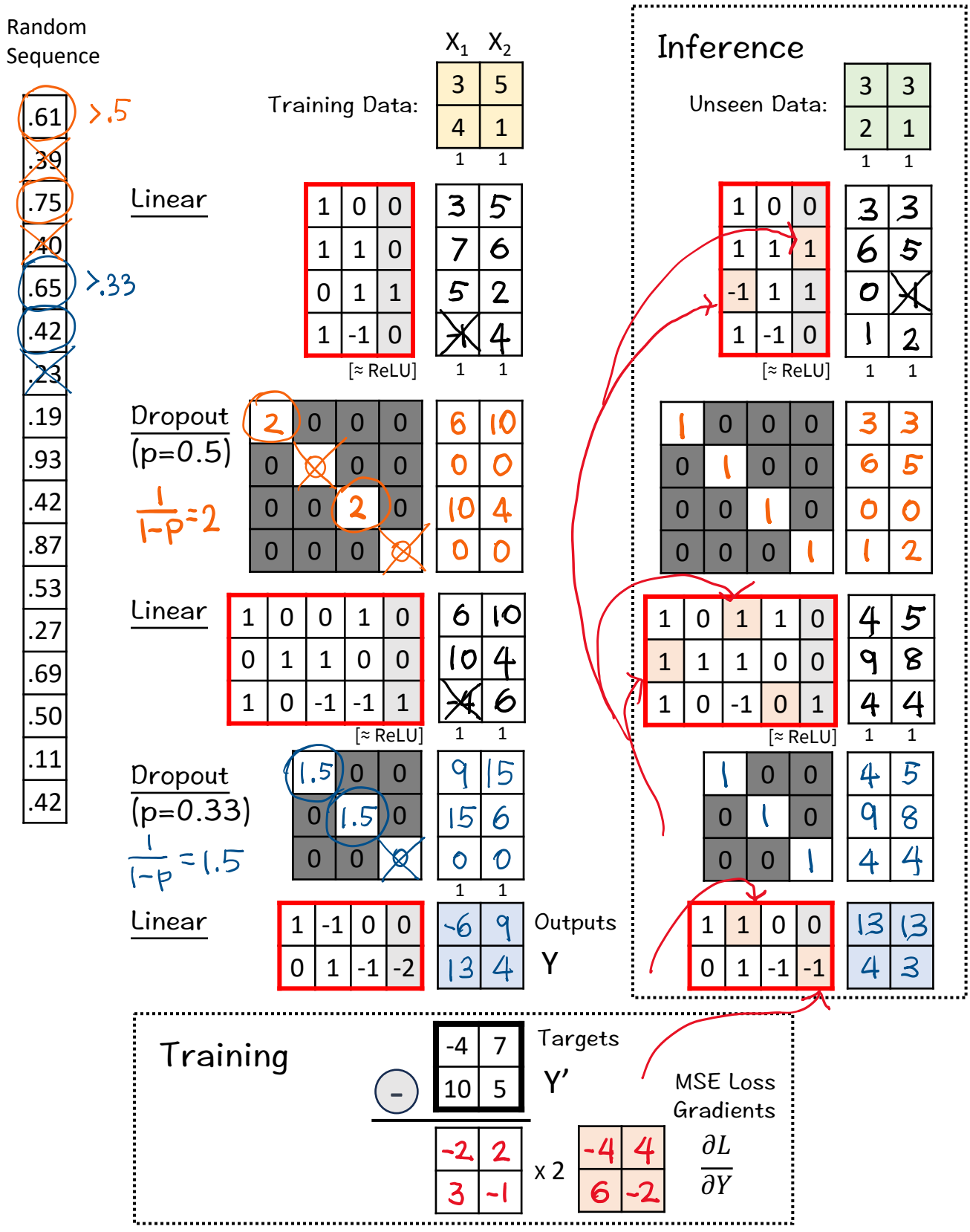
Targets Y'

MSE Loss Gradients $\frac{\partial L}{\partial Y}$

x 2

1.19.24

Dropout



Backpropagation

X

2
1
3

1

Layer 1

1	-1	1	-5	-1	ReLU ≈	0
1	1	0	0	3		3
0	1	1	1	5		5
1	0	1	-2	3		3

1

Layer 2

1	-1	1	0	0	2	ReLU ≈	2
0	1	-1	1	3	4		4

1

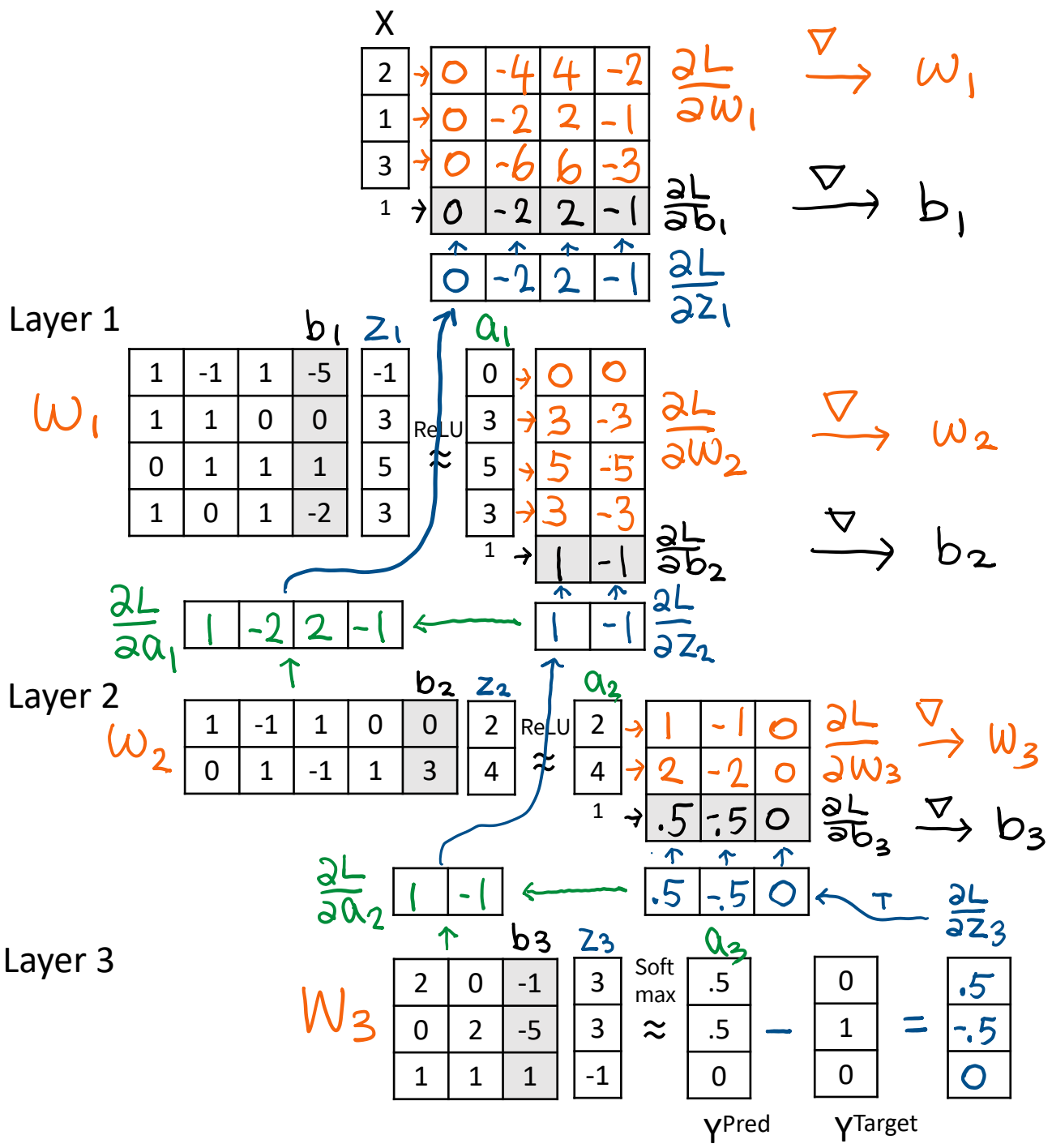
Layer 3

2	0	-1	3	Soft max ≈	.5	0
0	2	-5	3		.5	1
1	1	1	-1		0	0

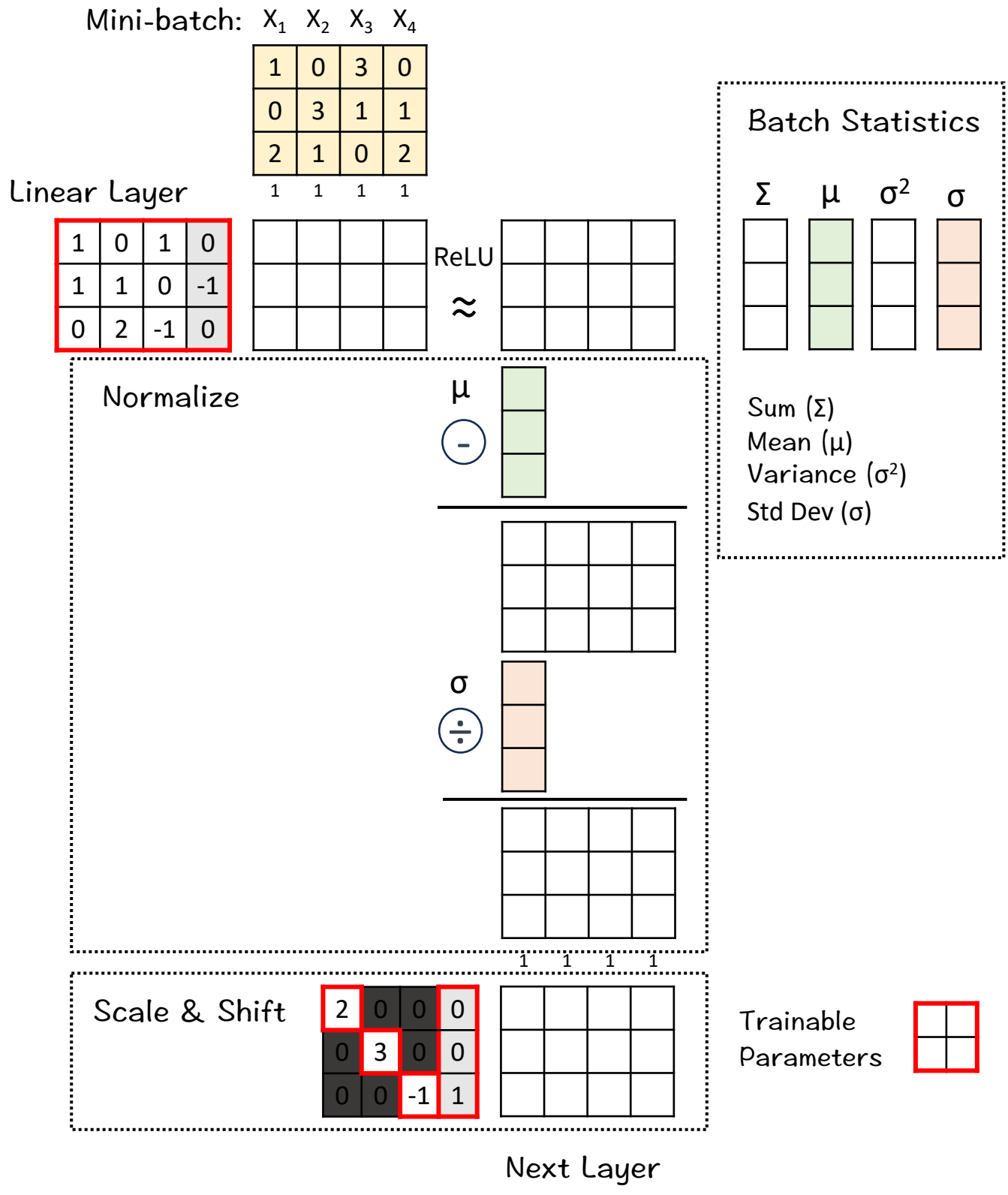
y^{Pred} y^{Target}

L: Cross-Entropy Loss

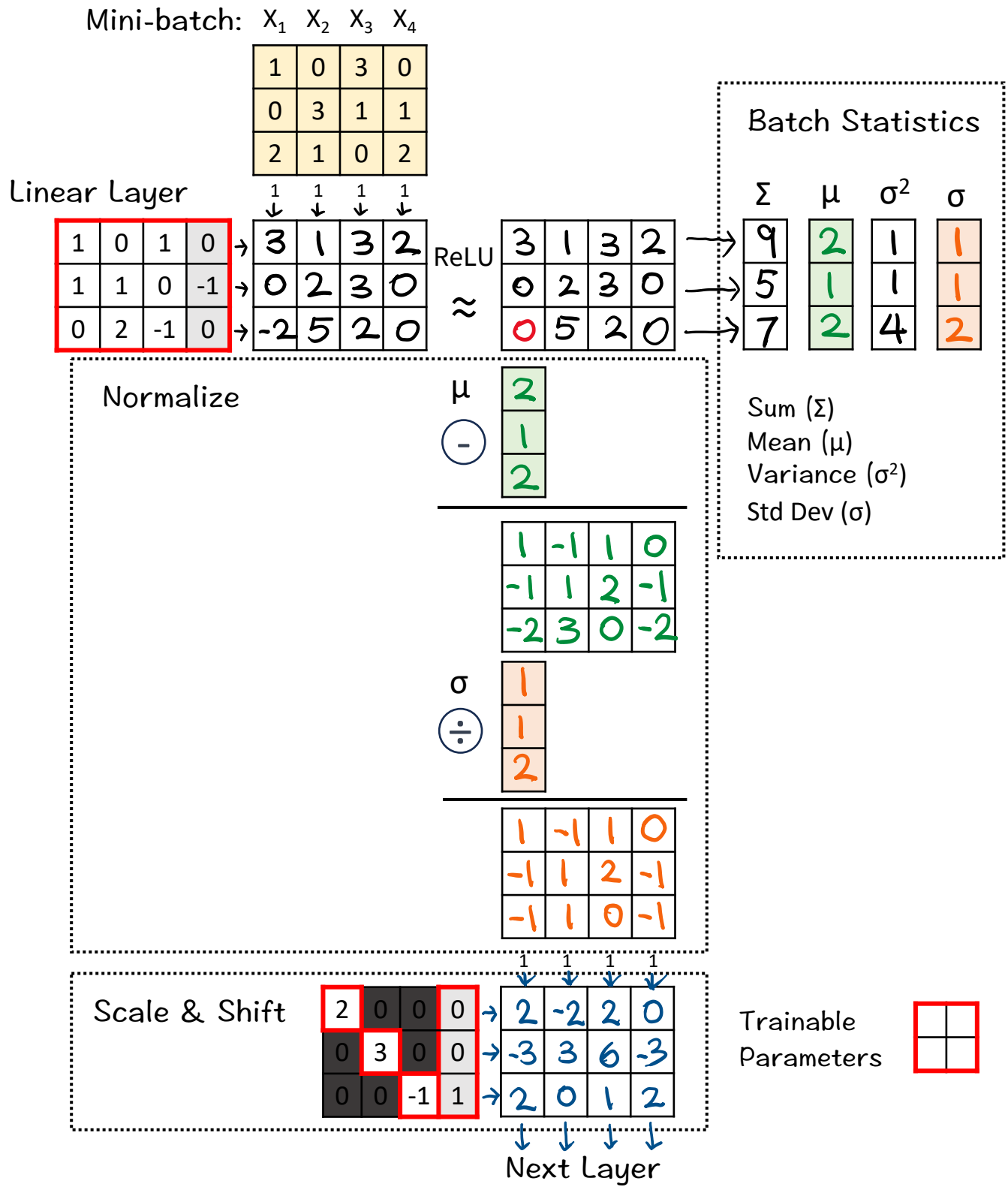
Backpropagation



Batch Normalization



Batch Normalization



SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i		
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1	b
$a_i\ y_i$							2

Decision Boundary

$b + \sum a_i\ y_i\ K$

sign

RBF

x'_j	2	1
	4	1
	3	1

L2 distance $||^2$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

x_i			$(x - x')^2$			$\sum$	$\sqrt{\ }$	$-\gamma\|\|^2$ (-0.1)	$e^{\square}$
1	2	1							
0	4	3							
0	3	0							
0	3	1							
5	0	3							
1	4	3							

a_i	20	10	10	10	0	10	b
$a_i\ y_i$						-1	

sign

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i		
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

13		
25		
12		
15		
19		
27		

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1
$a_i\ y_i$						2

b

sign

Decision Boundary

$b + \sum a_i\ y_i\ K$

RBF

x'_j	2
	4
	3

L2 distance $||^2$

x_i			$(x - x')^2$			$\sum$	$\sqrt{\ }$	$-\gamma\|\|^2$ (-0.1)	$e^{\square}$
1	2	1							
0	4	3							
0	3	0							
0	3	1							
5	0	3							
1	4	3							

$-\gamma\|\|^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	
1	
1	

a_i	20	10	10	10	0	10	b
$a_i\ y_i$						-1	

sign

Linear SVM

y_i		x_i		
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

x'_j

x'_1	x'_2
2	1
4	1
3	1

Kernel Matrix

$K(x_i, x'_j)$

x_1	13	2
x_2	25	7
x_3	12	3
x_4	15	4
x_5	19	8
x_6	27	8

a_i	2	0	1	1	1	1	b
$a_i\ y_i$						2	
sign							

Decision Boundary

$b + \sum a_i\ y_i\ K$

RBF

x'_j

2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ $\sum$ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	
0	4	3	
0	3	0	
0	3	1	
5	0	3	
1	4	3	

1
1
1

a_i	20	10	10	10	0	10	b
$a_i\ y_i$						-1	
sign							

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i		
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

13	2
25	7
12	3
15	4
19	8
27	8

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2
sign							

Decision Boundary

$b + \sum a_i y_i K$

RBF

	x'_j
	2
	4
	3

L2 distance $||^2$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1
1
1

x_i			$(x - x')^2$	$\sum$	$\sqrt{\ }$	$-\gamma||^2$ (-0.1)	$e^{\square}$
1	2	1					
0	4	3					
0	3	0					
0	3	1					
5	0	3					
1	4	3					

a_i	20	10	10	10	0	10	b
$a_i y_i$						-1	
sign							

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i					
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

$$K(x_i, x'_j)$$

a_i	2	0	1	1	1	1	b
$a_i\ y_i$	2	0	1	-1	1	-1	2

17	5

sign

Decision Boundary

$$b + \sum a_i\ y_i\ K$$

RBF

x'_j
2
4
3

L2 distance $||^2$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1
1
1

x_i | $(x - x')^2$ | $\sum$ | $\sqrt{\ }$ | $-\gamma||^2$ (-0.1) | $e^{\square}$

1	2	1
0	4	3
0	3	0
0	3	1
5	0	3
1	4	3

a_i	20	10	10	10	0	10	b
$a_i\ y_i$							-1

sign

Linear SVM

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1
$a_i y_i$	2	0	1	-1	1	-1

b = 2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
	17	5
sign	+	+

Decision Boundary

$b + \sum a_i y_i K$

RBF

x'_j
2
4
3

L2 distance $||^2$

x_i				$(x - x')^2$			$\sum$	$\sqrt{\quad}$	$-\gamma||^2$ (-0.1)	$e^{\square}$
1	2	1								
0	4	3								
0	3	0								
0	3	1								
5	0	3								
1	4	3								

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1
1
1

a_i	20	10	10	10	0	10	b
$a_i y_i$							-1
sign							

SVM

Linear

x'_j: $x'_1 = [2, 4, 3]$, $x'_2 = [1, 1, 1]$

y_i		x_i			$K(x_i, x'_1)$	$K(x_i, x'_2)$
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

$K(x_i, x'_j)$

							b
a_i	2	0	1	1	1	1	
$a_i y_i$	2	0	1	-1	1	-1	2

→ 17, 5

sign: +, +

Decision Boundary

$b + \sum a_i y_i K$

RBF

x'_j: [2, 4, 3] and [1, 1, 1]

L2 distance $||^2$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

x_i			$(x - x')^2$			$\sum$	$\sqrt{\ }$	$-\gamma||^2$ (-0.1)	$e^{\square}$
1	2	1	1	4	4				
0	4	3	4	0	0				
0	3	0	0	1	0				
0	3	1	4	1	4				
5	0	3	9	16	0				
1	4	3	1	0	0				

							b
a_i	20	10	10	10	0	10	
$a_i y_i$							-1

sign

Linear SVM

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1	b	
$a_i y_i$	2	0	1	-1	1	-1	2 →	17 / 5
sign								+ / +

Decision Boundary

$b + \sum a_i y_i K$

RBF

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

x'_j

2		
4		
3		

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1
0	4	3
0	3	0
0	3	1
5	0	3
1	4	3

1	4	4
4	0	0
0	1	0
4	1	4
9	16	0
1	0	0

a_i	20	10	10	10	0	10	b	
$a_i y_i$							-1	
sign								

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

a_i	2	0	1	1	1	1	b
$a_i\ y_i$	2	0	1	-1	1	-1	2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
→	17	5
sign	+	+

Kernel Matrix

$K(x_i, x'_j)$

Decision Boundary

$b + \sum a_i\ y_i\ K$

RBF

x'_j
2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3
0	4	3	4	0	0	4	2
0	3	0	0	1	0	1	1
0	3	1	4	1	4	9	3
5	0	3	9	16	0	25	5
1	4	3	1	0	0	1	1

a_i	20	10	10	10	0	10	b
$a_i\ y_i$							-1
sign							

1
1
1

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i			$K(x_i, x'_j)$	
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2

	17	5
sign	+	+

Decision Boundary

$$b + \sum a_i y_i K$$

RBF

L2 distance $||^2$

x'_j

2	4	3
4	1	1
3	1	1

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

x_i	$(x - x')^2$	Σ	$\sqrt{\ }$	$-\gamma\|\|^2$ (-0.1)	$e^{\square}$
1 2 1	1 4 4	9	3	-0.3	
0 4 3	4 0 0	4	2	-0.2	
0 3 0	0 1 0	1	1	-0.1	
0 3 1	4 1 4	9	3	-0.3	
5 0 3	9 16 0	25	5	-0.5	
1 4 3	1 0 0	1	1	-0.1	

$-\gamma\|\|^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

a_i	20	10	10	10	0	10	b
$a_i y_i$							-1
sign							

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i			$K(x_i, x'_j)$	
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2

	17	5
sign	+	+

Decision Boundary

$$b + \sum a_i y_i K$$

RBF

x'_j

2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

x_i	$(x - x')^2$	Σ	$\sqrt{\ }$	$-\gamma\|\|^2$ (-0.1)	$e^{\square}$
1 2 1	1 4 4	9	3	-0.3	0.7
0 4 3	4 0 0	4	2	-0.2	0.8
0 3 0	0 1 0	1	1	-0.1	0.9
0 3 1	4 1 4	9	3	-0.3	0.7
5 0 3	9 16 0	25	5	-0.5	0.6
1 4 3	1 0 0	1	1	-0.1	0.9

$-\gamma\|\|^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

a_i	20	10	10	10	0	10	b
$a_i y_i$							-1
sign							

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i			$K(x_i, x'_j)$	
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

							b
a_i	2	0	1	1	1	1	
$a_i y_i$	2	0	1	-1	1	-1	2

$a_i y_i$	17	5
sign	+	+

Decision Boundary

$$b + \sum a_i y_i K$$

RBF

x'_j

2		
4		
3		

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

$\gamma = 0.1$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0			
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4			
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1			
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0			
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4			
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4			

							b
a_i	20	10	10	10	0	10	
$a_i y_i$							-1

$a_i y_i$		
sign		

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1	b		
$a_i y_i$	2	0	1	-1	1	-1	2	→ 17	5
sign								+	+

Decision Boundary

$b + \sum a_i y_i K$

RBF

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

x'_j

2		
4		
3		

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13

a_i	20	10	10	10	0	10	b	
$a_i y_i$							-1	
sign								

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

a_i	2	0	1	1	1	1	b
$a_i\ y_i$	2	0	1	-1	1	-1	2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
→	17	5
sign	+	+

Kernel Matrix

$K(x_i, x'_j)$

Decision Boundary

$b + \sum a_i\ y_i\ K$

RBF

x'_j
2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1	1		1	
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14	3		1	
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6	2		1	
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5	2		1	
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21	4		1	
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13	3		1	

a_i	20	10	10	10	0	10	b
$a_i\ y_i$							-1
sign							

SVM

Linear

x'_j	x'_1	x'_2
1	2	1
2	4	1
3	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
→	17	5
sign	+	+

Kernel Matrix

$K(x_i, x'_j)$

Decision Boundary

$b + \sum a_i y_i K$

RBF

L2 distance $||^2$

x'_j

2		
4		
3		

x_i $(x - x')^2$ $\sum$ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

x_i	$(x - x')^2$	$\sum$	$\sqrt{\ }$	$-\gamma\|\|^2$ (-0.1)	$e^{\square}$
1 2 1	1 4 4	9	3	-0.3	0.7
0 4 3	4 0 0	4	2	-0.2	0.8
0 3 0	0 1 0	1	1	-0.1	0.9
0 3 1	4 1 4	9	3	-0.3	0.7
5 0 3	9 16 0	25	5	-0.5	0.6
1 4 3	1 0 0	1	1	-0.1	0.9

$-\gamma\|\|^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

0 1 0	1	1	-0.1				
1 9 4	14	3	-0.3				
1 4 1	6	2	-0.2				
1 4 0	5	2	-0.2				
16 1 4	21	4	-0.4				
0 9 4	13	3	-0.3				

a_i	20	10	10	10	0	10	b
$a_i y_i$							-1
sign							

SVM

Linear

	x'_1	x'_2
x'_j	2	1
	4	1
	3	1

y_i		x_i			$K(x_i, x'_1)$	$K(x_i, x'_2)$
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

$K(x_i, x'_j)$

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2 → 17 / 5
sign							+ / +

Decision Boundary

$b + \sum a_i y_i K$

RBF

x'_j

2	
4	
3	

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1	1	-0.1	0.9	1						
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14	3	-0.3	0.7	1						
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6	2	-0.2	0.8	1						
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5	2	-0.2	0.8	1						
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21	4	-0.4	0.67	1						
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13	3	-0.3	0.9	1						

a_i	20	10	10	10	0	10	b	
$a_i y_i$							-1	
sign								

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i				$x_i \cdot x'_1$	$x_i \cdot x'_2$
1	x_1	1	2	1	13	2
-1	x_2	0	4	3	25	7
1	x_3	0	3	0	12	3
-1	x_4	0	3	1	15	4
1	x_5	5	0	3	19	8
-1	x_6	1	4	3	27	8

Kernel Matrix

$$K(x_i, x'_j)$$

a_i	2	0	1	1	1	1	b
$a_i y_i$	2	0	1	-1	1	-1	2

$a_i y_i$	17	5
sign	+	+

Decision Boundary

$$b + \sum a_i y_i K$$

RBF

x'_j
2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ $\sum$ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1	1	-0.1	0.9	1	1	1
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14	3	-0.3	0.7	1	1	1
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6	2	-0.2	0.8			
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5	2	-0.2	0.8			
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21	4	-0.4	0.67			
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13	3	-0.3	0.9			

a_i	20	10	10	10	0	10	b
$a_i y_i$	20	-10	10	-10	0	-10	-1

$a_i y_i$		
sign		

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

a_i	2	0	1	1	1	1
$a_i y_i$	2	0	1	-1	1	-1

b = 2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
$a_i y_i$	17	5
sign	+	+

Kernel Matrix

$K(x_i, x'_j)$

Decision Boundary

$b + \sum a_i y_i K$

RBF

x'_j

	2	
	4	
	3	

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ $e^{\square}$

$\gamma = 0.1$

$-\gamma\|\|^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1	1	-0.1	0.9	1						
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14	3	-0.3	0.7	1						
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6	2	-0.2	0.8	1						
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5	2	-0.2	0.8	1						
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21	4	-0.4	0.67	1						
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13	3	-0.3	0.9	1						

a_i	20	10	10	10	0	10	
$a_i y_i$	20	-10	10	-10	0	-10	-1

b = -1

	-2	1
sign		

SVM

Linear

x'_j	x'_1	x'_2
	2	1
	4	1
	3	1

y_i	x_i			
1	x_1	1	2	1
-1	x_2	0	4	3
1	x_3	0	3	0
-1	x_4	0	3	1
1	x_5	5	0	3
-1	x_6	1	4	3

a_i	2	0	1	1	1	1
$a_i y_i$	2	0	1	-1	1	-1

b = 2

	13	2
	25	7
	12	3
	15	4
	19	8
	27	8
$a_i y_i$	17	5
sign	+	+

Kernel Matrix

$K(x_i, x'_j)$

Decision Boundary

$b + \sum a_i y_i K$

RBF

x'_j
2
4
3

L2 distance $||^2$

x_i $(x - x')^2$ Σ $\sqrt{\ }$ $-\gamma||^2$ (-0.1) $e^{\square}$

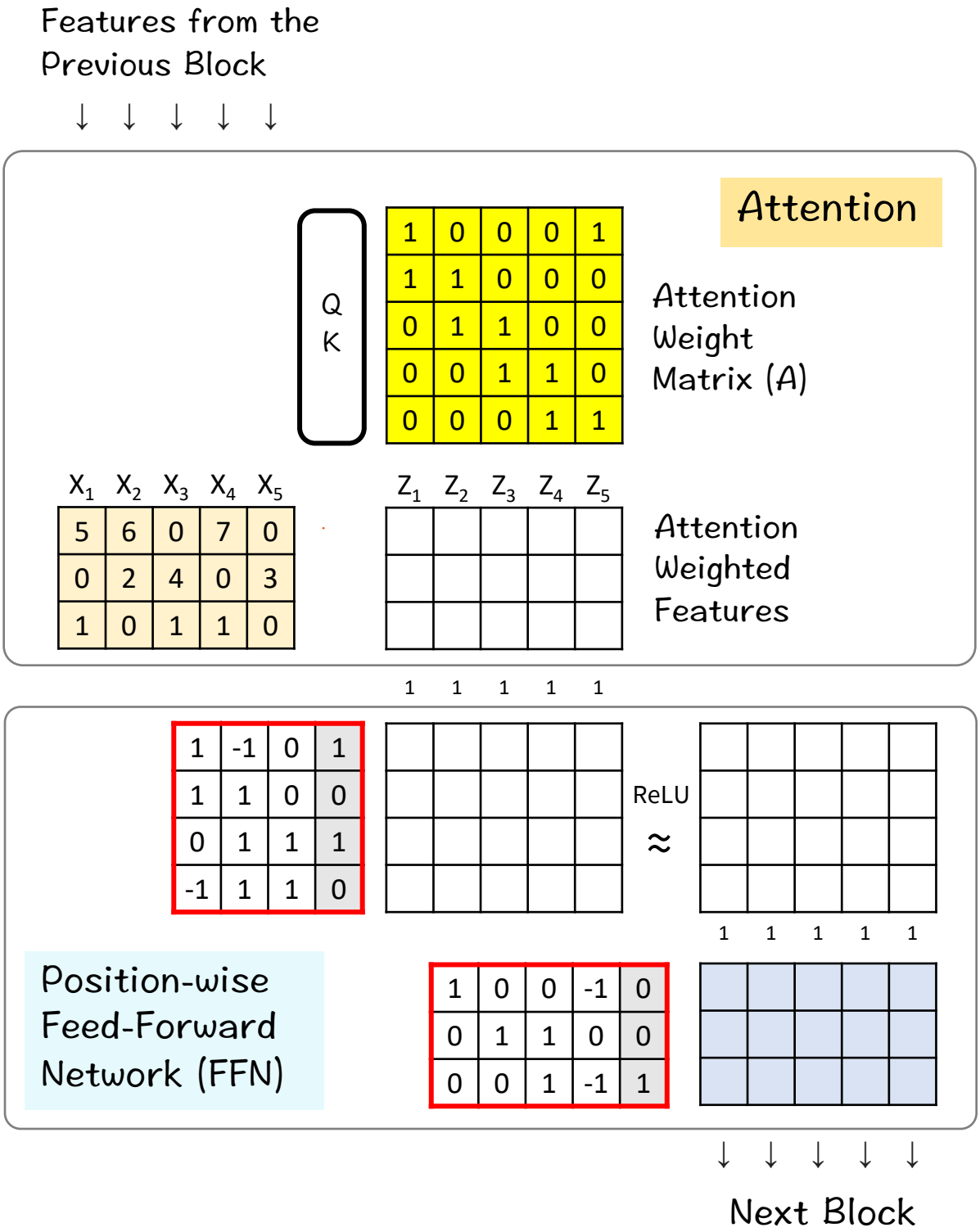
$-\gamma||^2$	$e^{\square}$
-0.5	0.6
-0.4	0.67
-0.3	0.7
-0.2	0.8
-0.1	0.9

1	2	1	1	4	4	9	3	-0.3	0.7	0	1	0	1	1	-0.1	0.9	1	1	1
0	4	3	4	0	0	4	2	-0.2	0.8	1	9	4	14	3	-0.3	0.7	1	1	1
0	3	0	0	1	0	1	1	-0.1	0.9	1	4	1	6	2	-0.2	0.8			
0	3	1	4	1	4	9	3	-0.3	0.7	1	4	0	5	2	-0.2	0.8			
5	0	3	9	16	0	25	5	-0.5	0.6	16	1	4	21	4	-0.4	0.67			
1	4	3	1	0	0	1	1	-0.1	0.9	0	9	4	13	3	-0.3	0.9			

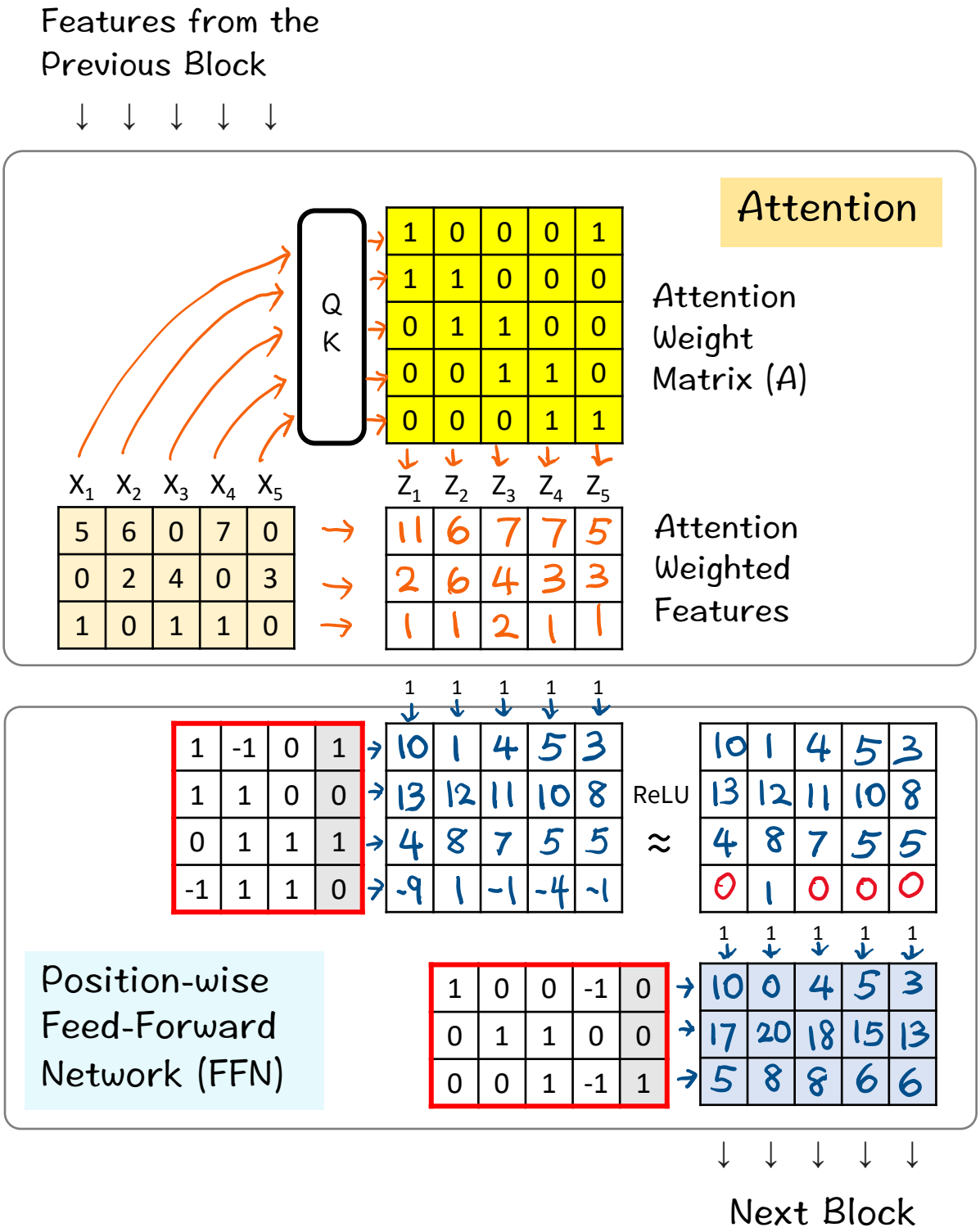
a_i	20	10	10	10	0	10	b = -1
$a_i y_i$	20	-10	10	-10	0	-10	-1

	-2	1
sign	−	+

Transformer



Transformer



Self Attention

Features

x_1	x_2	x_3	x_4
2	0	0	2
0	1	0	0
0	2	1	0
0	0	1	1
2	0	0	0
1	0	1	1

W_Q

1	1	0	0	0	0
0	1	0	1	0	0
0	0	1	0	1	1

W_K

0	0	1	0	0	0
0	1	0	0	0	0
1	0	0	0	0	-1

W_V

10	0	0	0	0	0
0	0	0	10	0	0
0	10	0	0	0	0

q_1	q_2	q_3	q_4

k_1	k_2	k_3	k_4

v_1	v_2	v_3	v_4

MatMul

(K^TQ)

q_1	q_2	q_3	q_4

k_1^T			
k_2^T			
k_3^T			
k_4^T			

Scale

$\frac{\square}{\sqrt{d_k}}$

Softmax

$e^{\square}$

$\div \Sigma$

MatMul

Attention Weight Matrix (A)

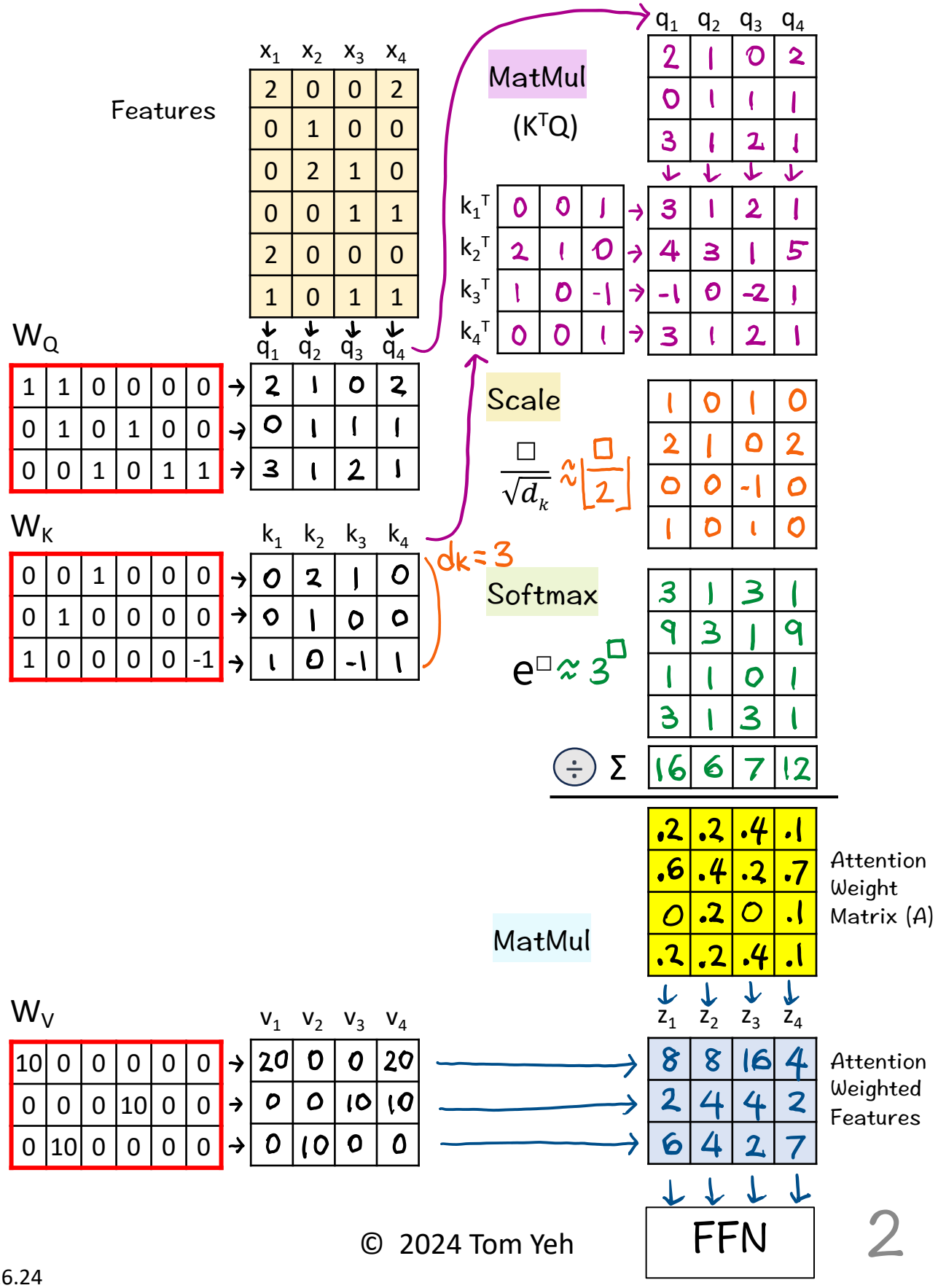
z_1	z_2	z_3	z_4

Attention Weighted Features

FFN

1.16.24

Self Attention



SORA's Diffusion Transformer

Training Video

1	0	2	0	0	1	0	1
0	1	1	0	3	0	4	0

Diffusion Step t = 3

Prompt

"sora is sky"

Text Encoder

0
1
-1

Spacetime Patches (Pixels)

1 1 1 1

Visual Encoder

1	0	-1	0	0
0	1	0	1	1

[ReLU]

Latent

1	0	0	1	0	0
0	1	0	-1	1	0
1	1	0	0	1	0
0	2	0	1	0	2

1

Sampled Noise

0	2	1	-1
-1	0	-2	1

Noised Latent

Adaptive Layer Norm

Q K

Self-Attention

1	0	0	0
1	1	0	0
0	1	1	0
0	0	1	1

1 1 1 1

Predicted Noise

Pointwise FFN

-1	1	-2
0	1	-5

Noise-free Latent

1 1 1 1

Visual Decoder

1	0	1
0	1	0
1	1	0
-1	1	0

[ReLU]

Train

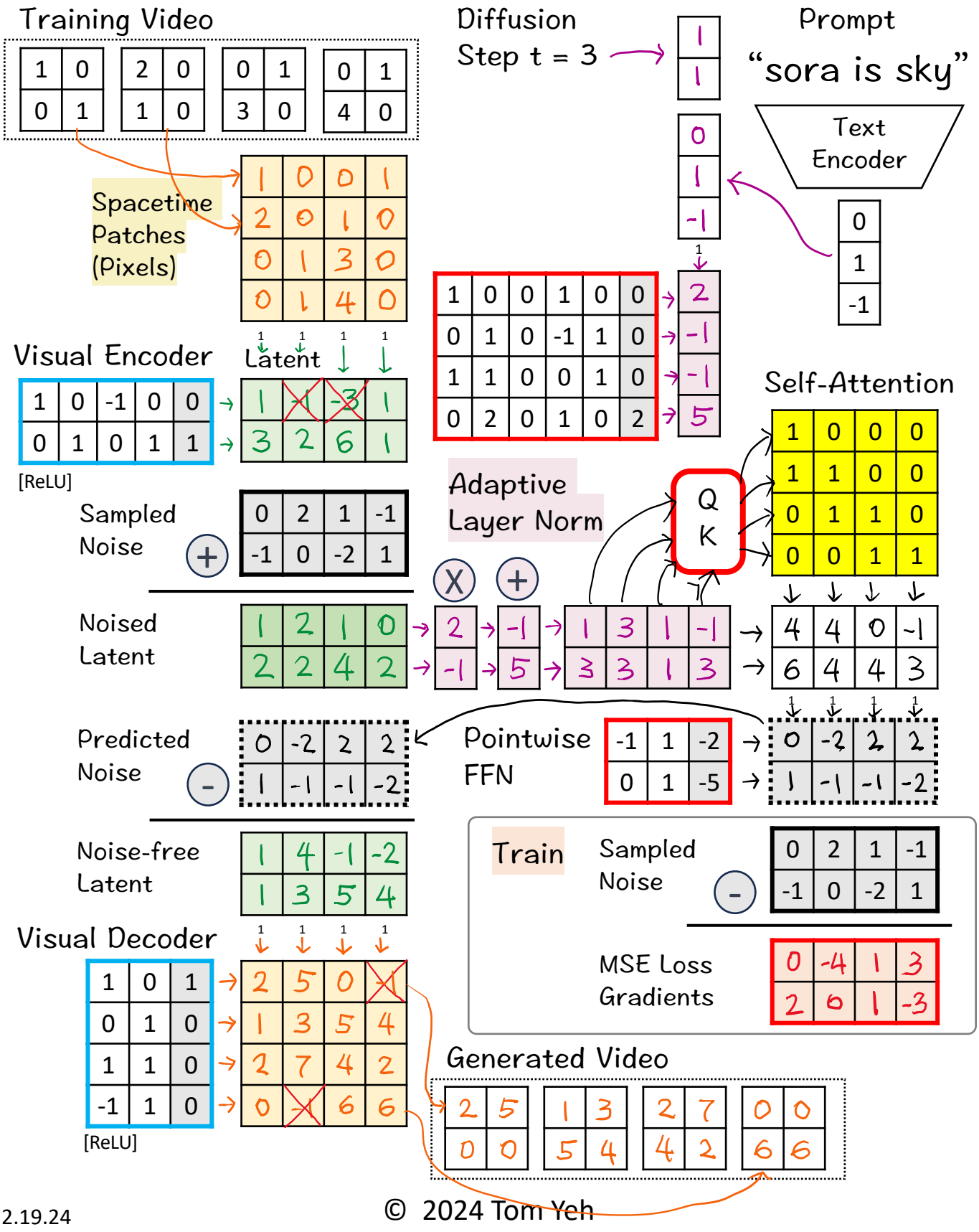
Sampled Noise

0	2	1	-1
-1	0	-2	1

MSE Loss Gradients

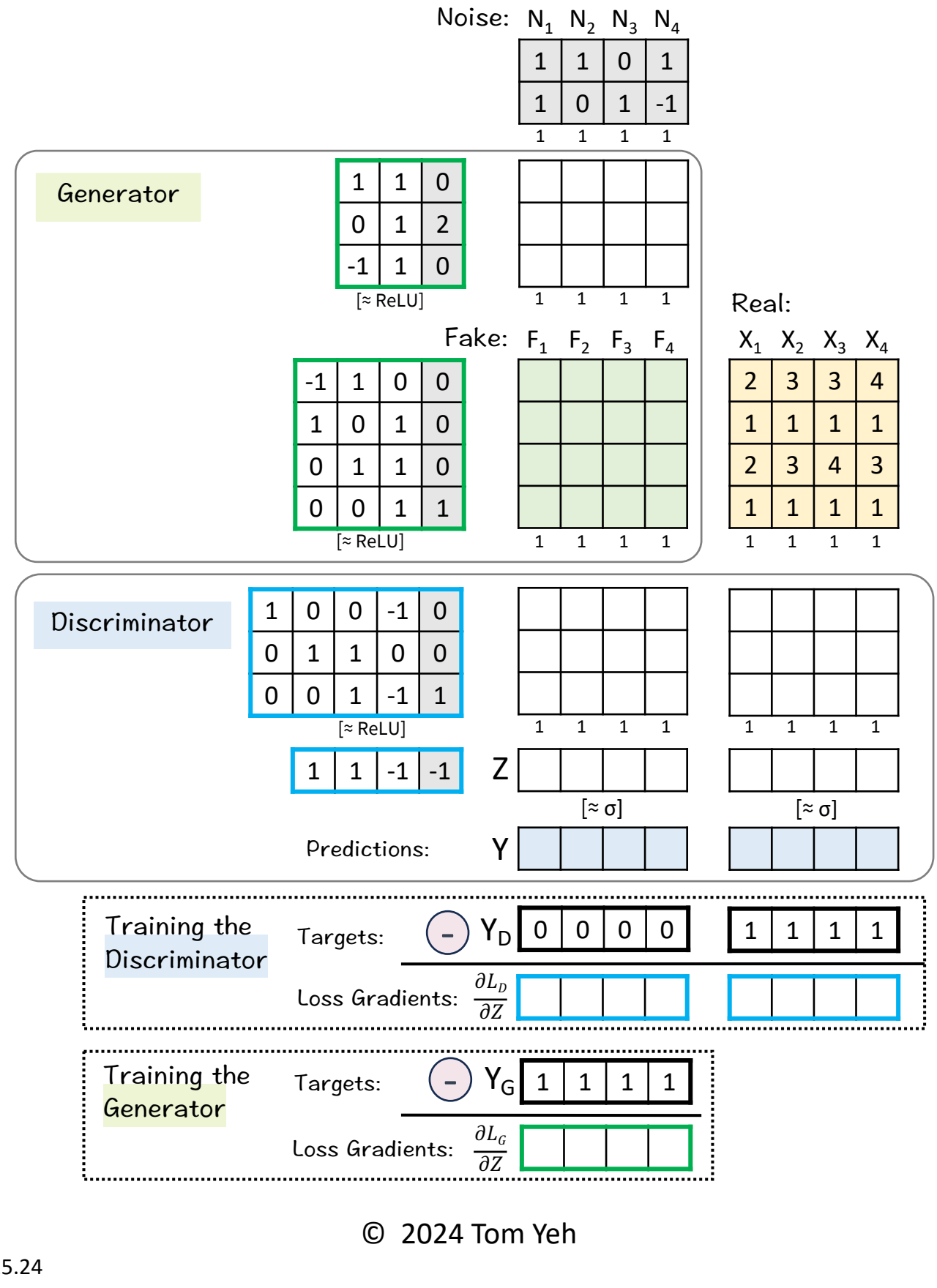
Generated Video

SORA's Diffusion Transformer

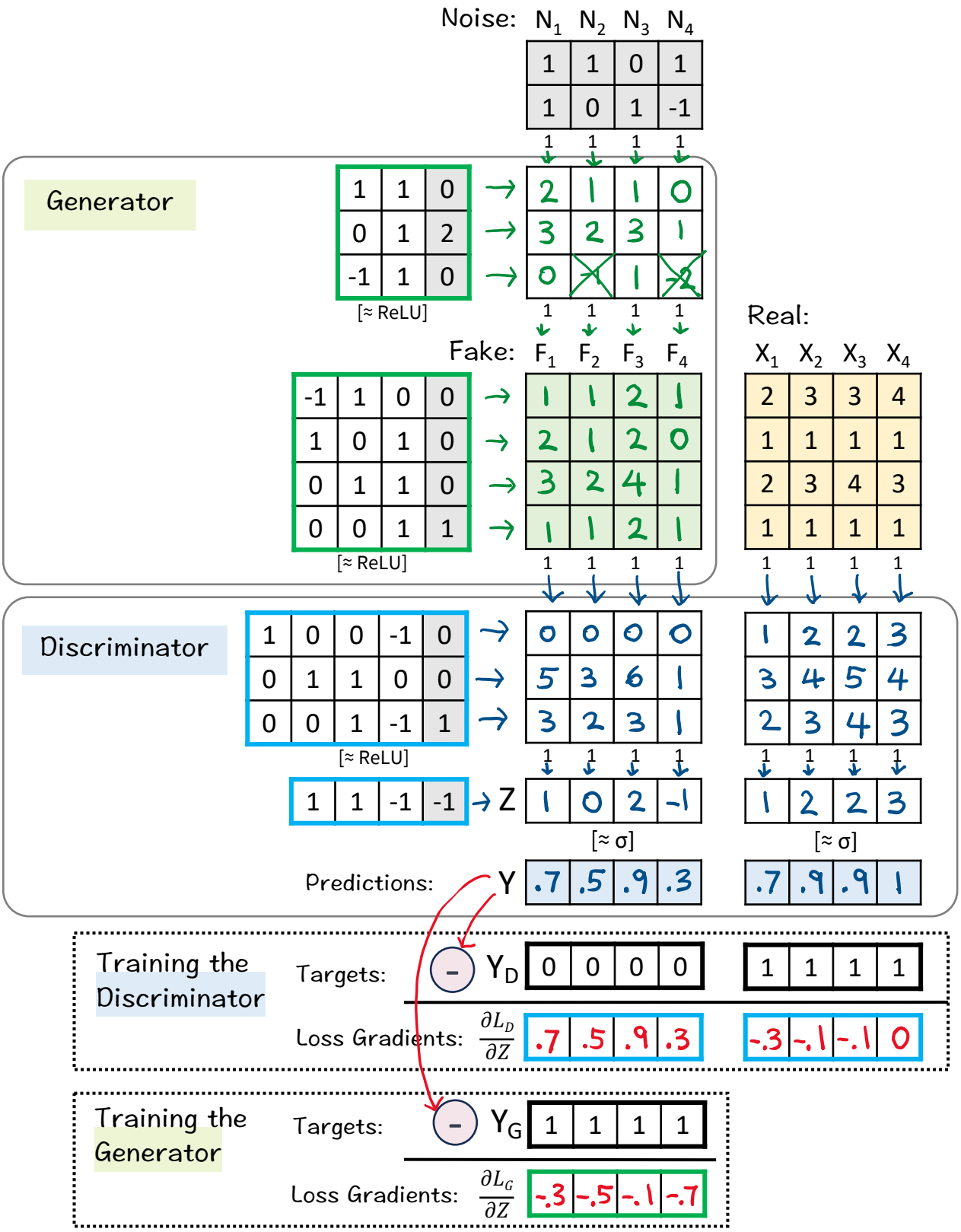


2.19.24

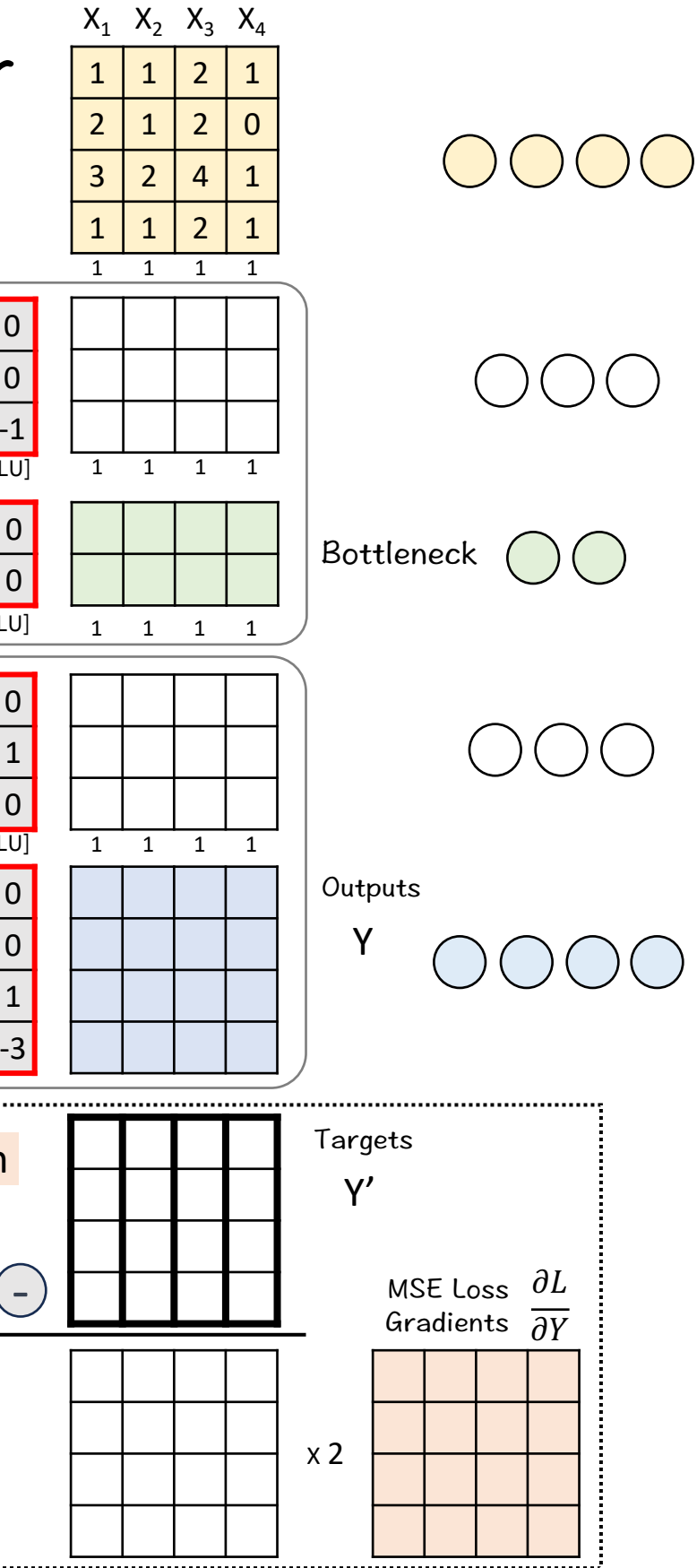
Generative Adversarial Network (GAN)



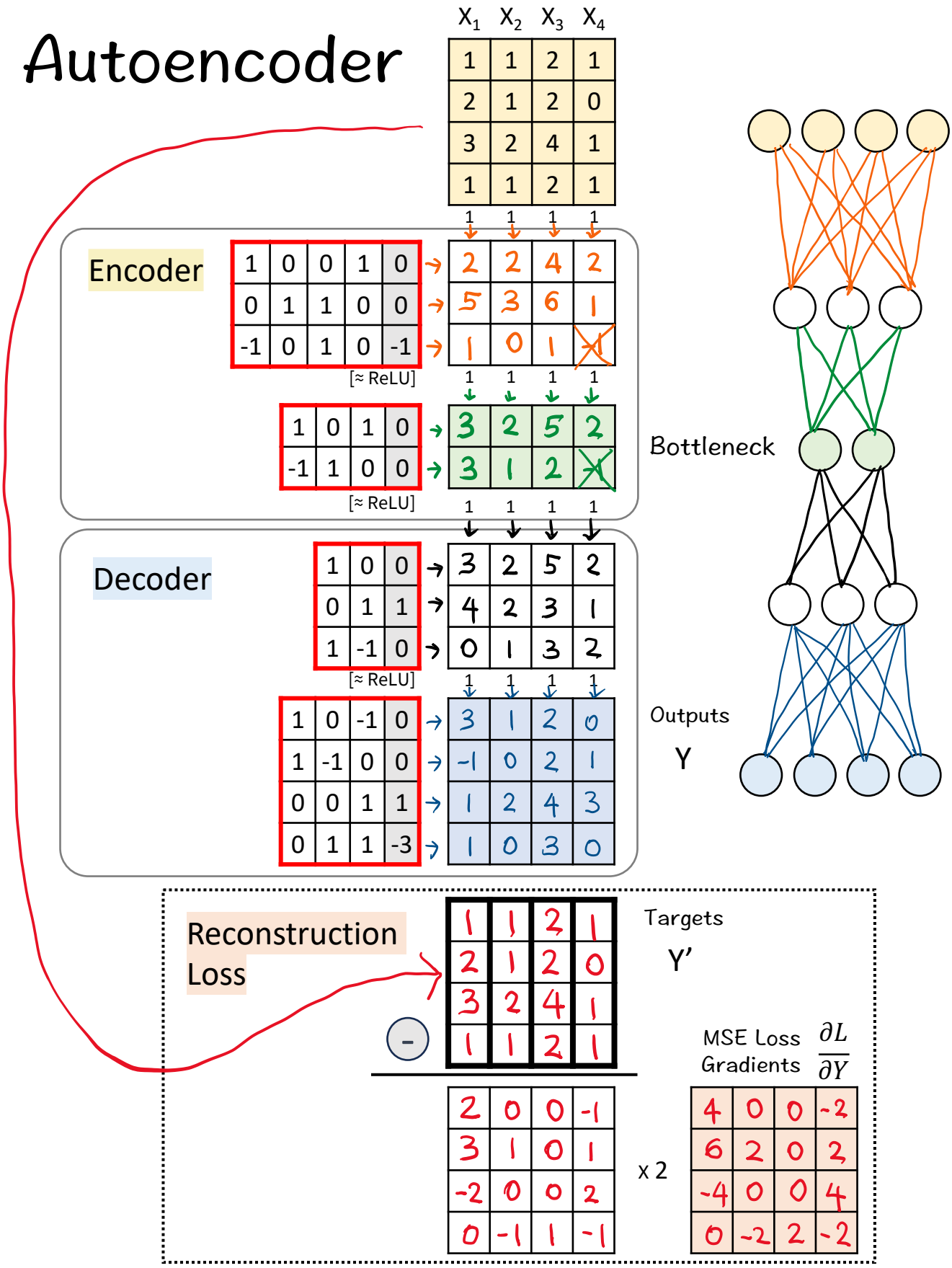
Generative Adversarial Network (GAN)



Autoencoder



Autoencoder



Variational Auto Encoder

X_1	X_2	X_3
1	1	2
2	1	2
3	2	4
1	1	2

1 1 1

Encoder

1	0	0	1	0
0	1	1	0	0
-1	0	1	0	-1

[≈ ReLU]

1 1 1

1	0	1	-3
-1	1	0	0

μ

1	1	0	0
0	1	1	-2

σ

Reparameterization Trick

ε

0.2	1	0.5
0.5	-1	0.2

$\varepsilon \odot \sigma$

$\mu + \varepsilon \odot \sigma$

1 1 1

KL Divergence Loss Gradients (SGVB)

$$= \begin{array}{|c|c|c|}\hline & & \\ \hline & & \\ \hline \end{array} \frac{\partial L}{\partial \mu}$$

$$- \frac{1}{\begin{array}{|c|c|c|}\hline & & \\ \hline & & \\ \hline \end{array}} = \begin{array}{|c|c|c|}\hline & & \\ \hline & & \\ \hline \end{array} \frac{\partial L}{\partial \sigma}$$

σ

Decoder

1	0	0
0	1	1
1	-1	0

[≈ ReLU]

1 1 1

1	0	-1	0
1	-1	0	0
0	0	1	1
0	1	1	-3

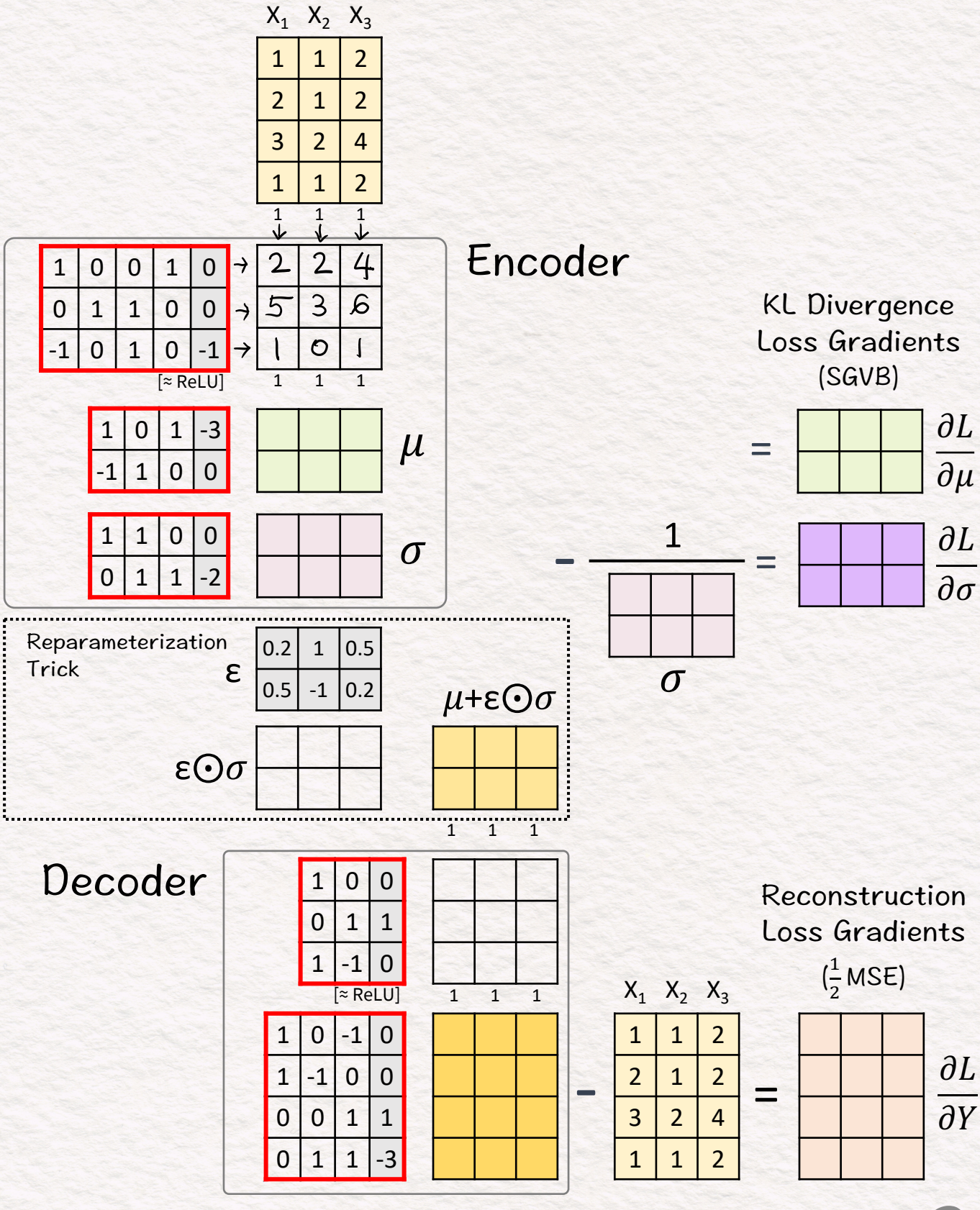
$-$

X_1	X_2	X_3
1	1	2
2	1	2
3	2	4
1	1	2

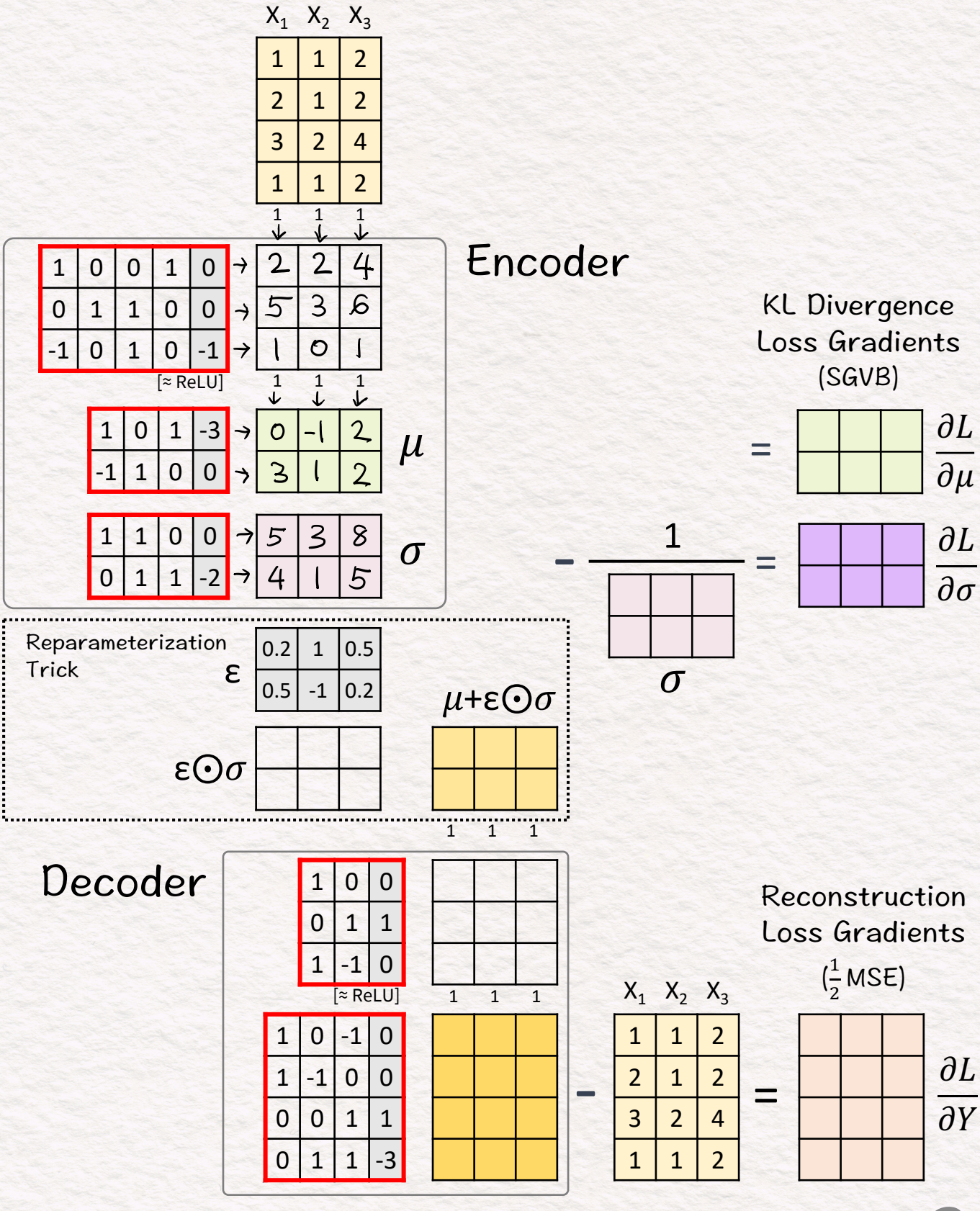
Reconstruction Loss Gradients ($\frac{1}{2}$ MSE)

$$= \begin{array}{|c|c|c|}\hline & & \\ \hline & & \\ \hline & & \\ \hline & & \\ \hline \end{array} \frac{\partial L}{\partial Y}$$

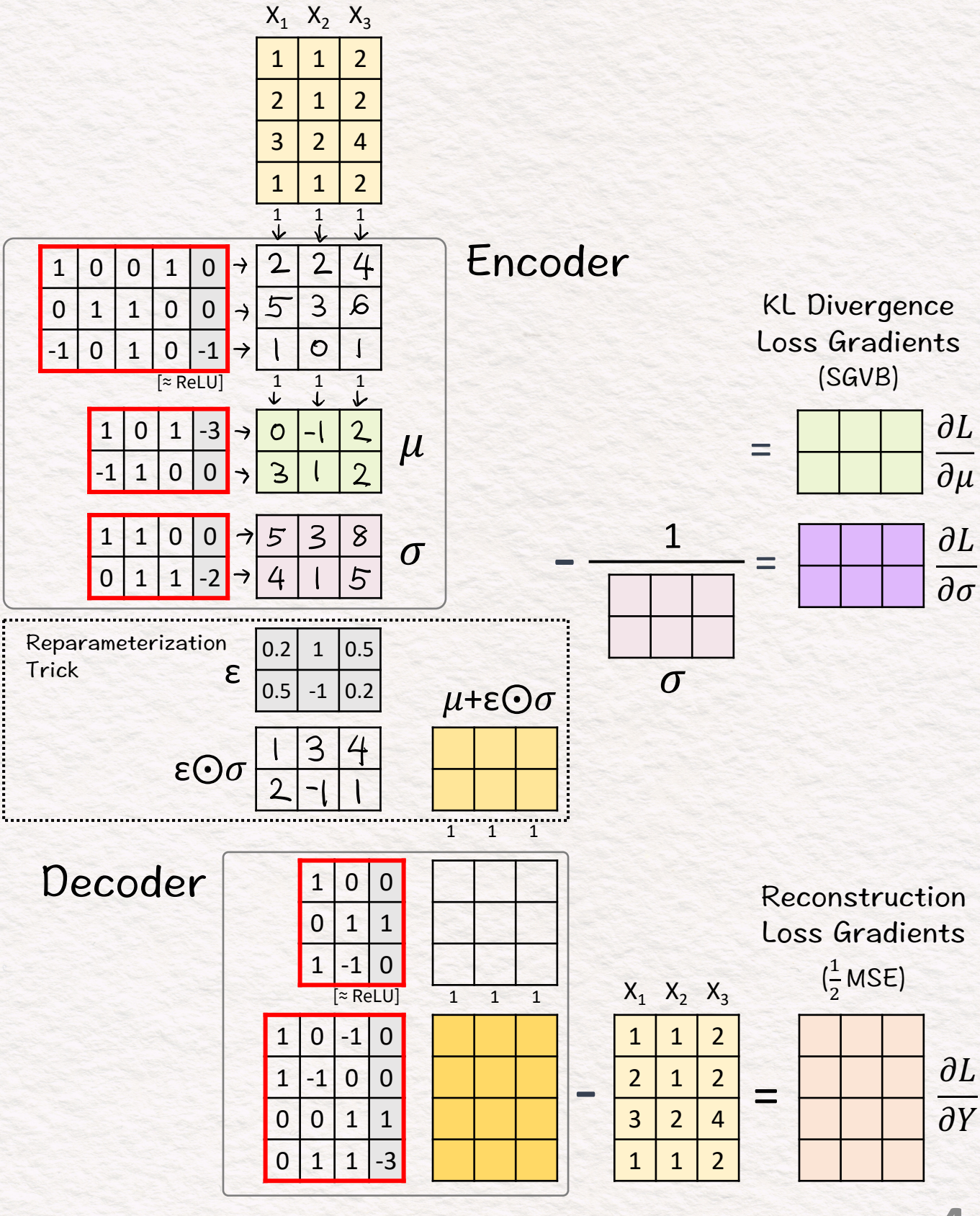
Variational Auto Encoder



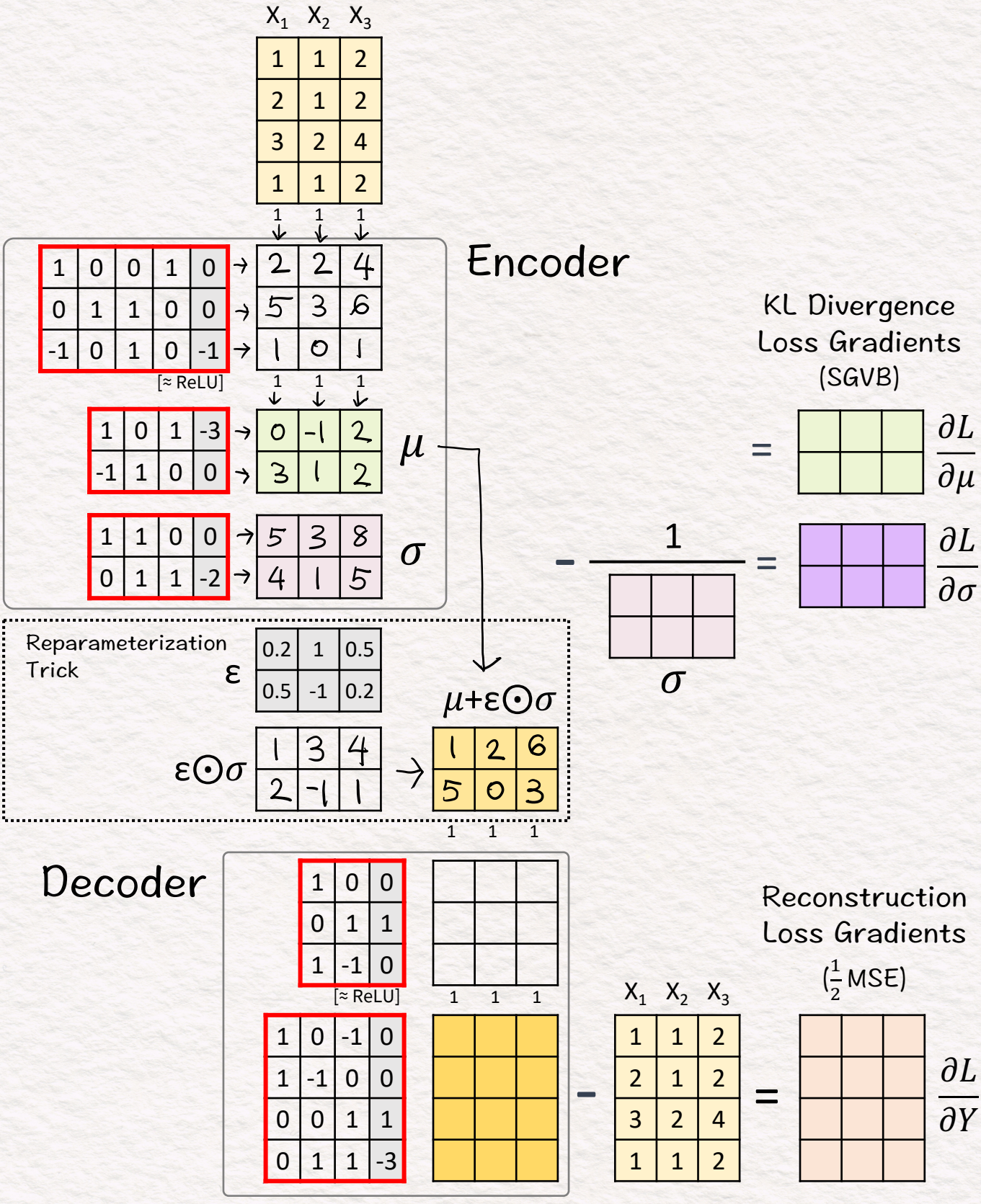
Variational Auto Encoder



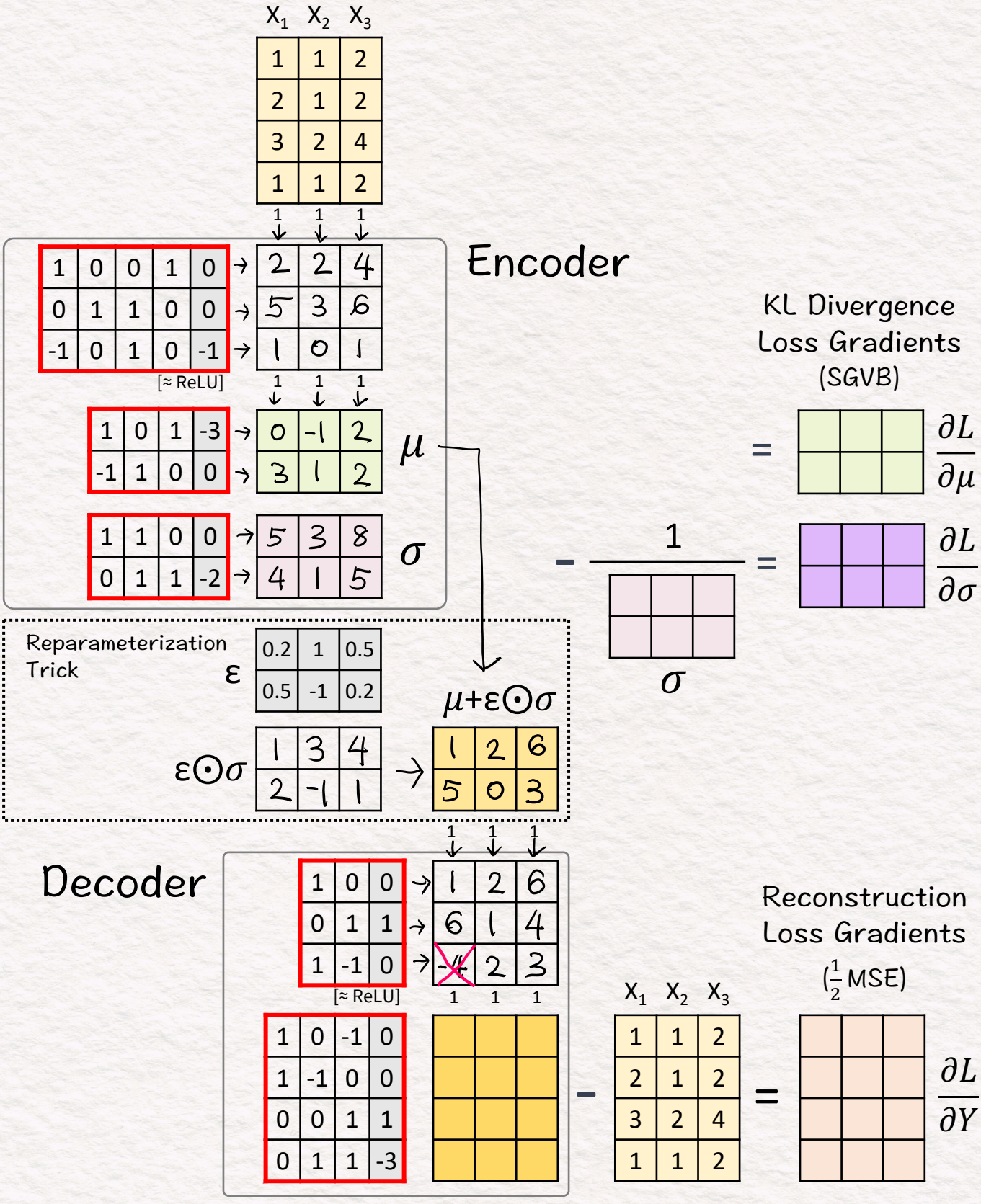
Variational Auto Encoder



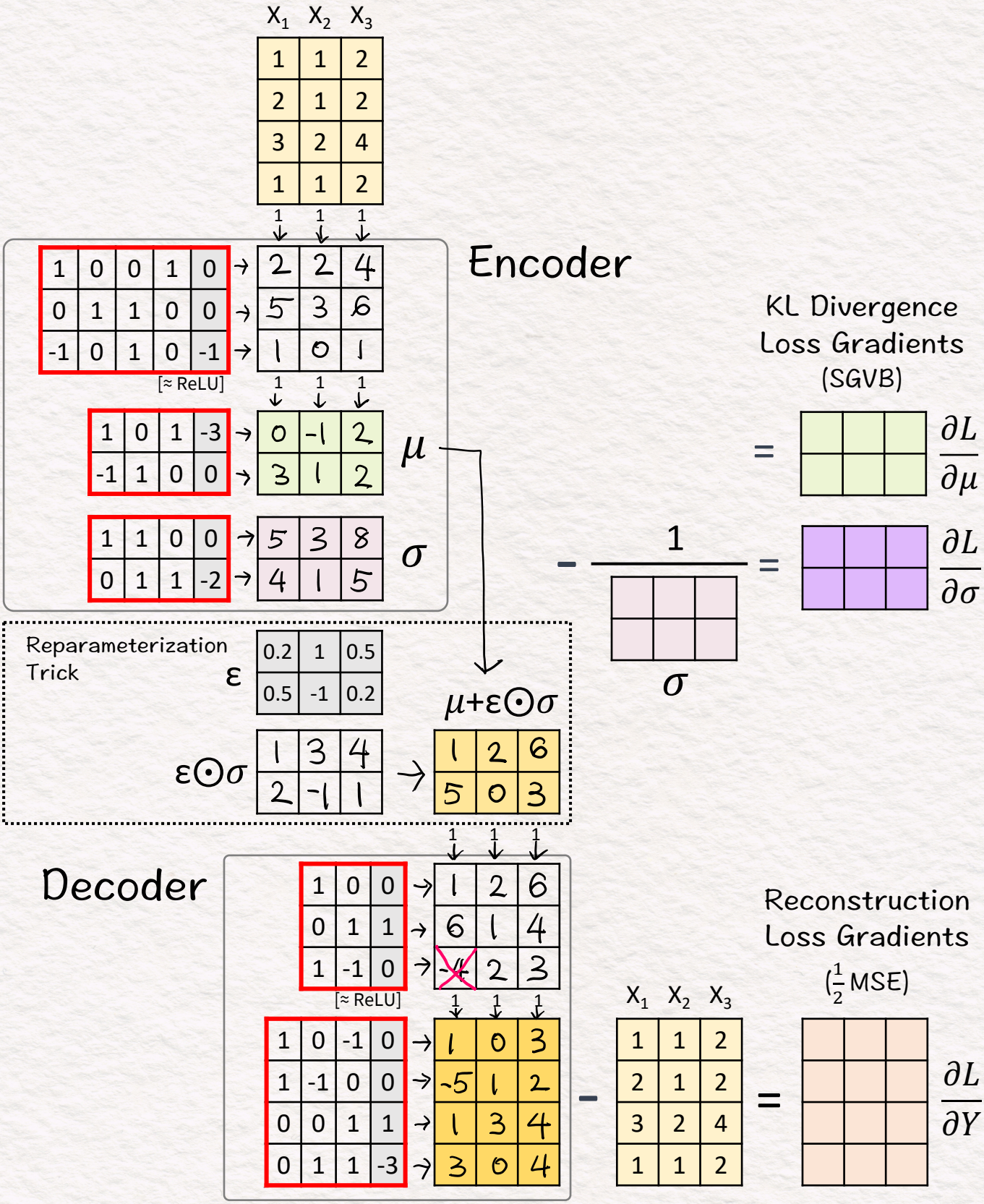
Variational Auto Encoder



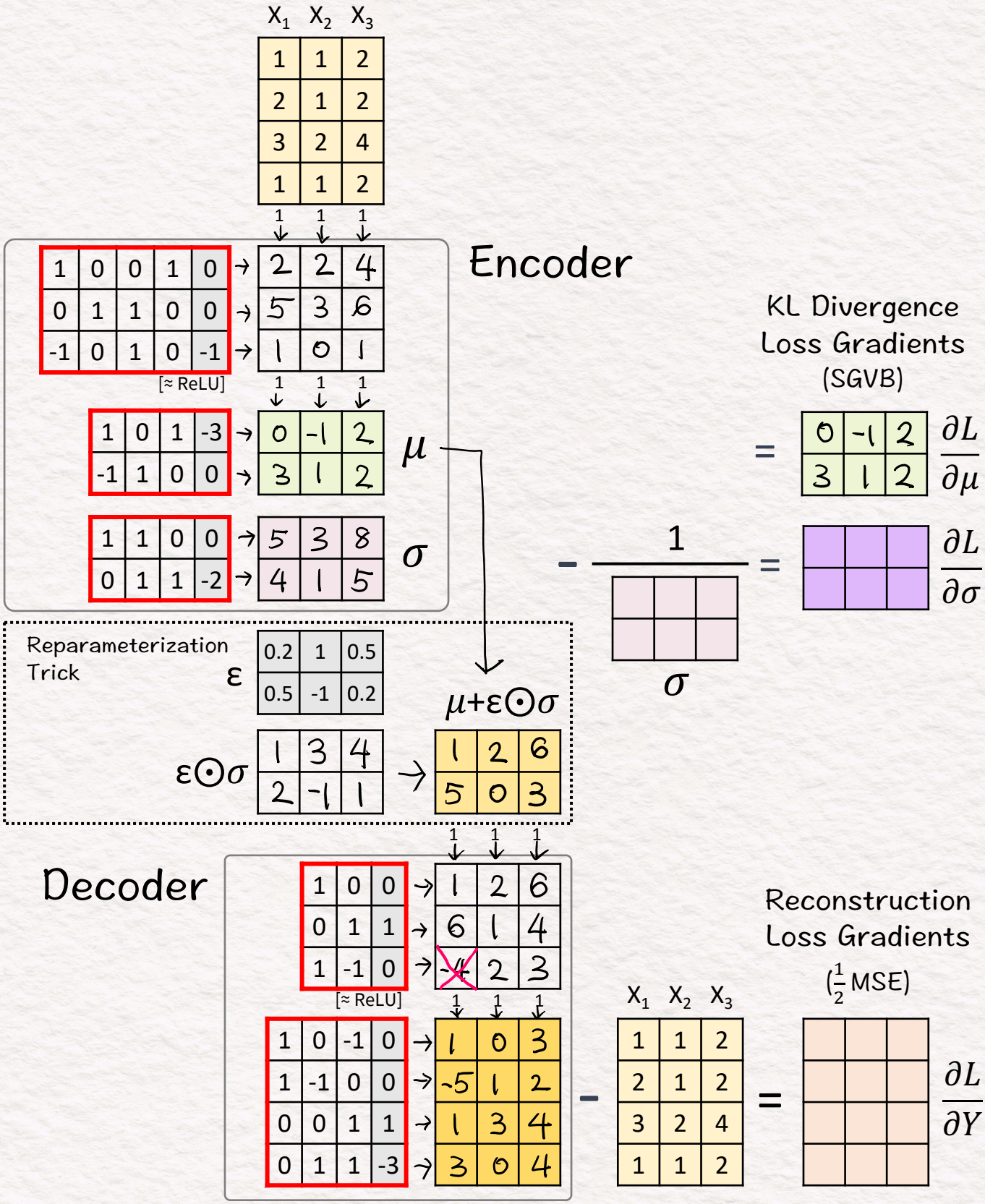
Variational Auto Encoder



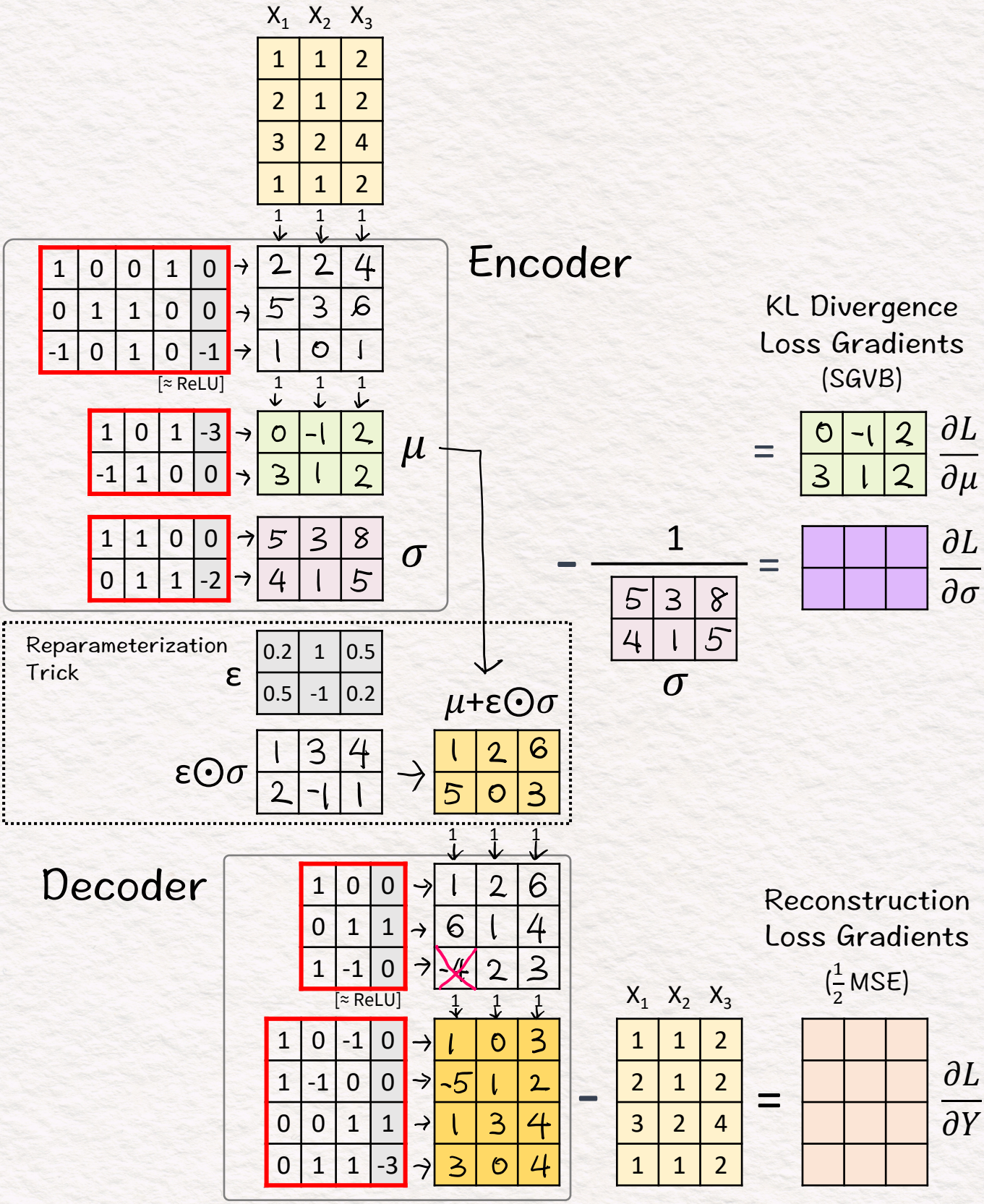
Variational Auto Encoder



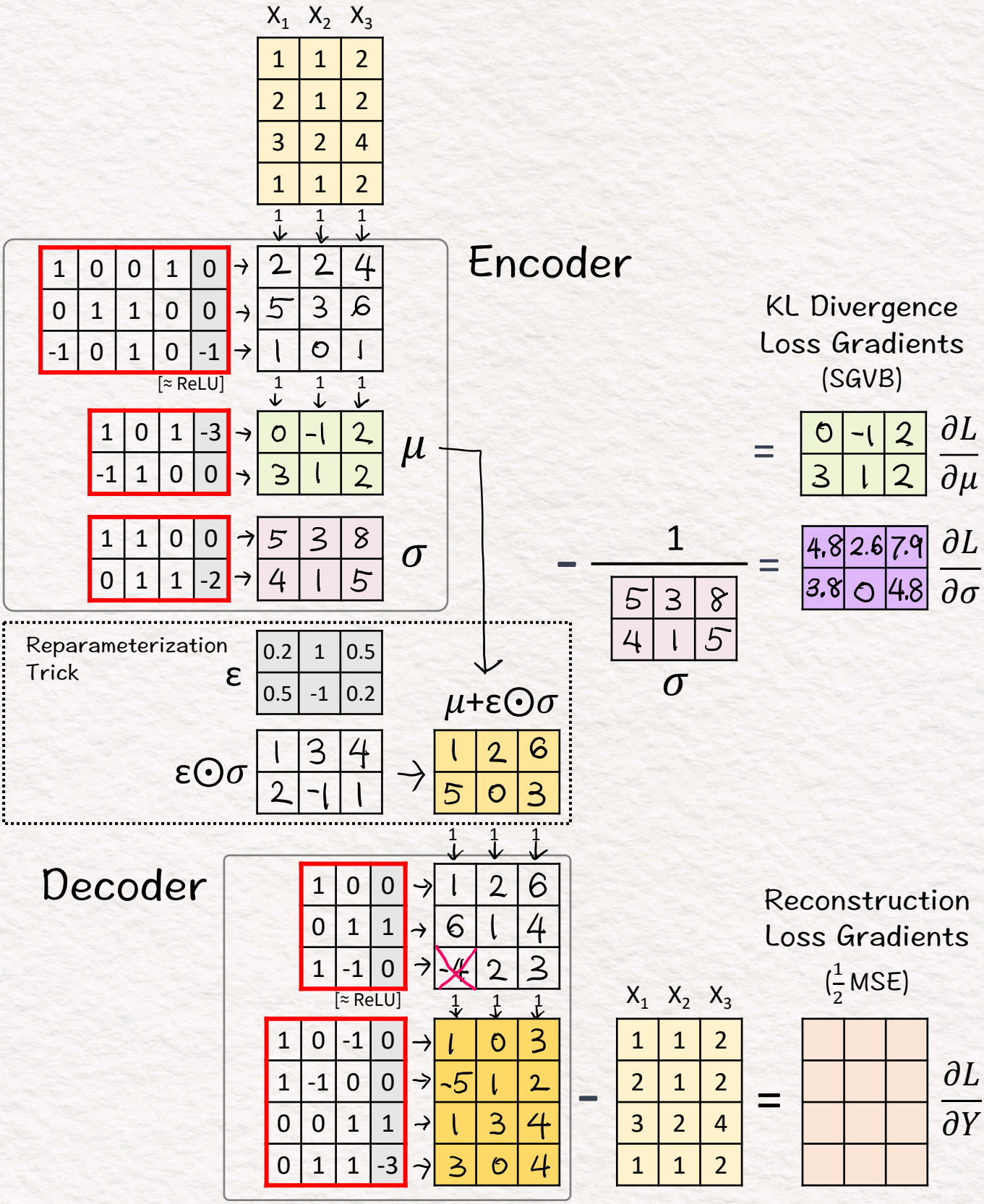
Variational Auto Encoder



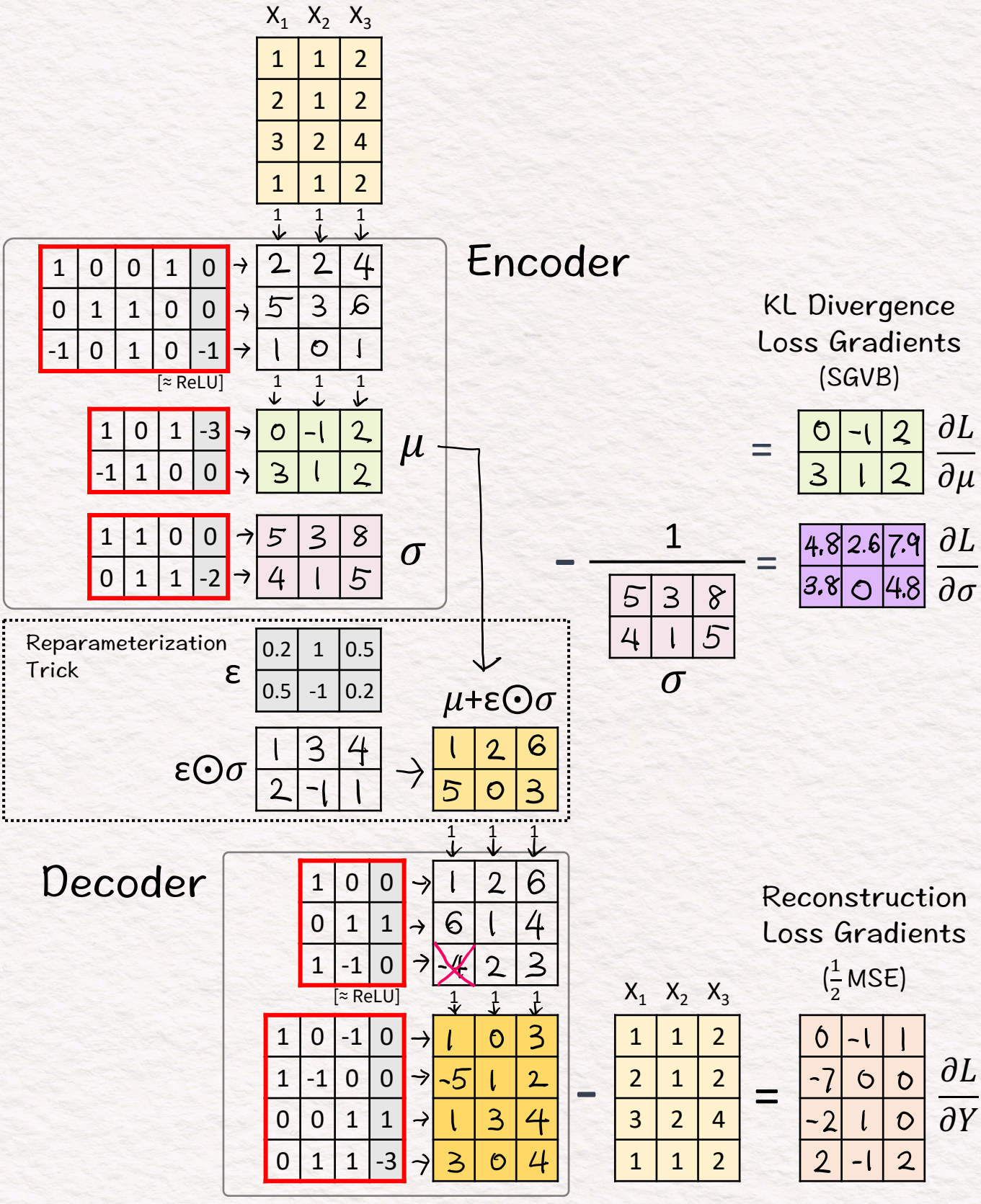
Variational Auto Encoder



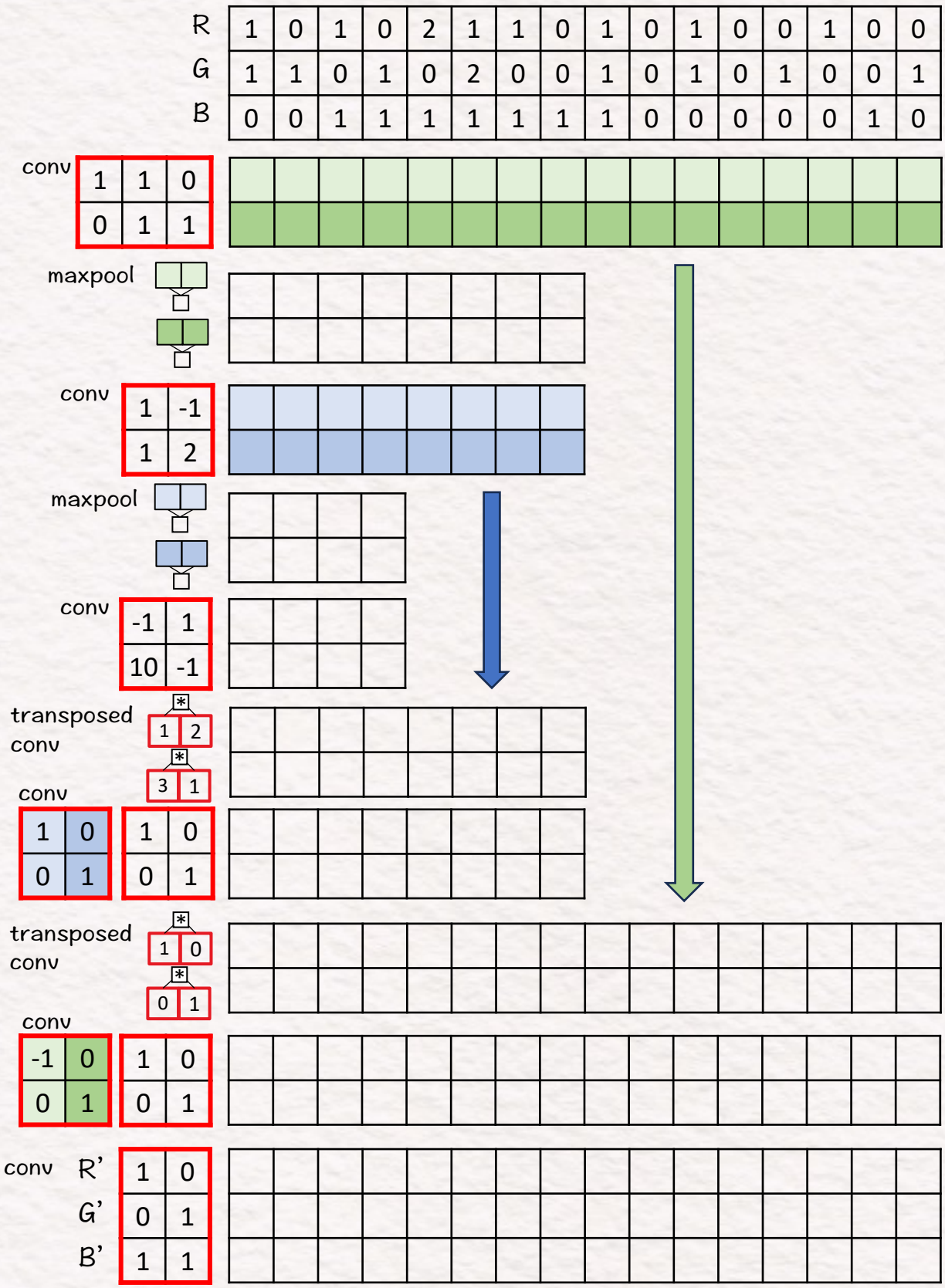
Variational Auto Encoder



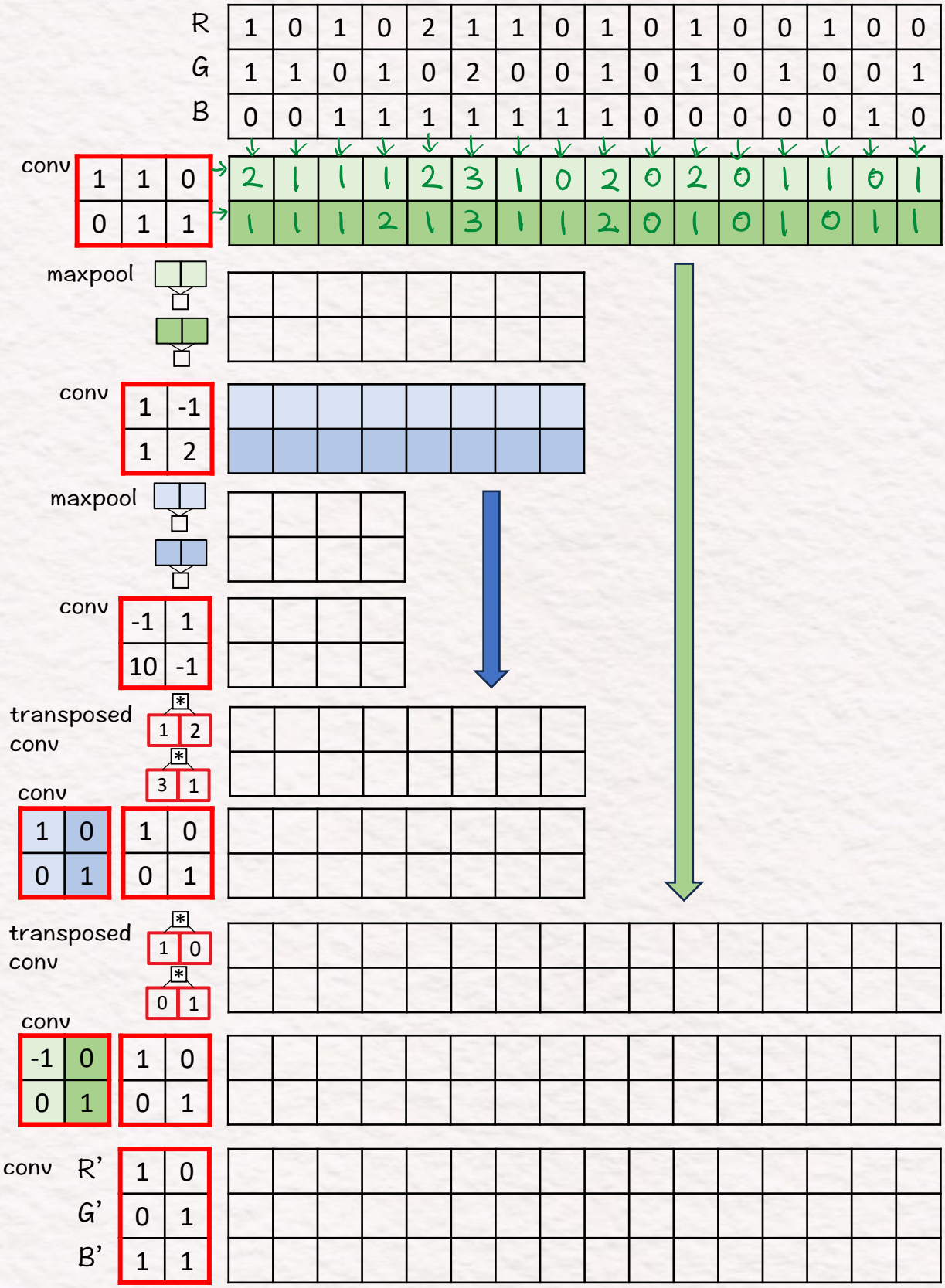
Variational Auto Encoder



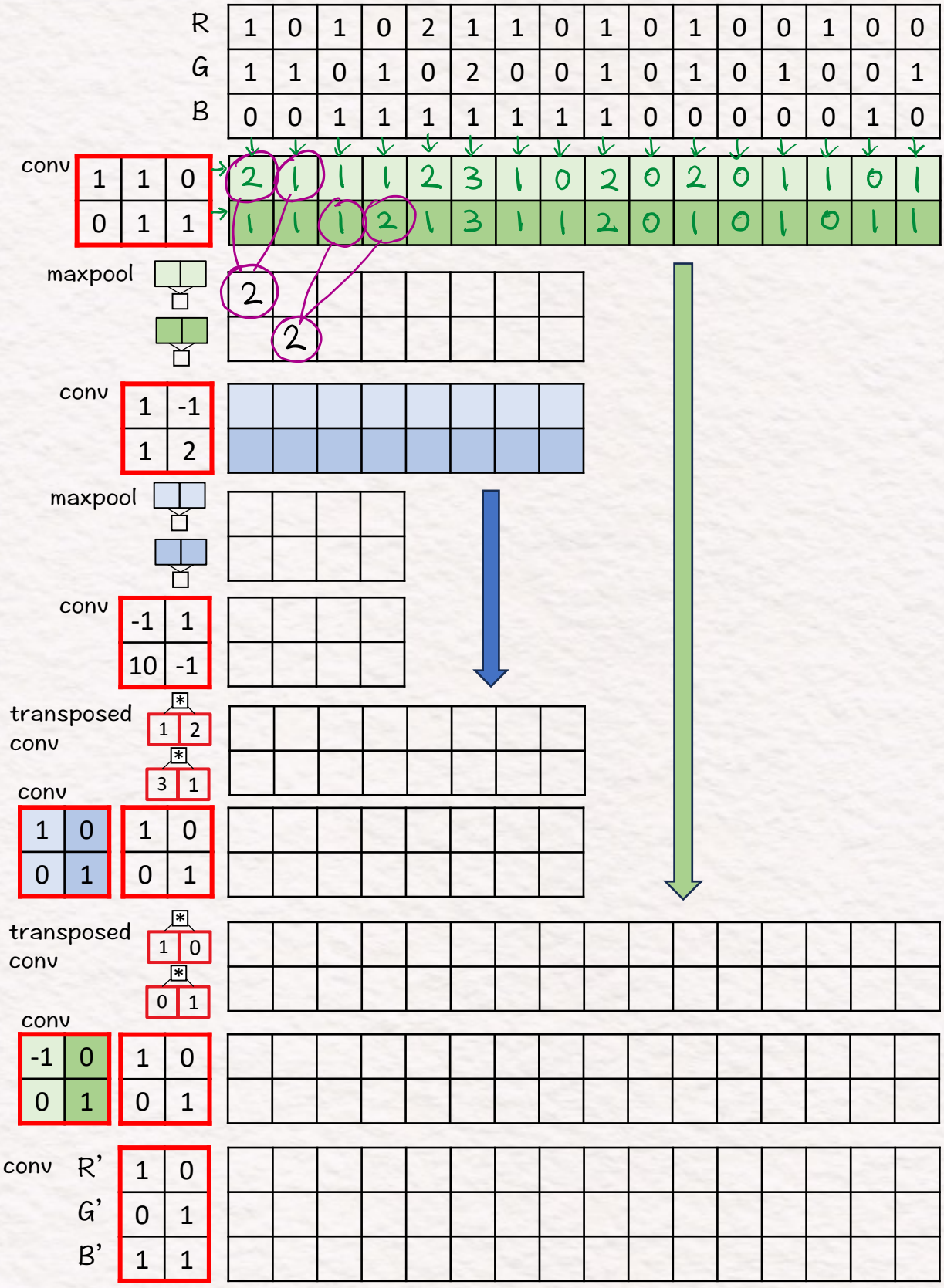
UNet



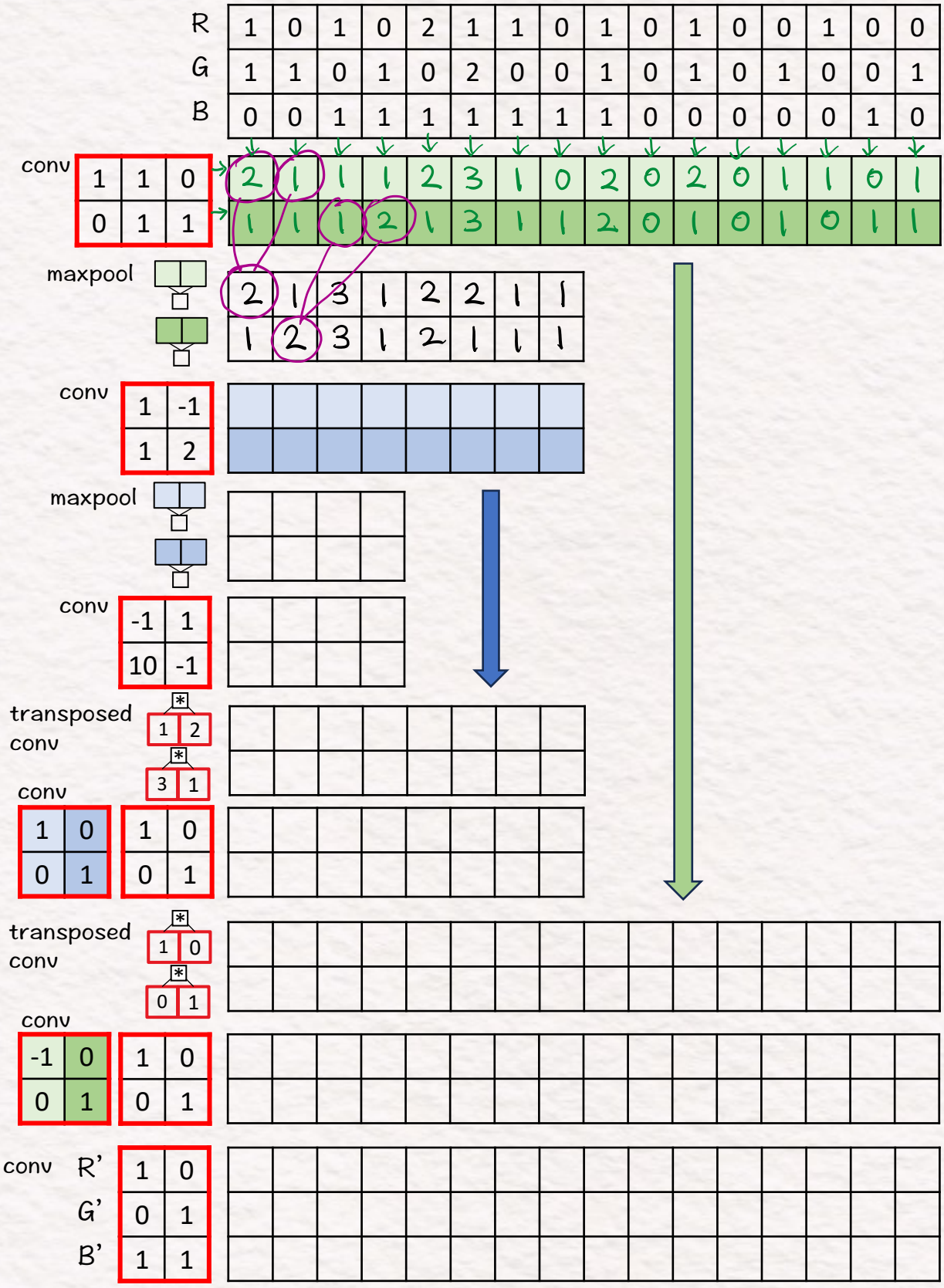
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

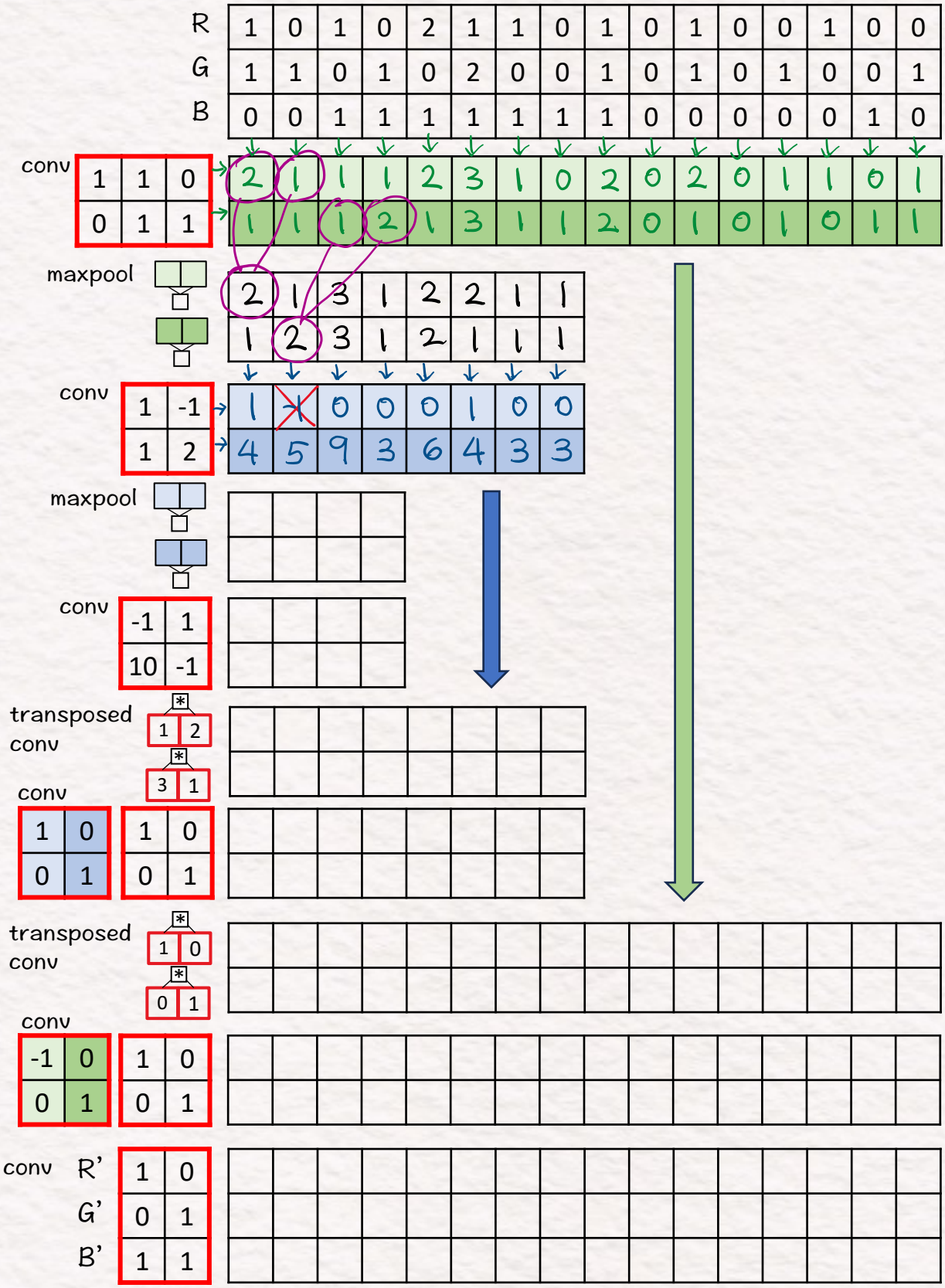
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

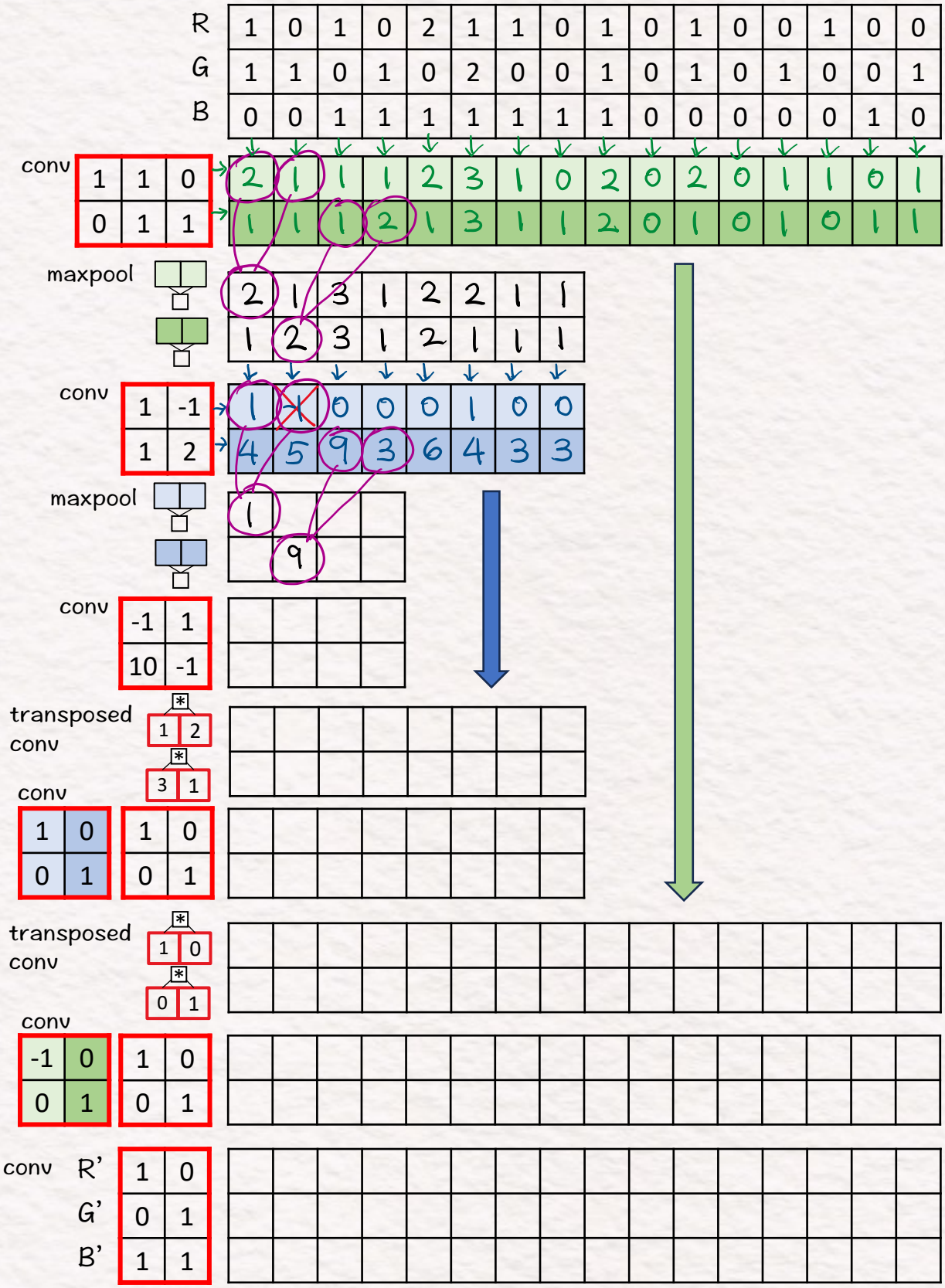
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

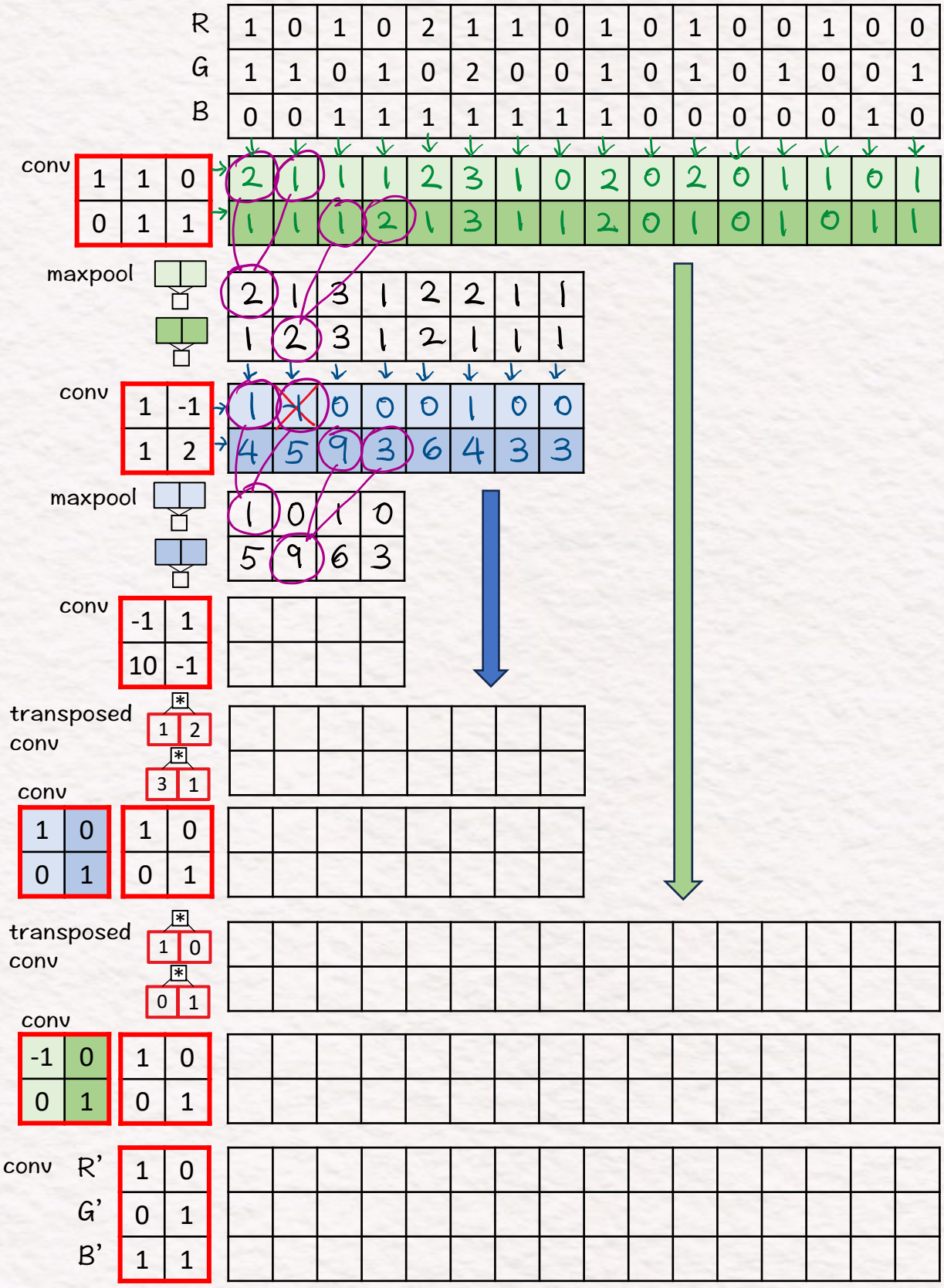
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

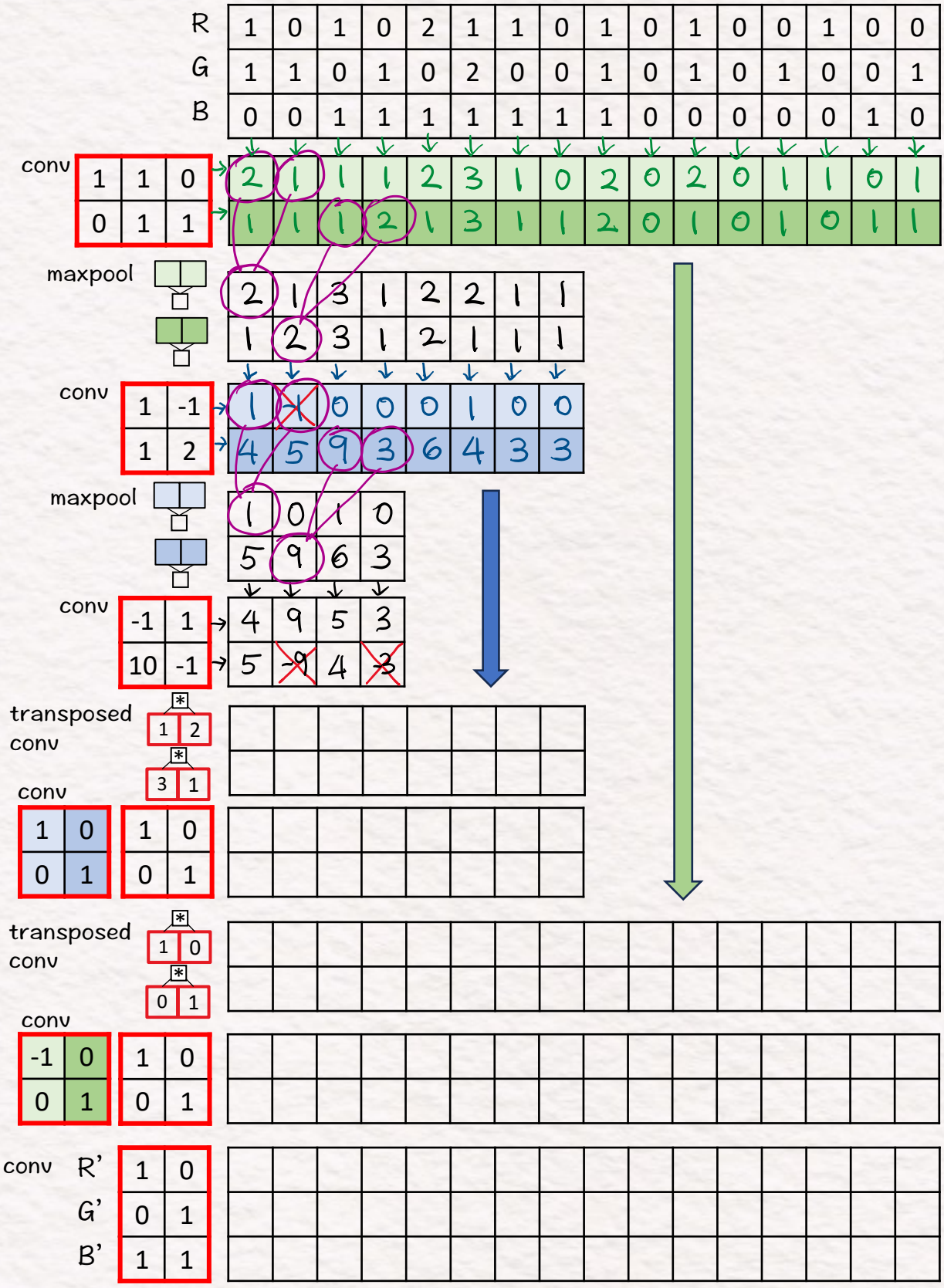
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

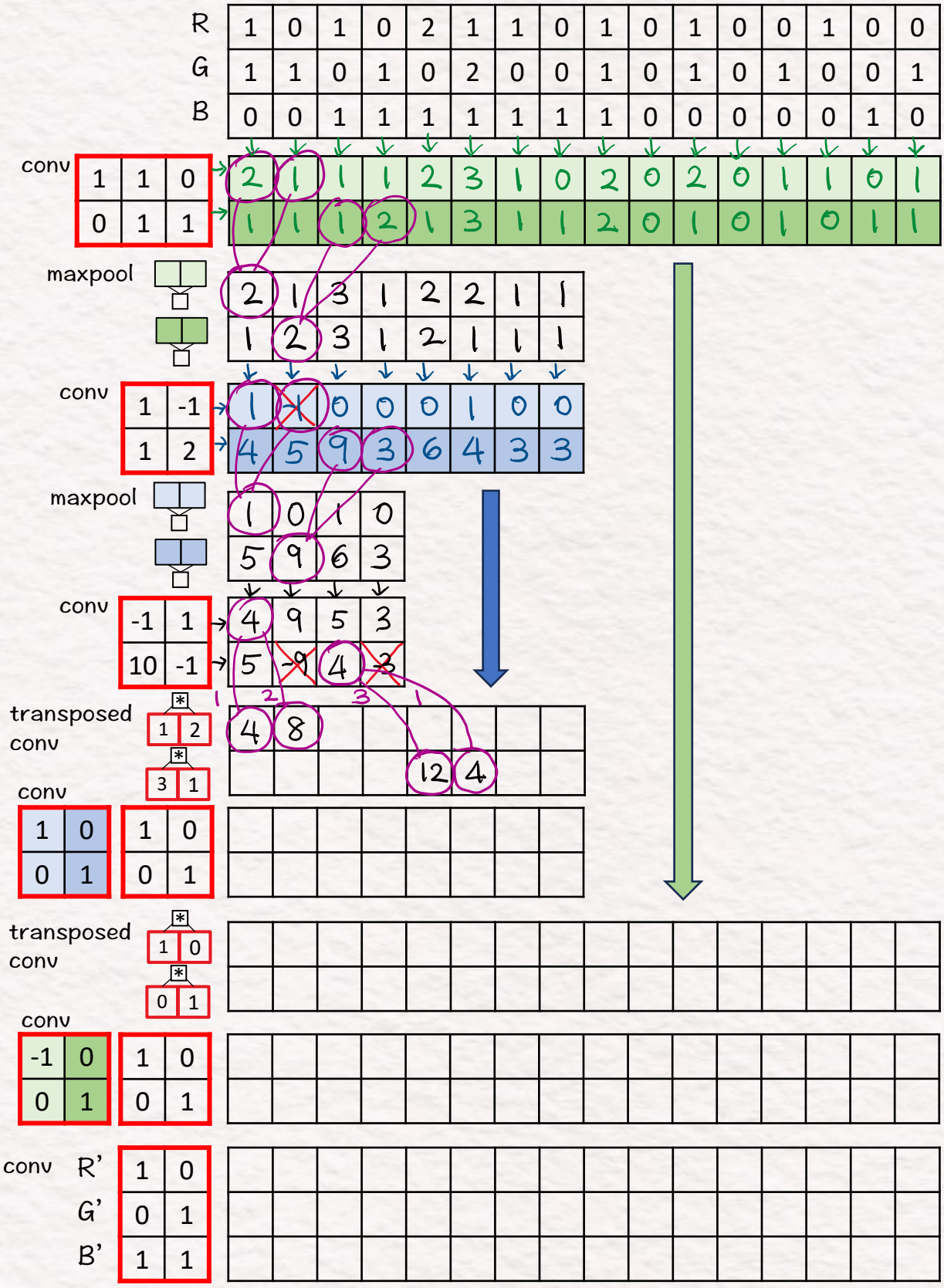
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

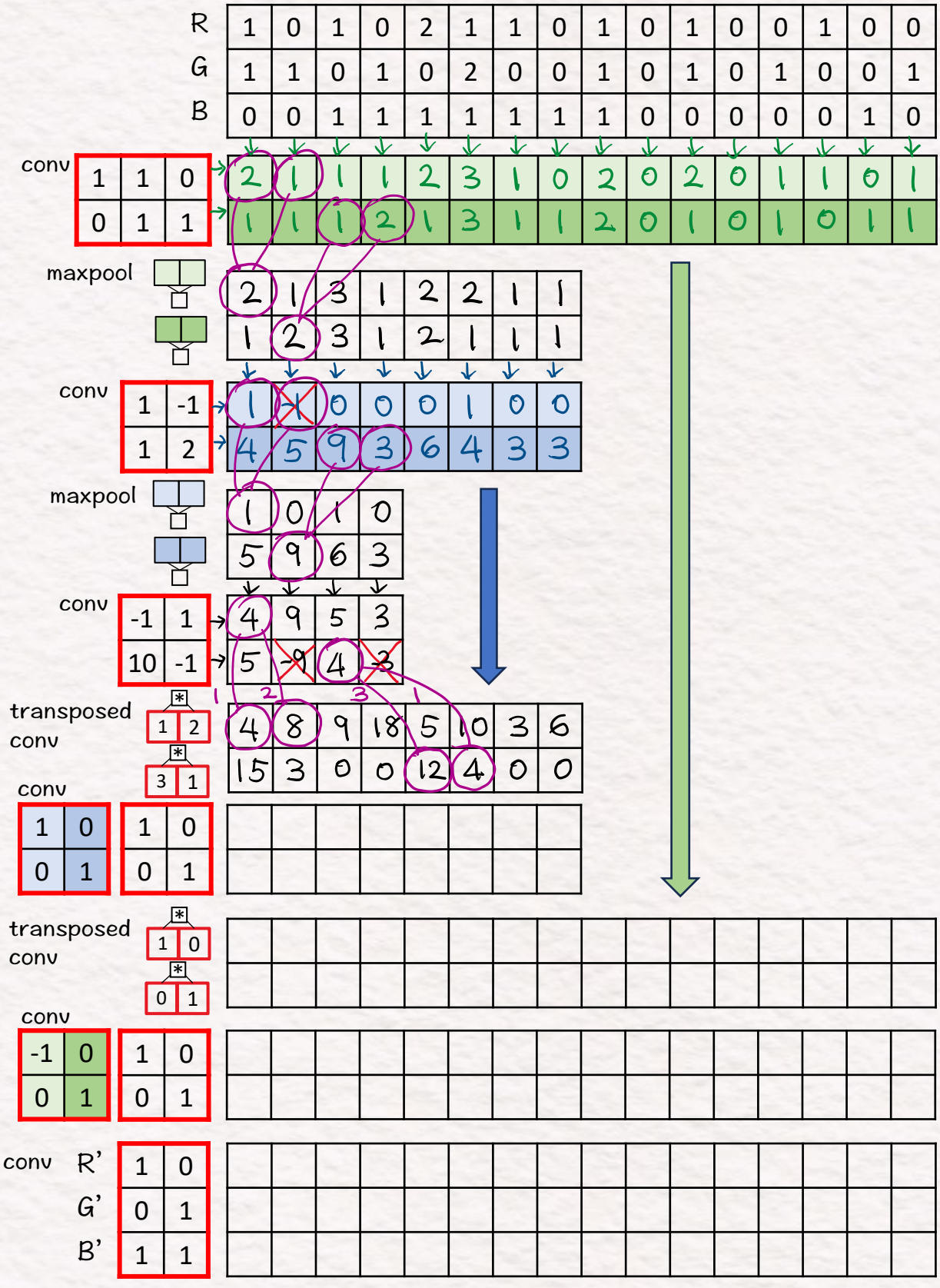
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

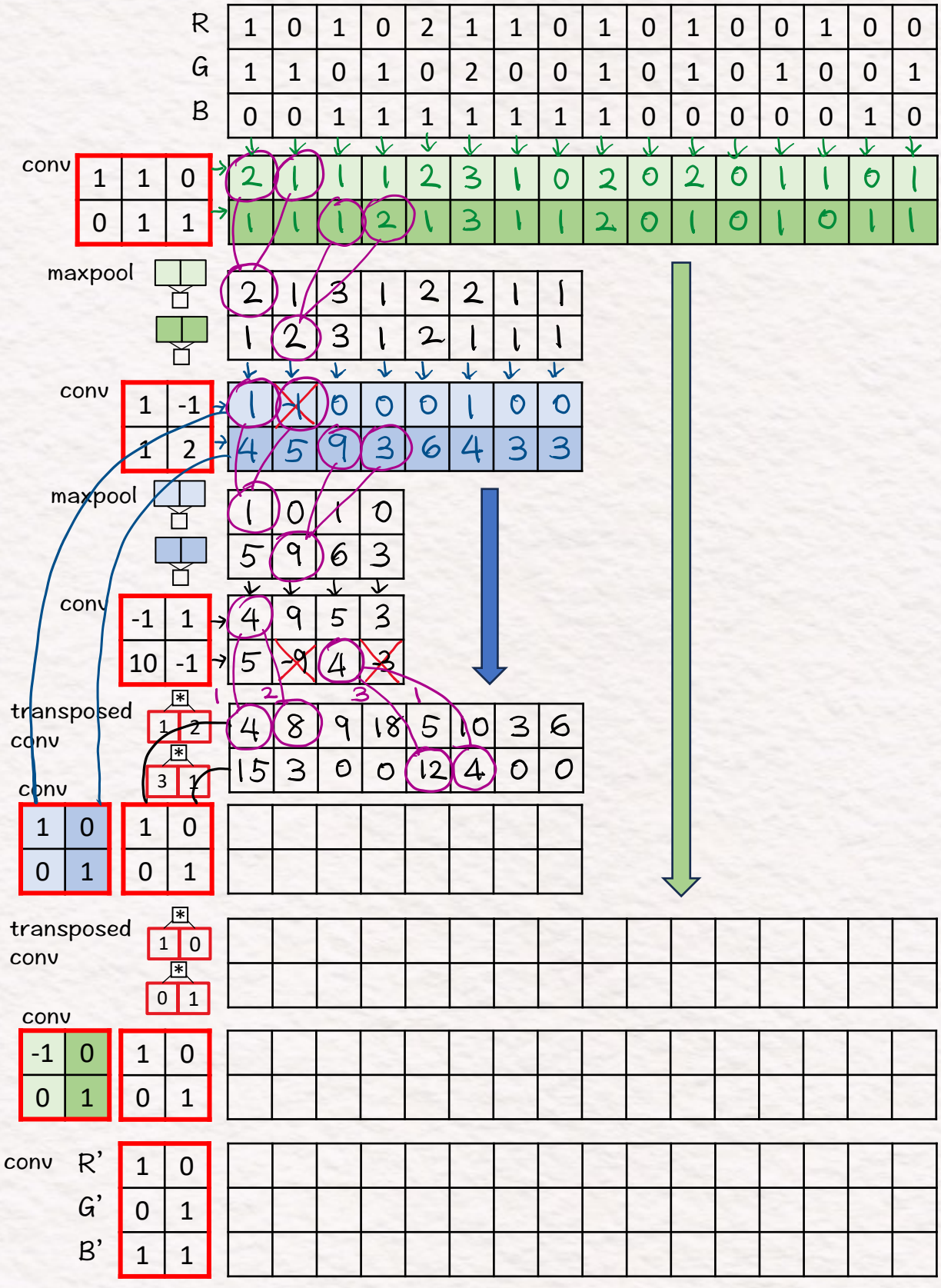
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

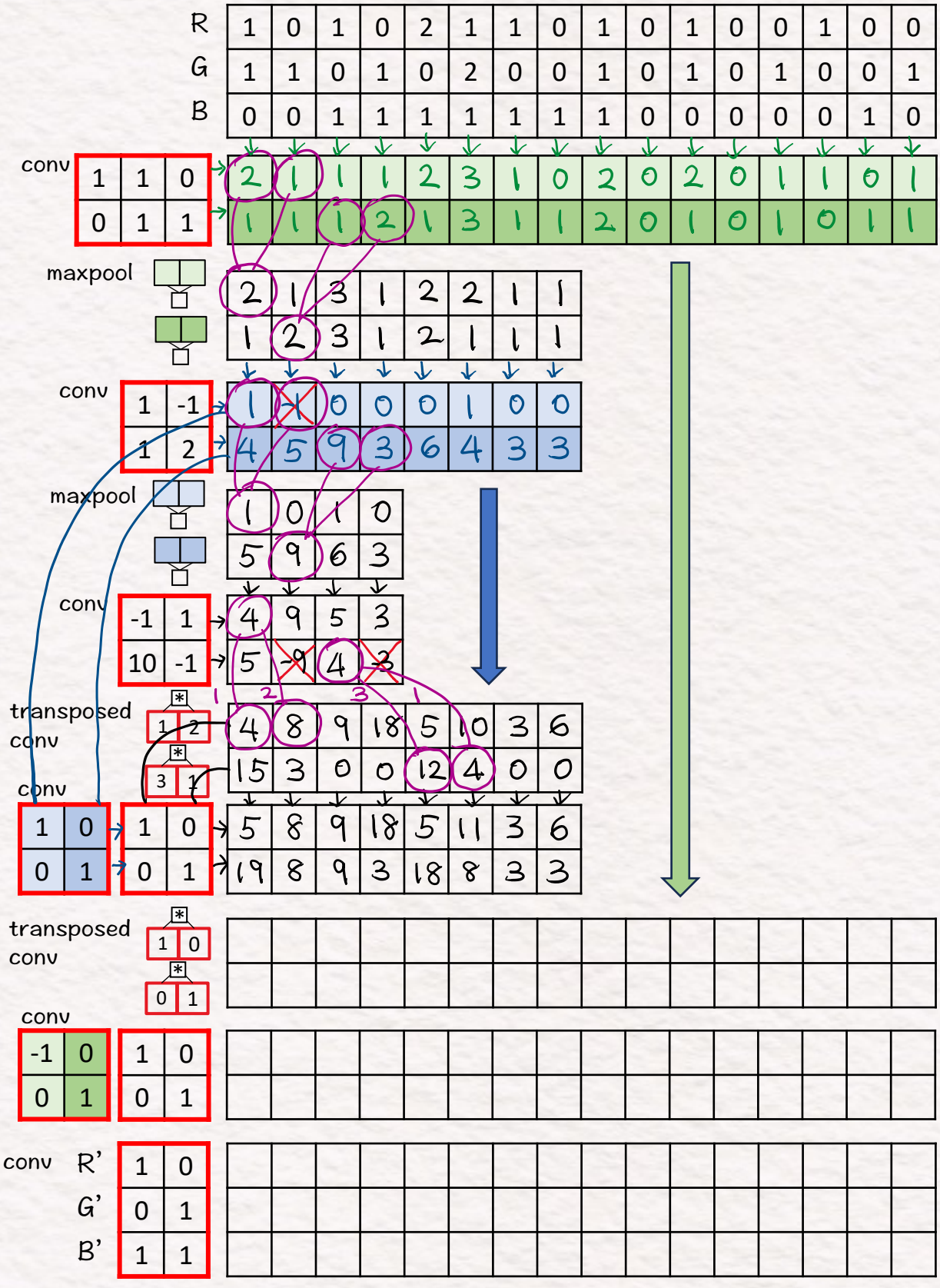
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

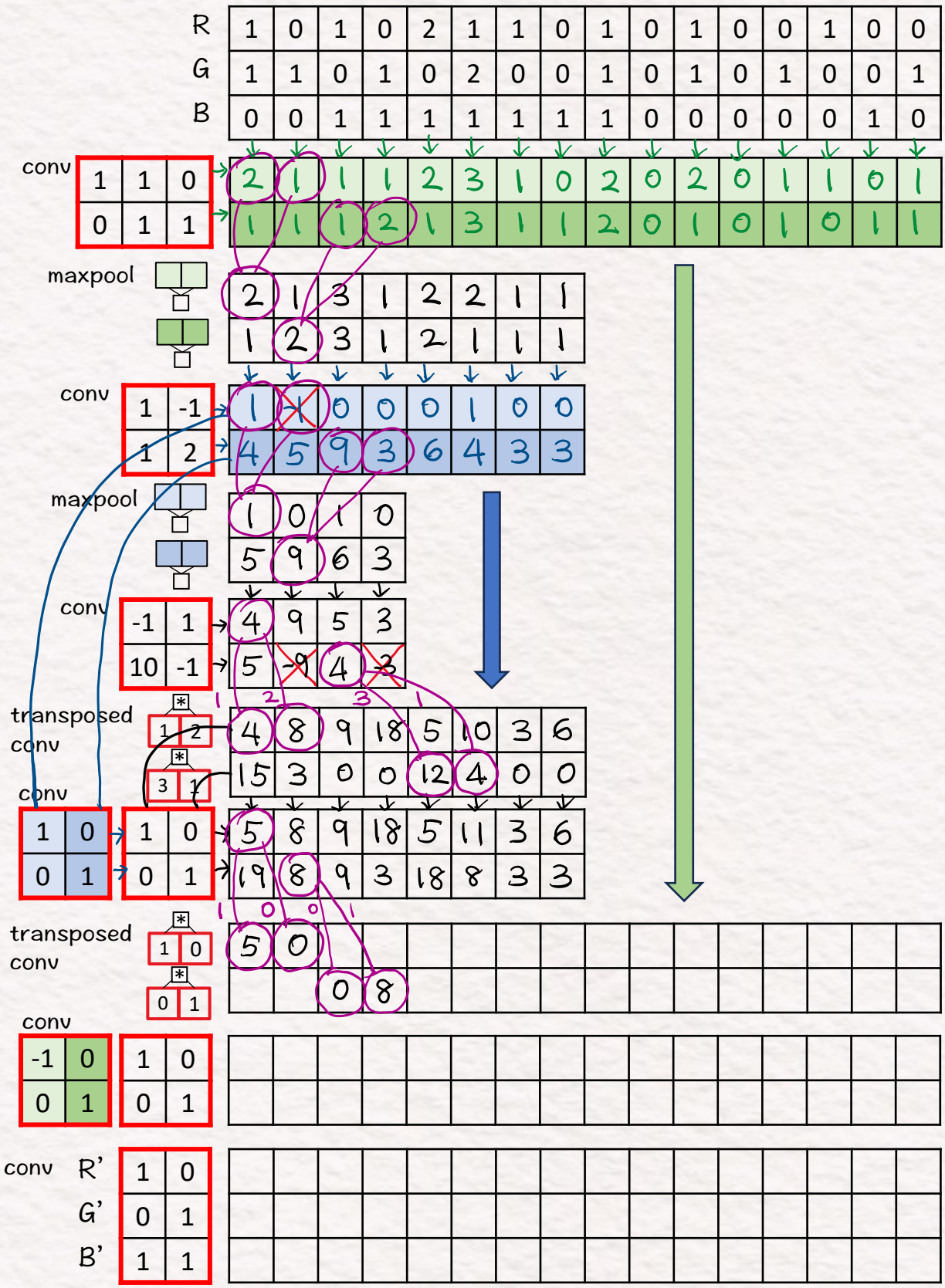
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

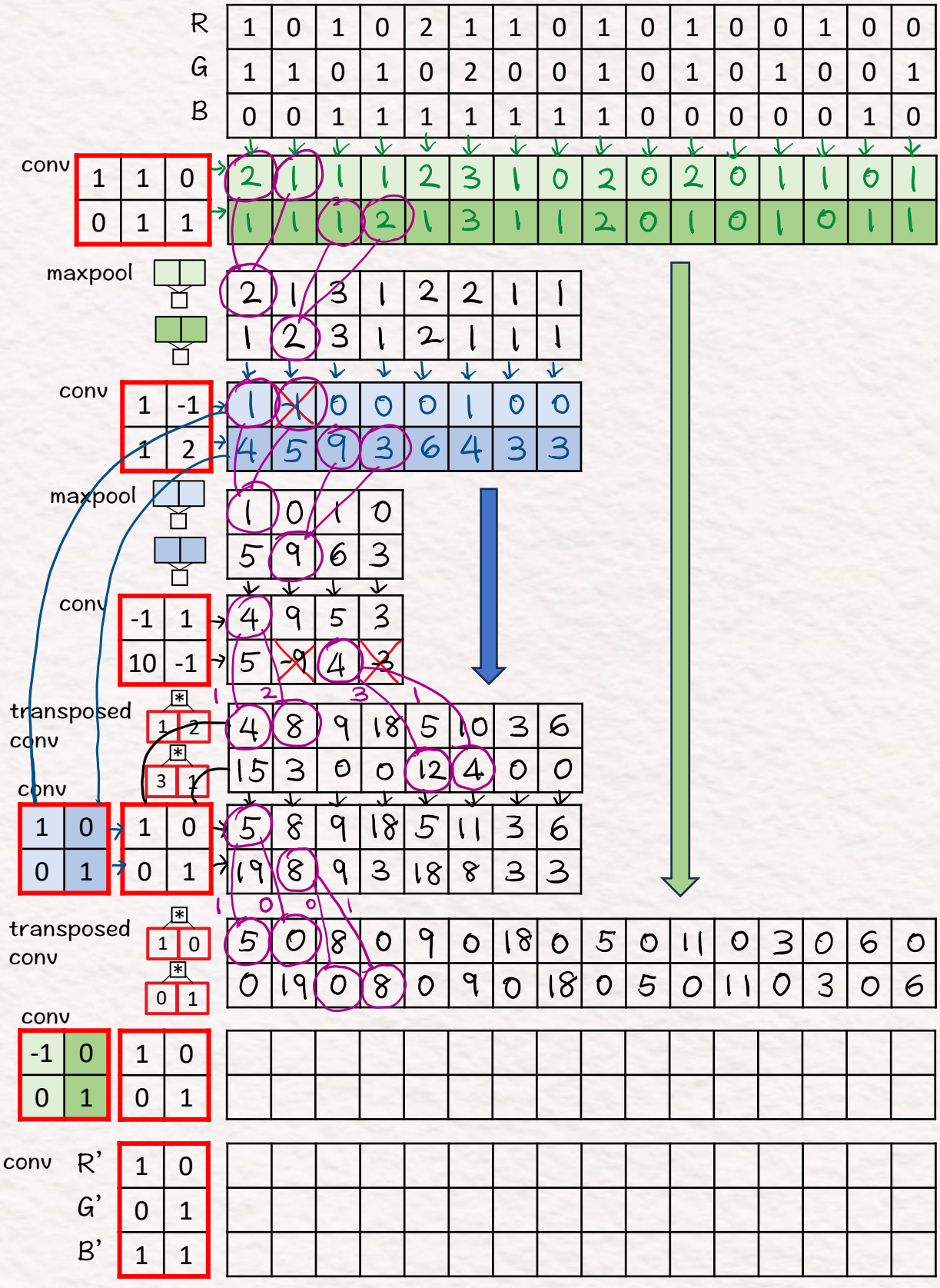
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

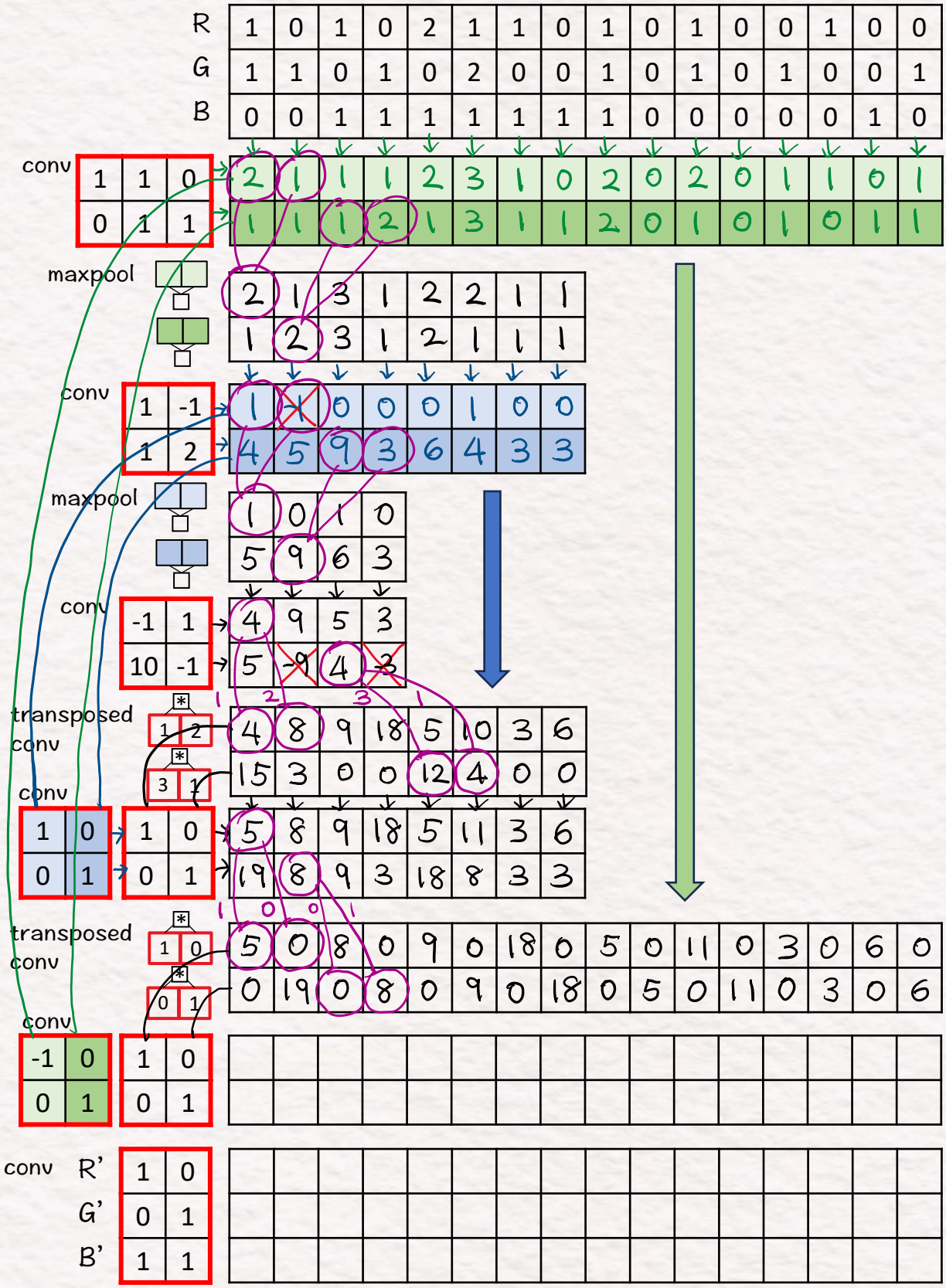
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

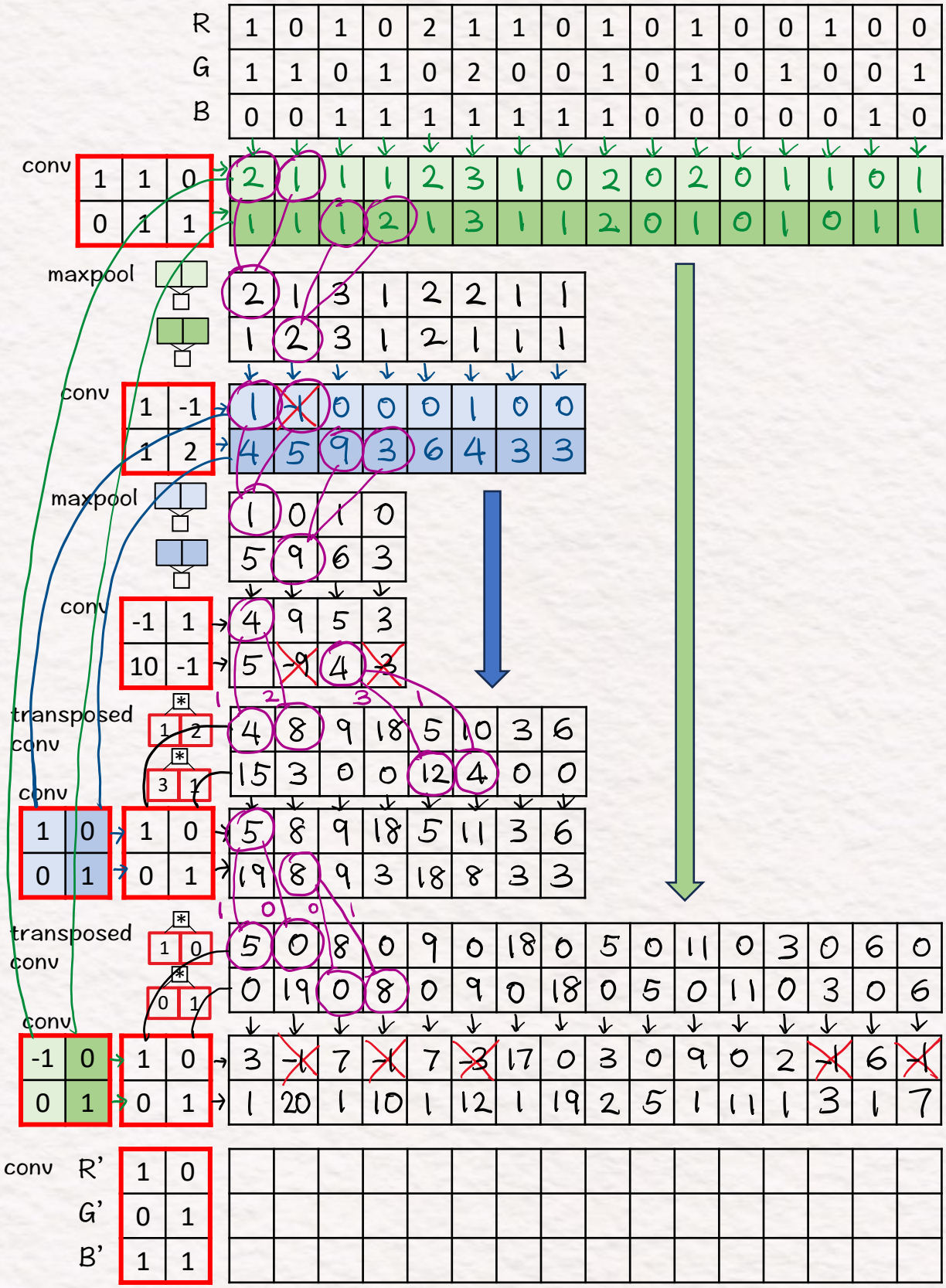
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

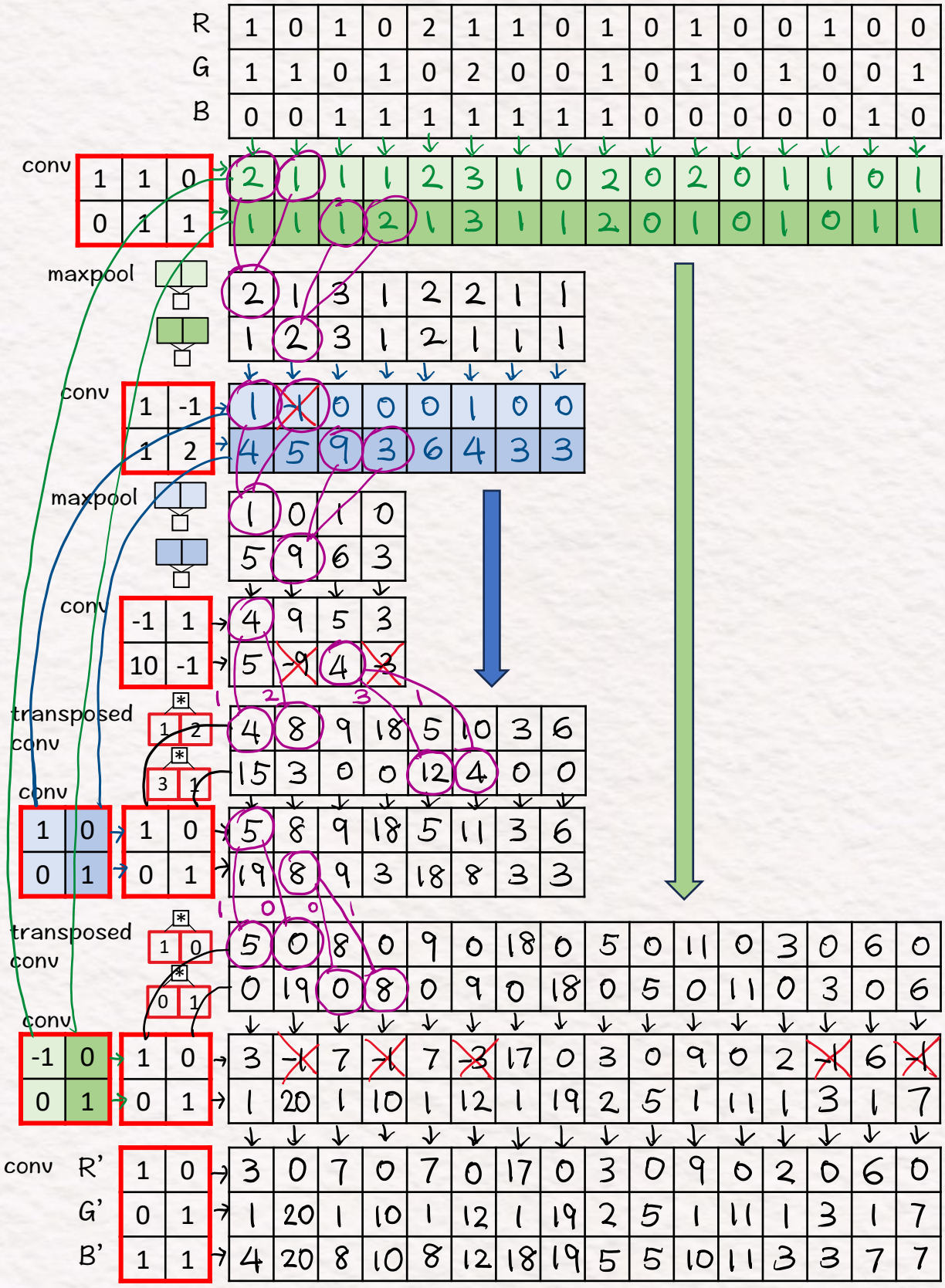
UNet


R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

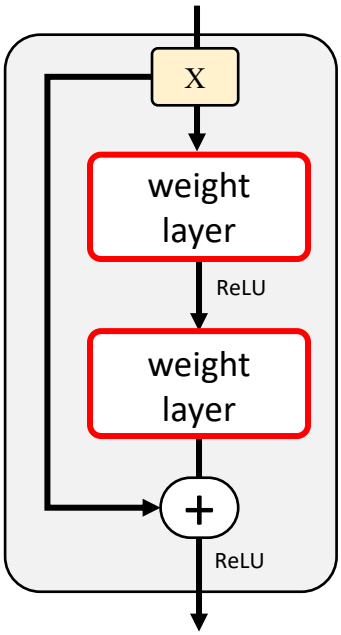
UNet

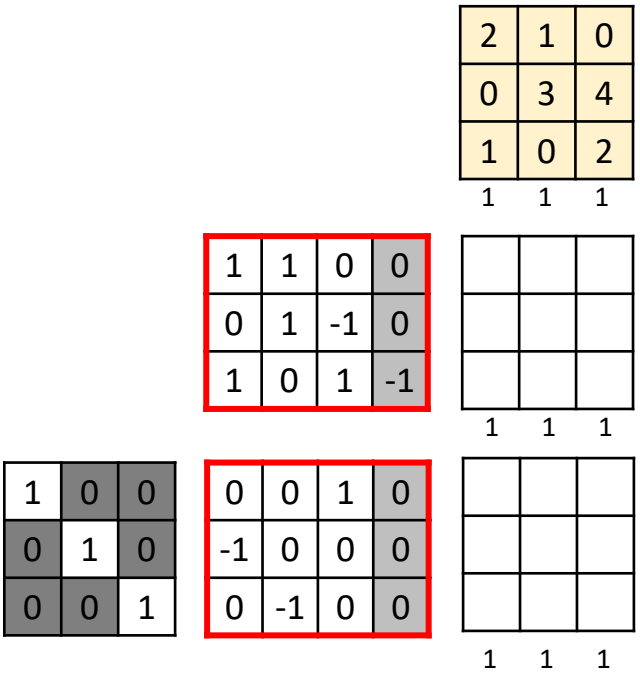

R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv
R'
G'
B'

UNet

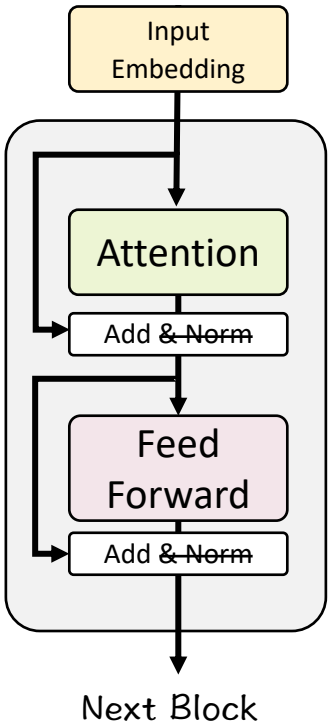

R
G
B
conv
maxpool
conv
maxpool
conv
transposed conv
conv
transposed conv
conv
conv R'
G'
B'

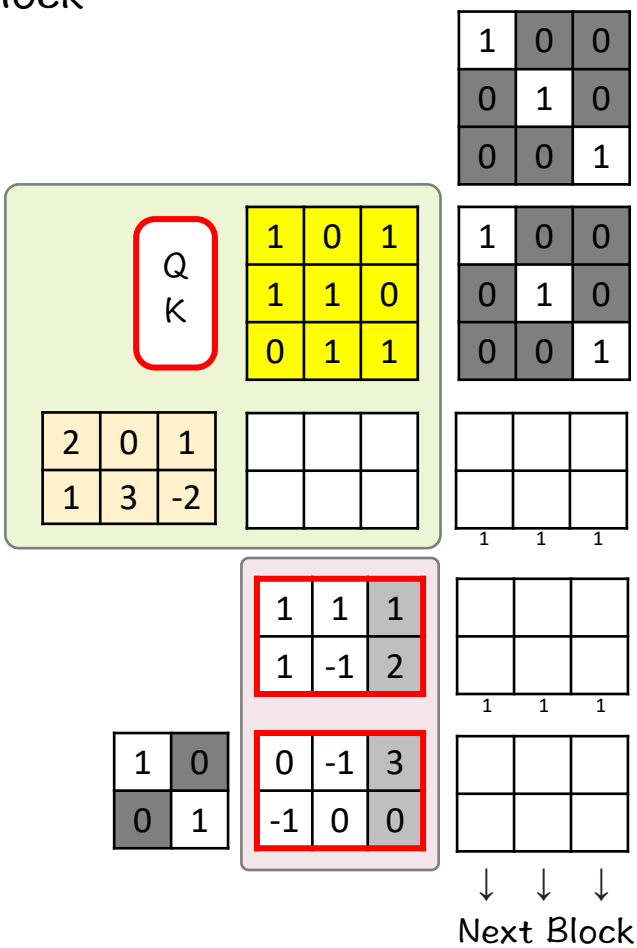
Residual Network



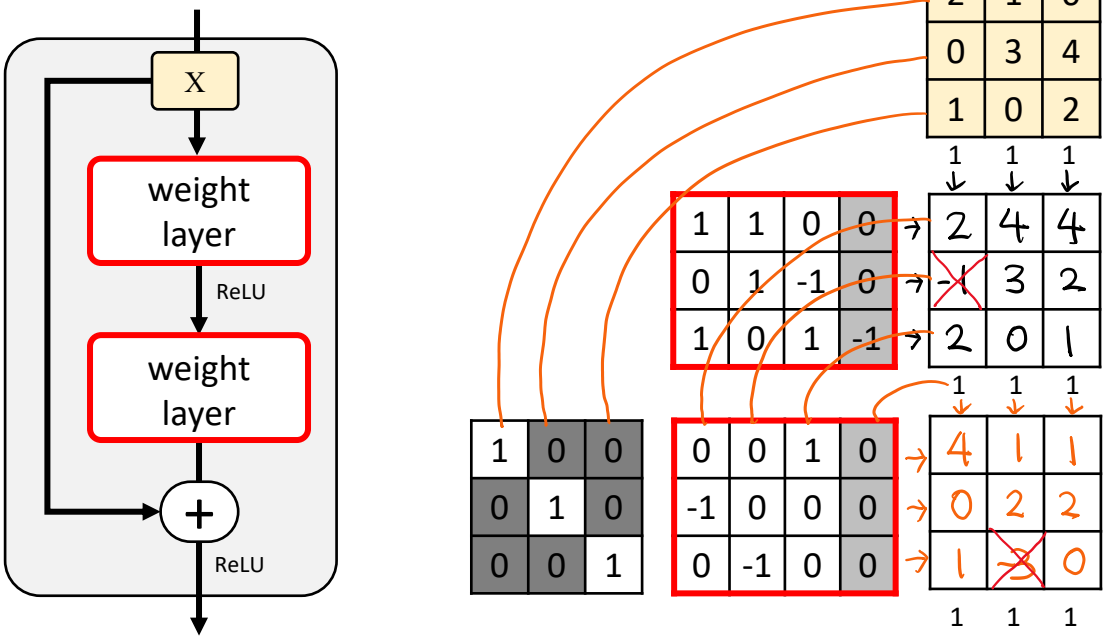


Transformer's Encoder Block

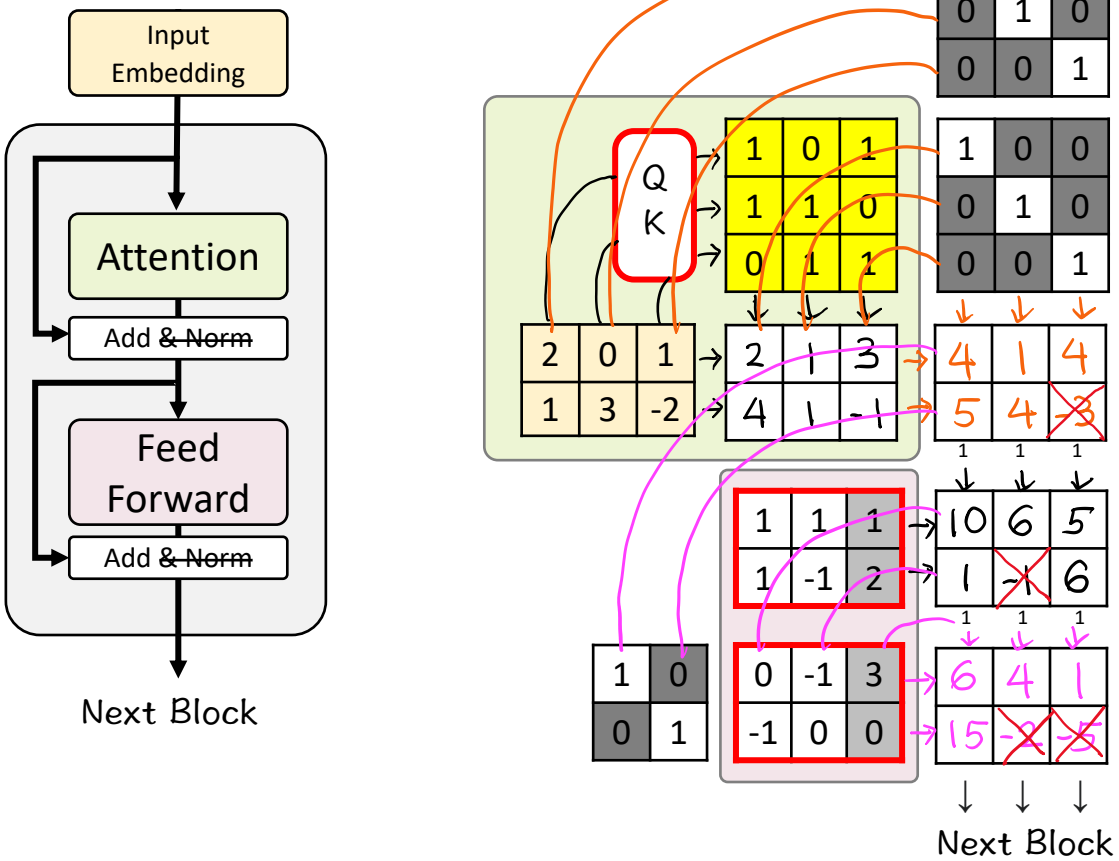




Residual Network



Transformer's Encoder Block



Vector Database

Data

how	are	you

who	are	you

who	am	I

Query

am	I	you

Word Embeddings

a	an	the	how	why	who	what	are	is	am	be	was	you	we	I	they	she	he	she	me	him	her
0	-1	0	1	0	1	0	0	-1	1	0	0	0	3	1	0	-1	0	0	0	-1	0
2	0	2	0	0	0	-1	1	0	0	0	2	1	0	2	0	2	0	0	2	0	0
-1	0	-1	1	2	0	0	1	0	1	-1	0	0	-1	0	3	0	0	-1	0	2	-1
0	1	0	0	1	0	1	0	1	0	1	-2	0	0	0	1	0	1	0	1	0	1

Text Embeddings

Encoder

1	1	0	0	0
0	1	0	1	0
1	0	1	0	-1
1	-1	0	0	0

Linear & ReLU

1 1 1 1 1 1 1 1 1 1 1 1

Mean Pooling

Indexing

Projection

1	1	0	0
0	0	1	1

Vector Storage

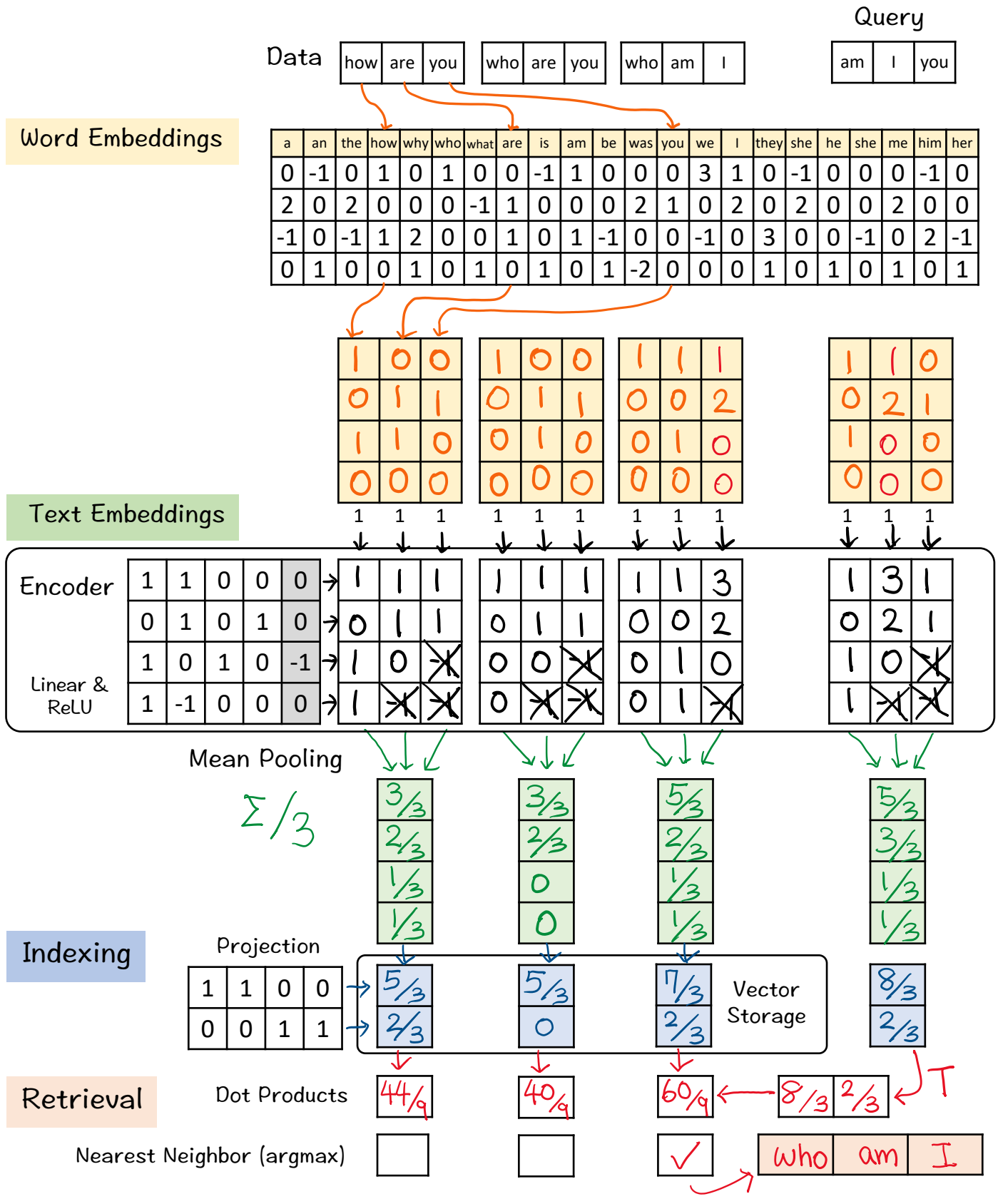
Retrieval

Dot Products

Nearest Neighbor (argmax)

2.1.24

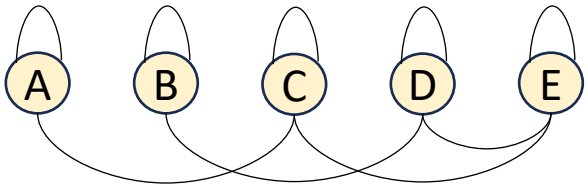
Vector Database



2.1.24

Graph Convolutional Network

Graph Data



Graph Convolutional Network

	A	B	C	D	E
	2	0	1	0	1
	1	1	0	0	0
	0	0	-1	1	1
	0	3	0	1	0
	1	1	1	1	1

	A	B	C	D	E
A					
B					
C					
D					
E					

Adjacency Matrix

	A	B	C	D	E

1	1	0	0	0
0	1	0	-1	0
1	0	0	1	-1

[ReLU]

Messages

1	1	1	1	1

0	1	1	0
1	-1	0	-2
1	0	0	0

[ReLU]

Messages

1	1	1	1	1

Fully Connected Network

1	0	0	-2
0	1	0	-2
0	0	1	-5
1	-1	0	0
1	0	-1	0

[ReLU]

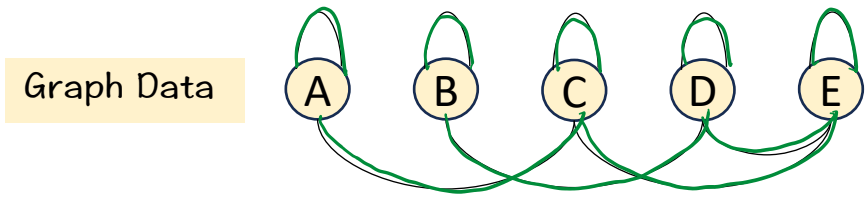
1	1	1	1	1

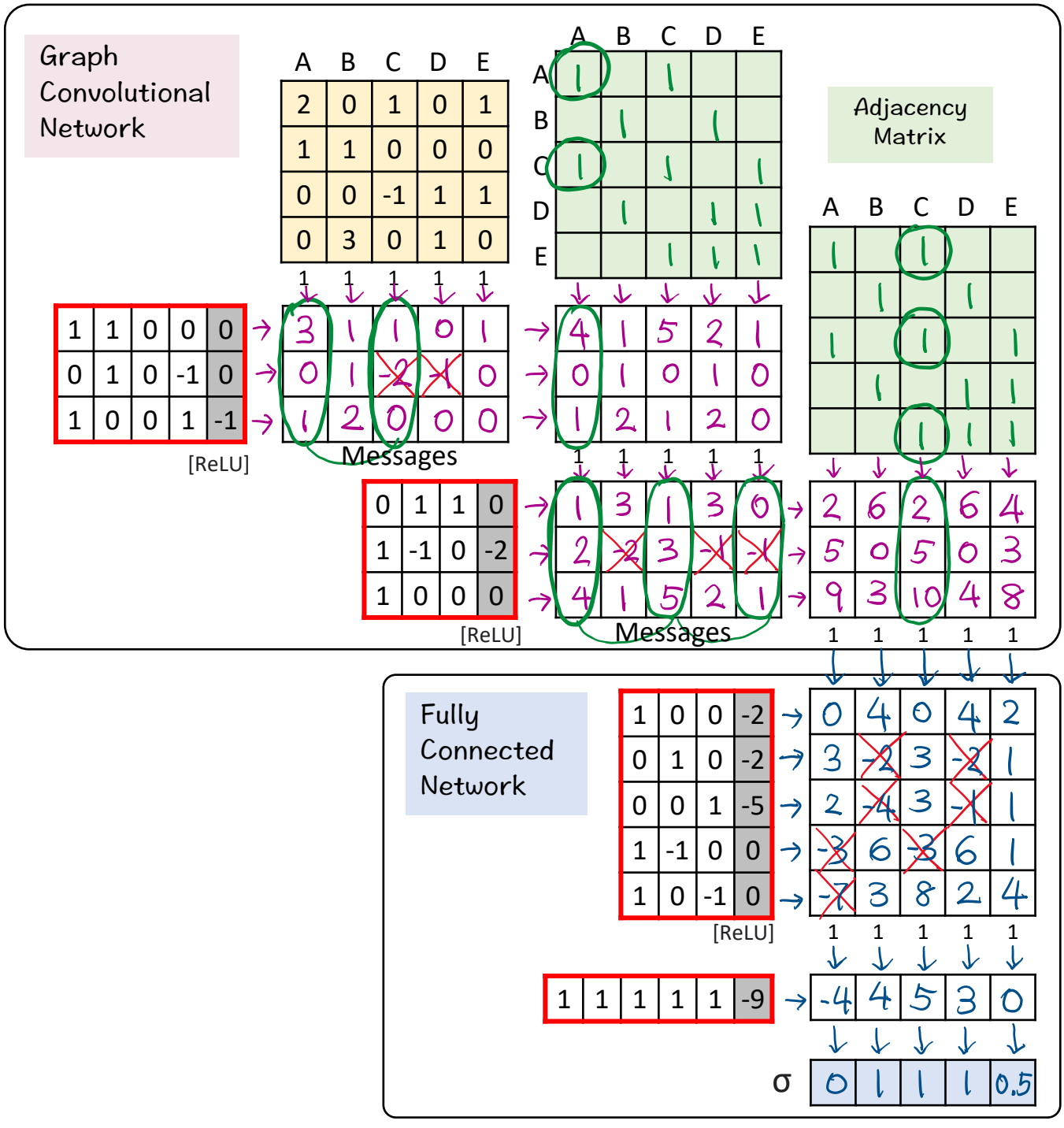
1	1	1	1	1	-9

σ

2.17.24

Graph Convolutional Network





Recurrent Neural Network (RNN)

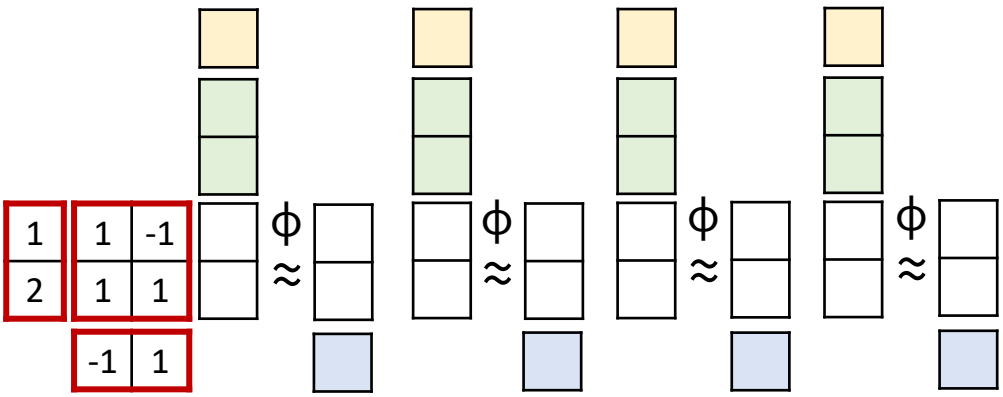
Input Sequence X | 3 | 4 | 5 | 6 |

Parameters A | 1 | -1 | / | 1 | 1 | B | 1 | / | 2 | C | -1 | 1 |

Activation Function ϕ: ReLU

Hidden States H_0 | 0 | / | 0 |

Output Sequence Y | | | | |



Recurrent Neural Network (RNN)

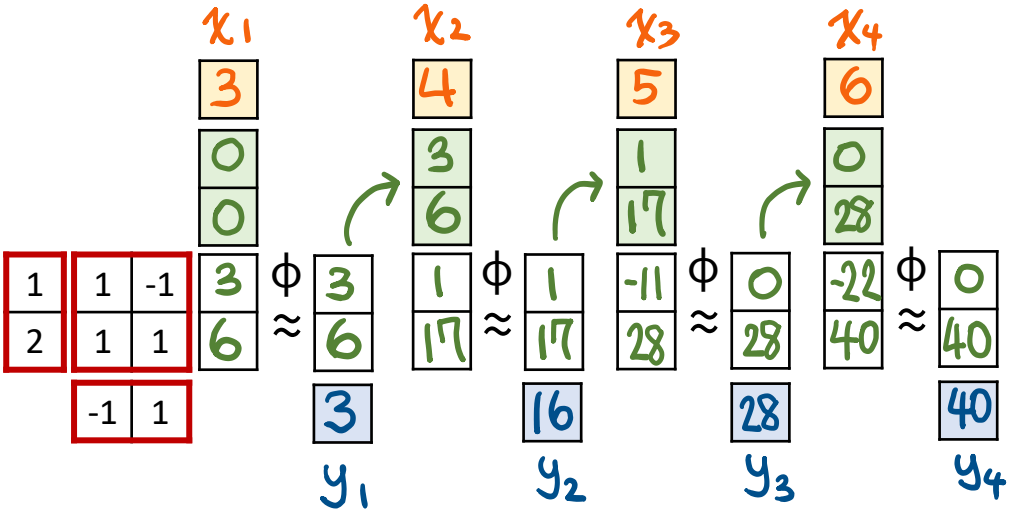
Input Sequence X | 3 | 4 | 5 | 6 |

Parameters A | 1 | -1 | / | 1 | 1 | B | 1 | / | 2 | C | -1 | 1 |

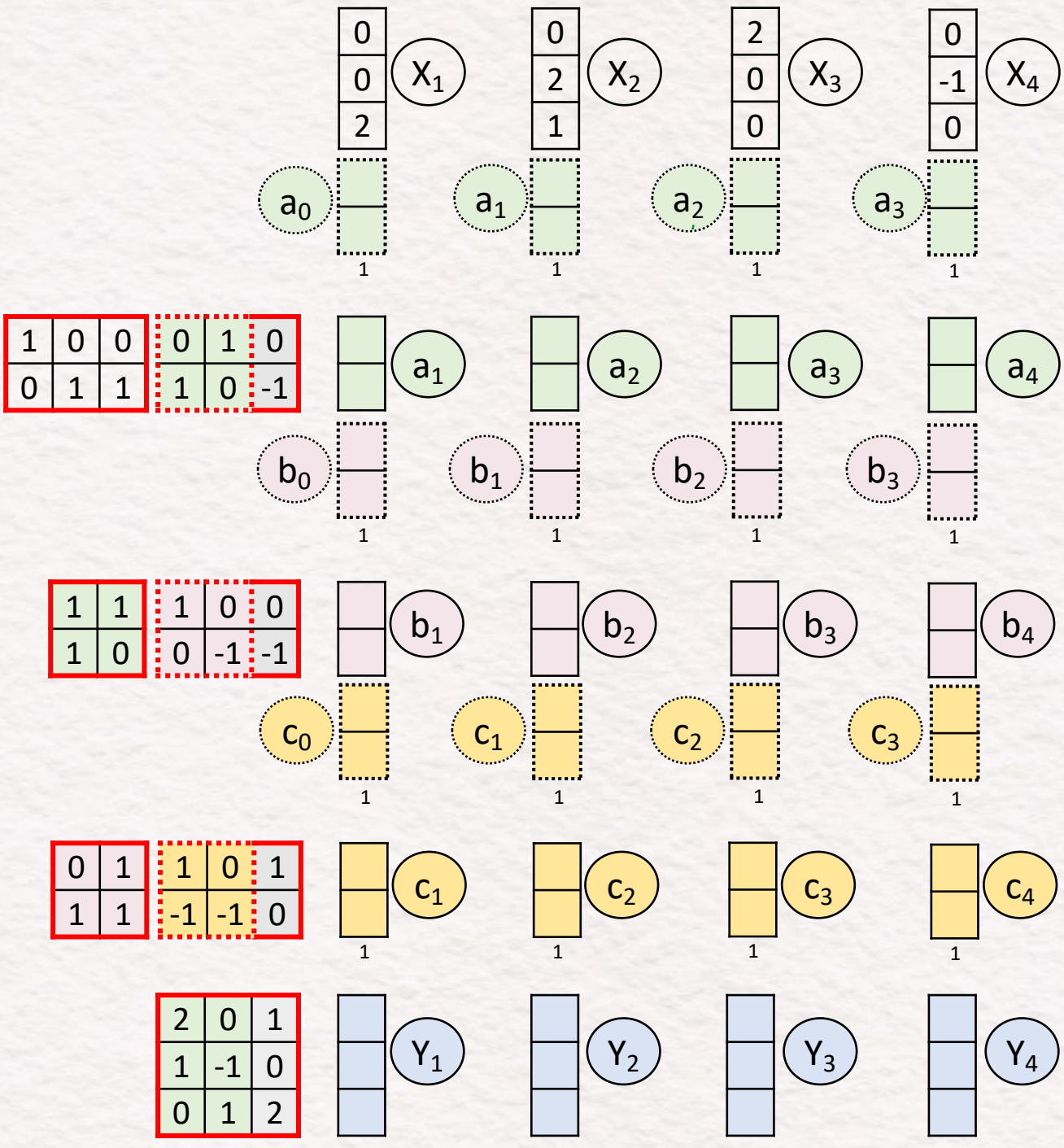
Activation Function ϕ: ReLU

Hidden States H_0 | 0 | / | 0 |

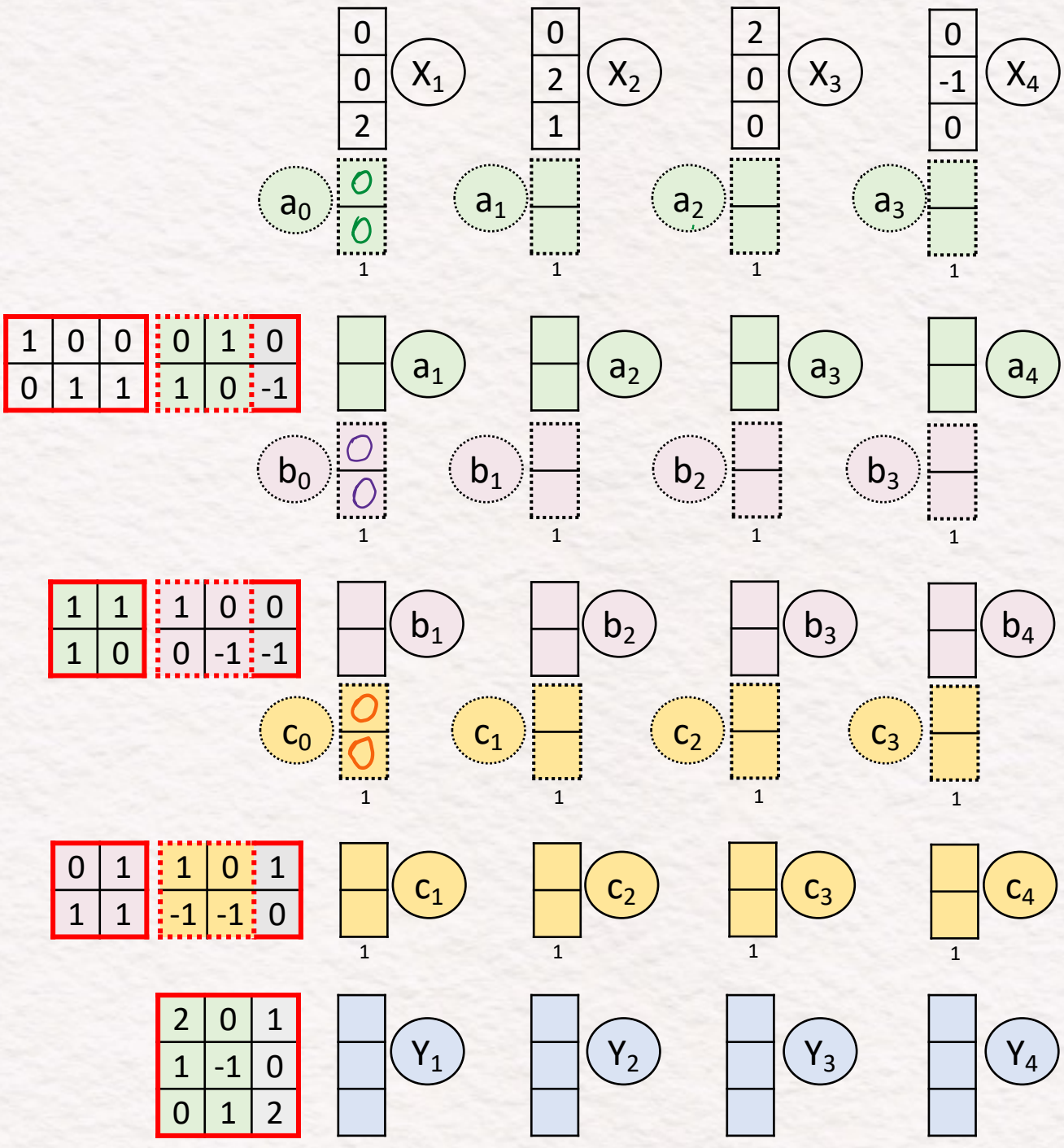
Output Sequence Y | | | | |



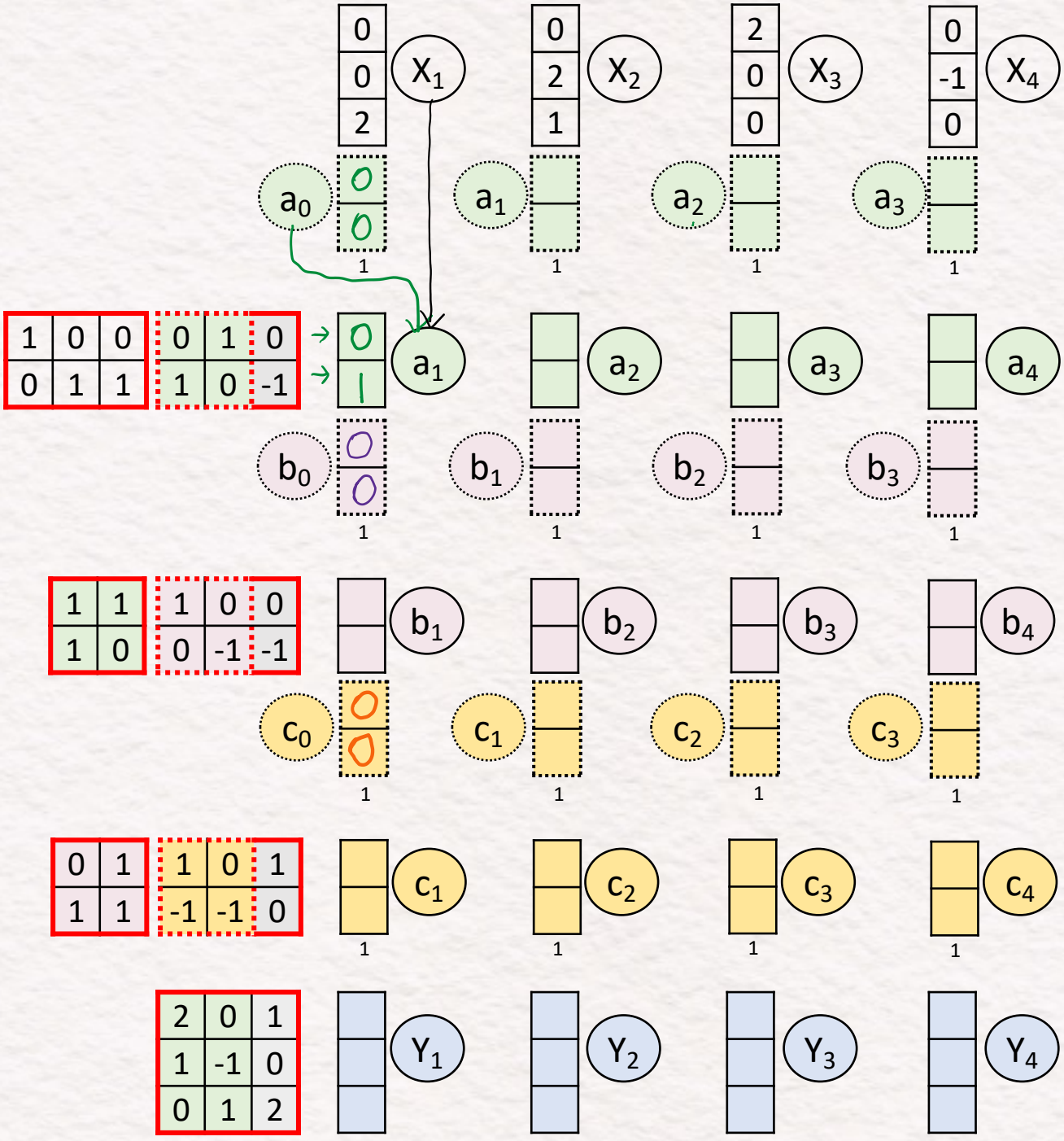
Deep RNN


0
0
2
X1
0
2
1
X2
2
0
0
X3
0
-1
0
X4
a0
a1
a2
a3
1
1
1
1
1 0 0
0 1 1
0 1 0
1 0 -1
a1
a2
a3
a4
b0
b1
b2
b3
1
1
1
1
1 1
1 0
1 0 0
0 -1 -1
b1
b2
b3
b4
c0
c1
c2
c3
1
1
1
1
0 1
1 1
1 0 1
-1 -1 0
c1
c2
c3
c4
1
1
1
1
2 0 1
1 -1 0
0 1 2
Y1
Y2
Y3
Y4

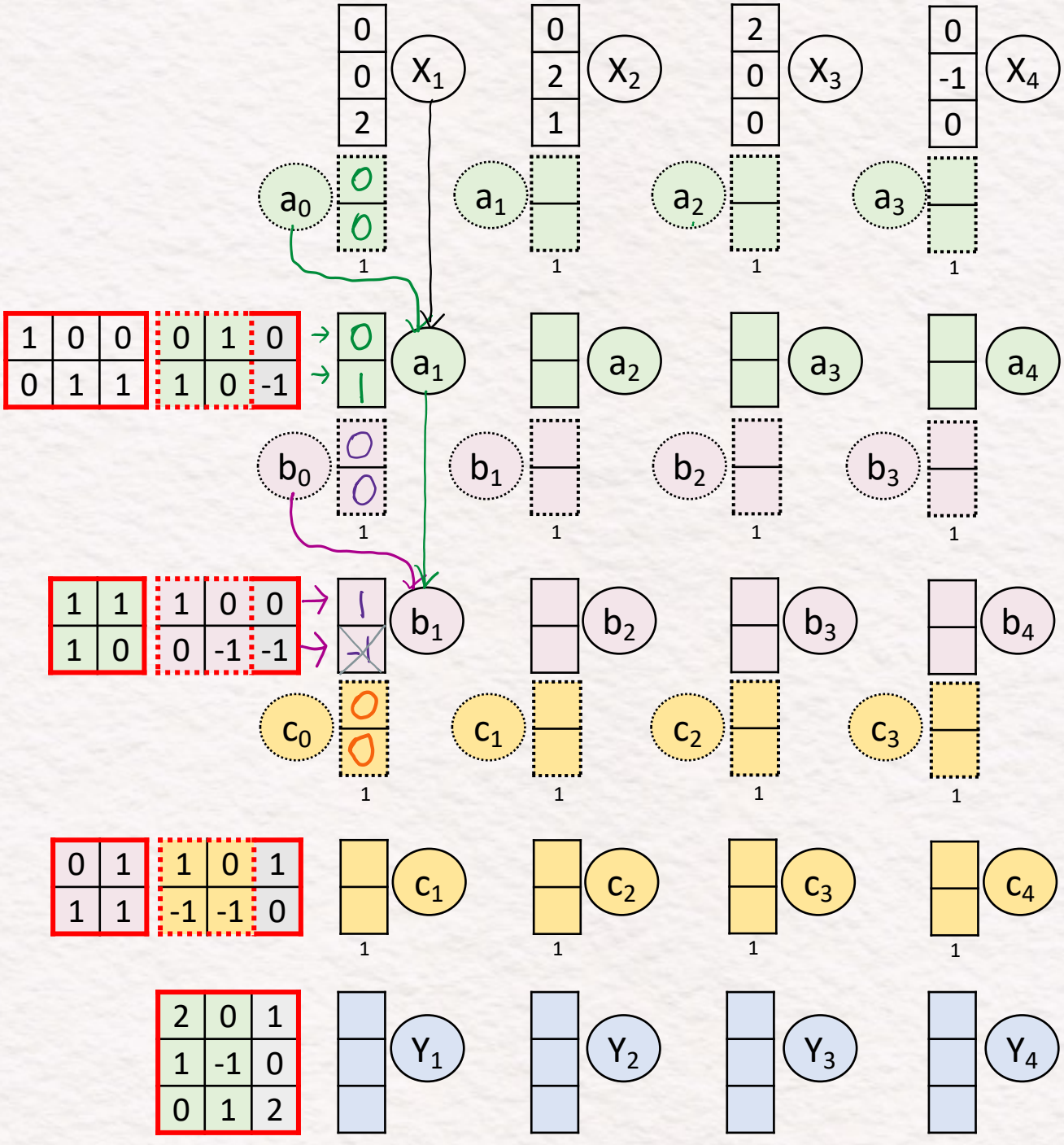
Deep RNN


0 0 2
X1
0 2 1
X2
2 0 0
X3
0 -1 0
X4
a0
a1
a2
a3
1
1 0 0 0 1 0
0 1 1 1 0 -1
a1
a2
a3
a4
b0
b1
b2
b3
1 1 1 0 0
1 0 0 -1 -1
b1
b2
b3
b4
c0
c1
c2
c3
0 1 1 0 1
1 1 -1 -1 0
c1
c2
c3
c4
2 0 1
1 -1 0
0 1 2
Y1
Y2
Y3
Y4

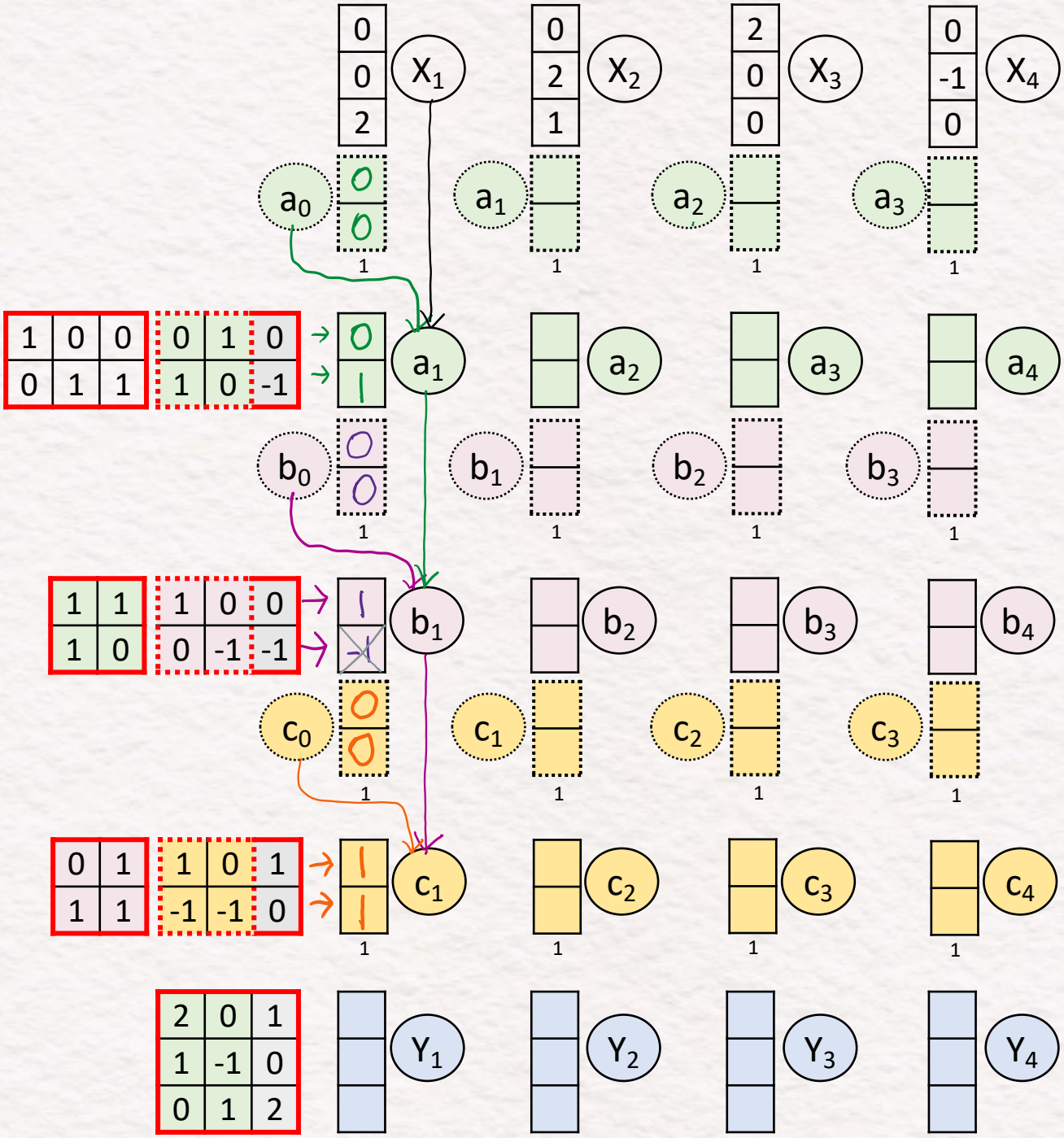
Deep RNN


0 0 2
X1
0 2 1
X2
2 0 0
X3
0 -1 0
X4
a0
a1
a2
a3
1 0 0 0 1 0
0 1 1 1 0 -1
a1
a2
a3
a4
b0
b1
b2
b3
1 1 1 0 0
1 0 0 -1 -1
b1
b2
b3
b4
c0
c1
c2
c3
0 1 1 1 1 0 1
-1 -1 0
c1
c2
c3
c4
2 0 1 1 -1 0 0 1 2
Y1
Y2
Y3
Y4

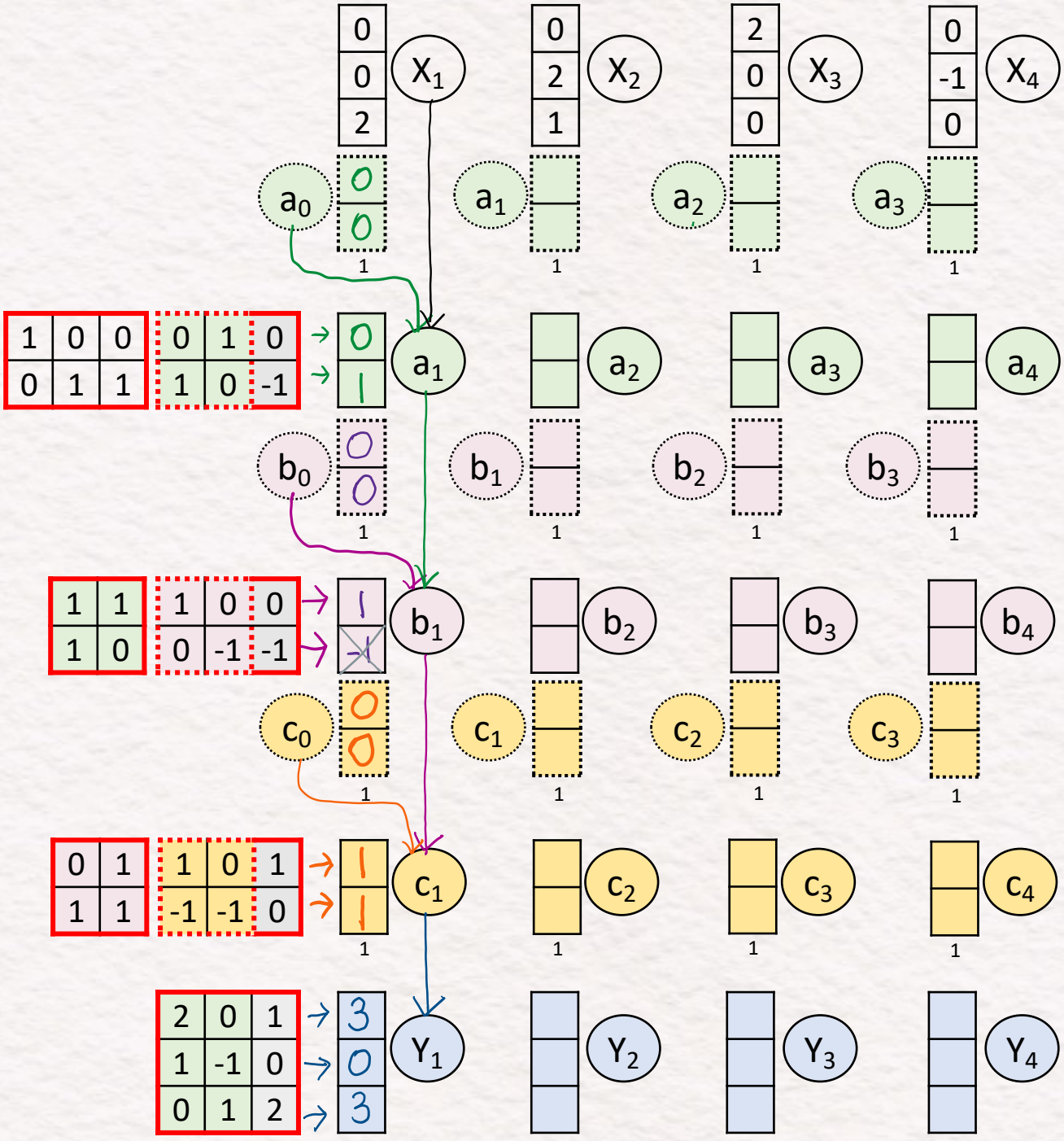
Deep RNN



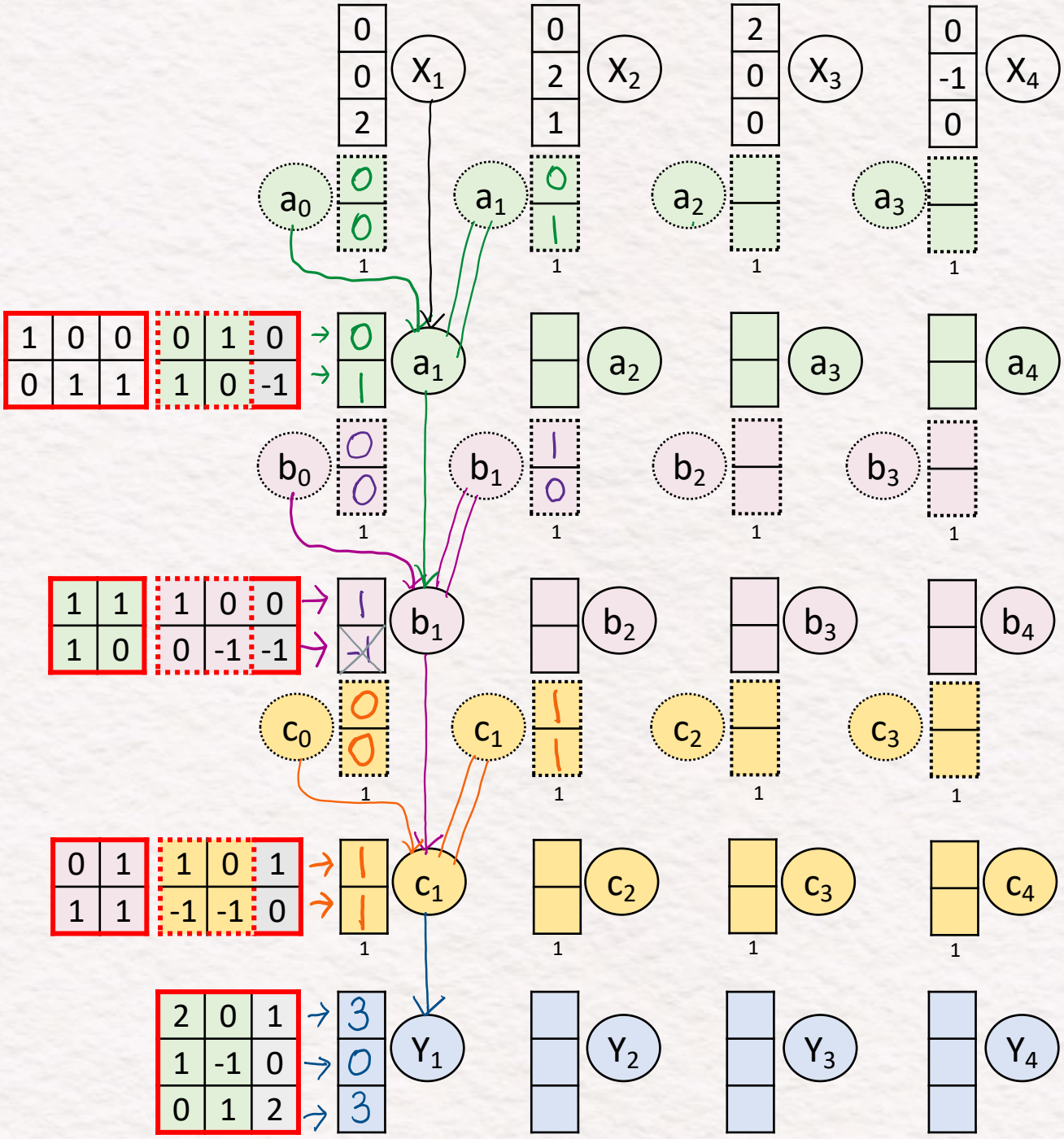
Deep RNN


0
0
2
X1
0
2
1
X2
2
0
0
X3
0
-1
0
X4
a0
a1
a2
a3
1
1
1
1
1 0 0
0 1 1
0 1 0
1 0 -1
a1
a2
a3
a4
b0
b1
b2
b3
1
1
1
1
1 1
1 0
1 0 0
0 -1 -1
b1
b2
b3
b4
c0
c1
c2
c3
1
1
1
1
0 1
1 1
1 0 1
-1 -1 0
c1
c2
c3
c4
1
1
1
1
2 0 1
1 -1 0
0 1 2
Y1
Y2
Y3
Y4

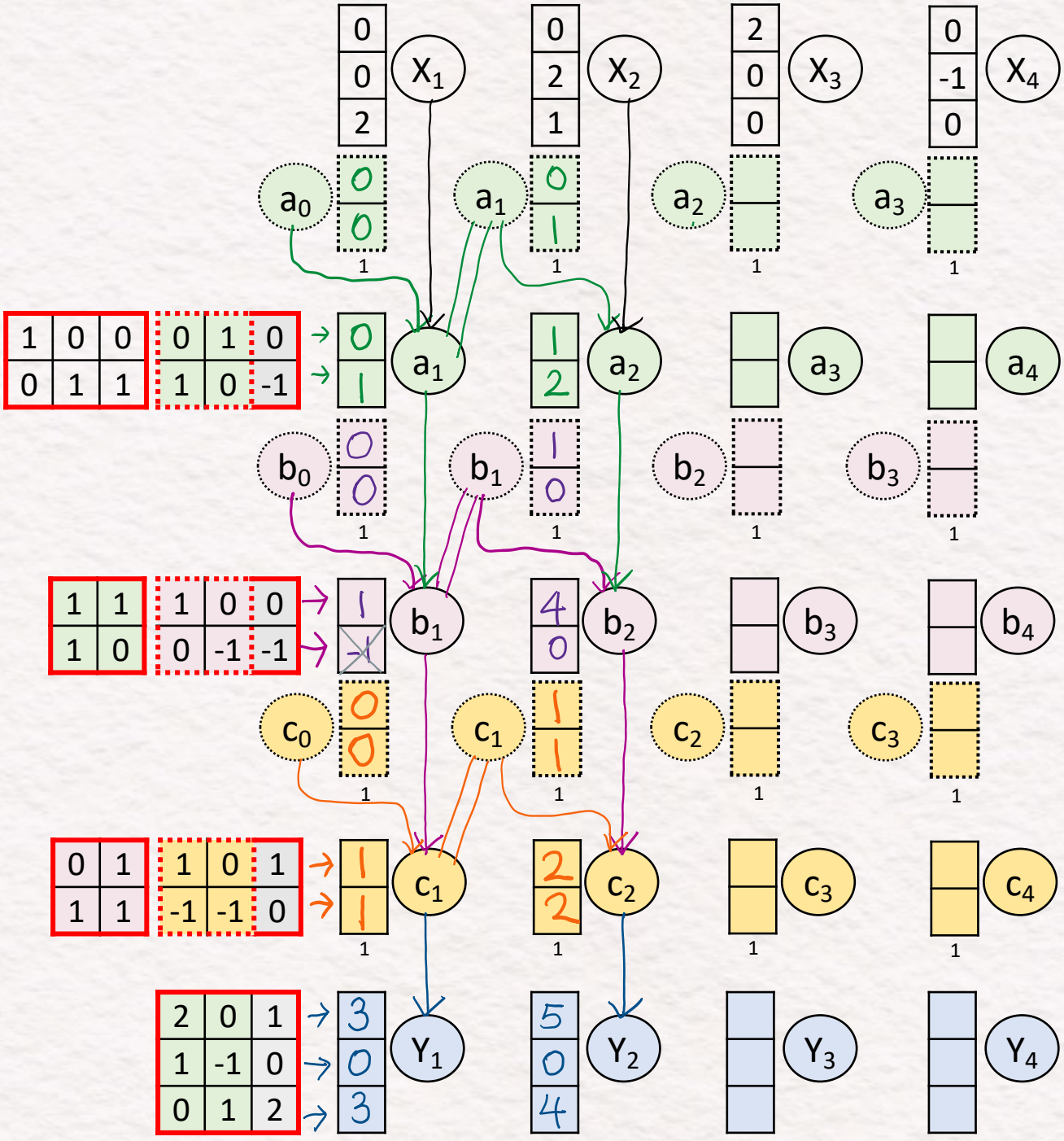
Deep RNN



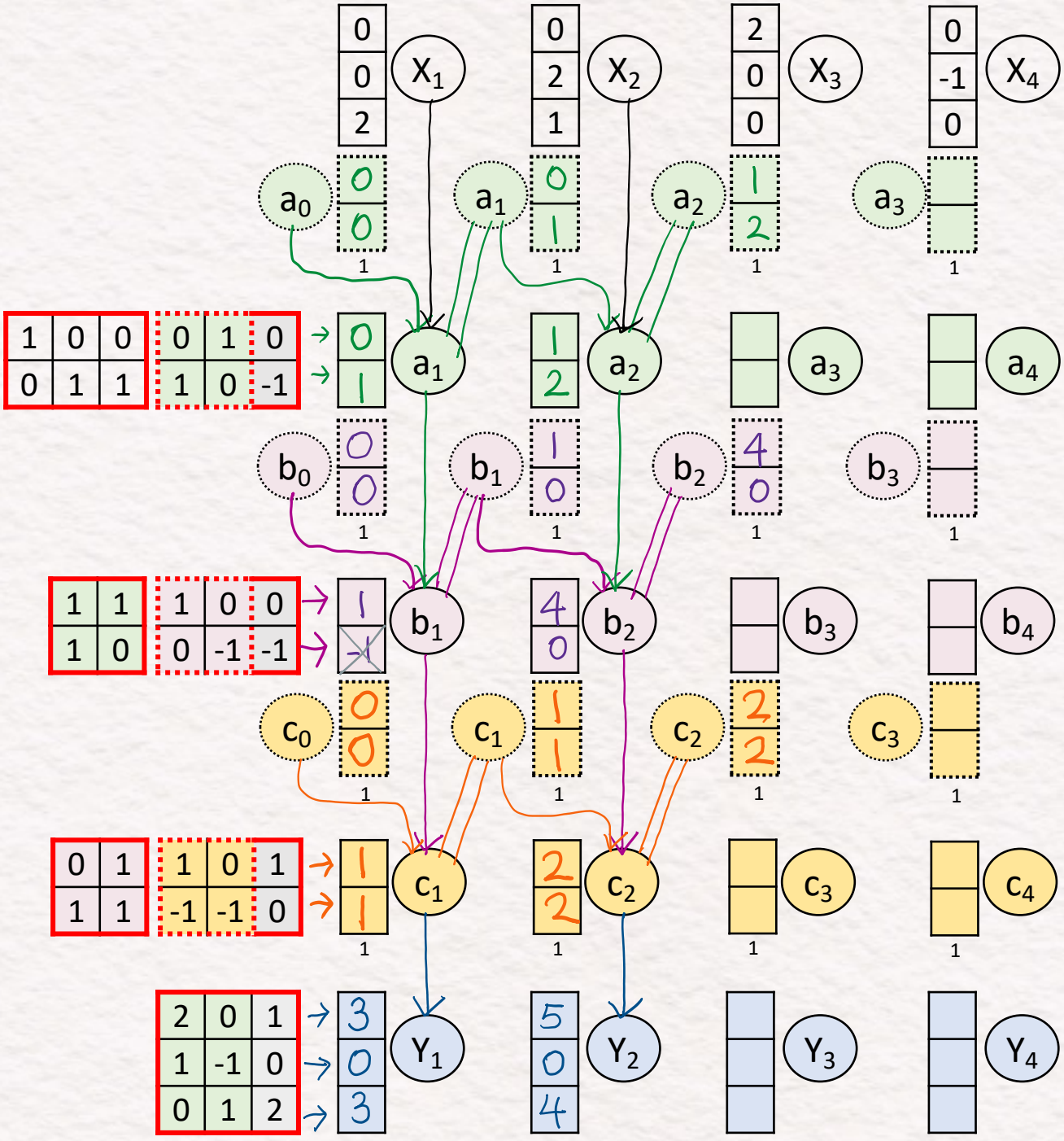
Deep RNN



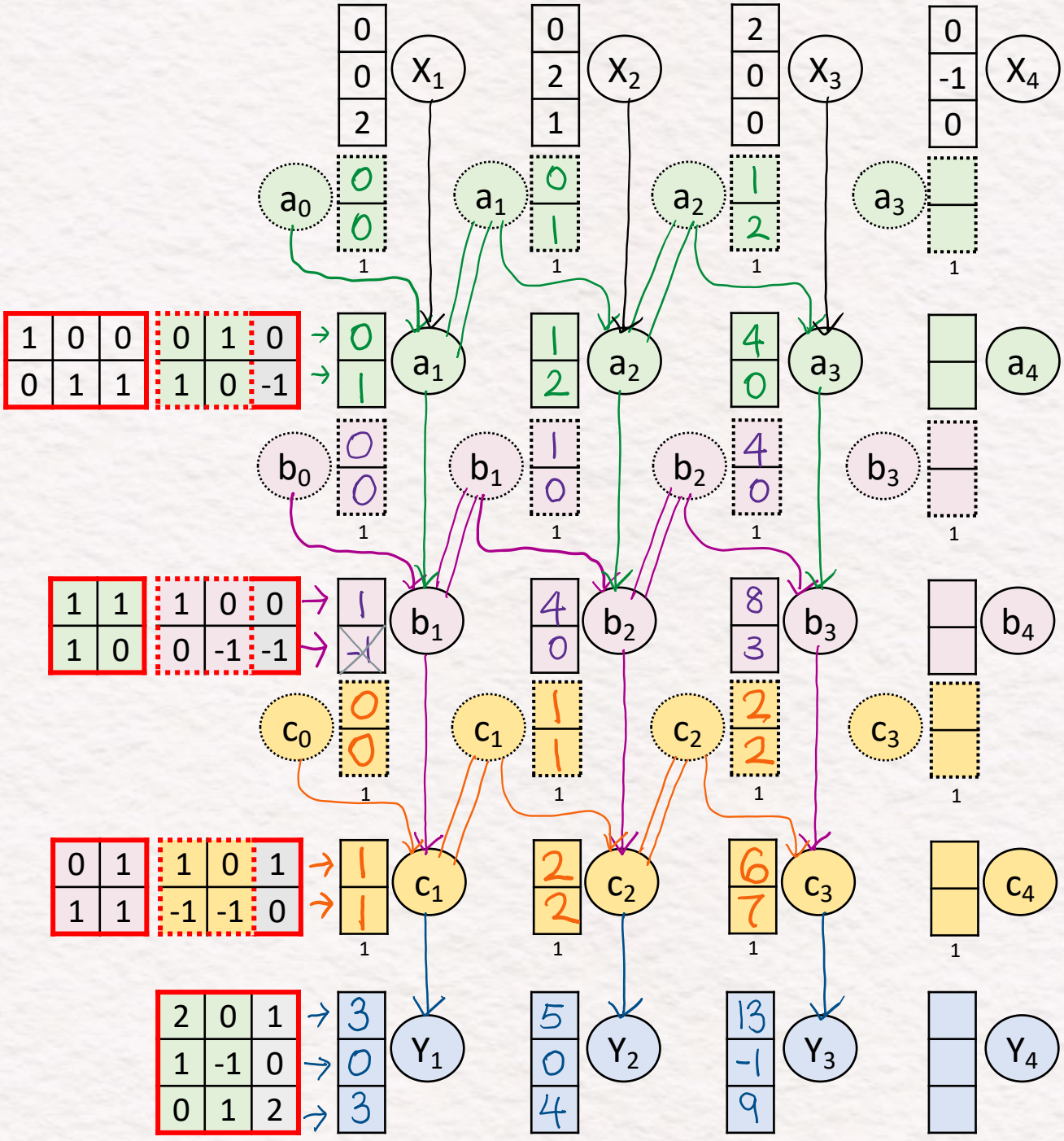
Deep RNN



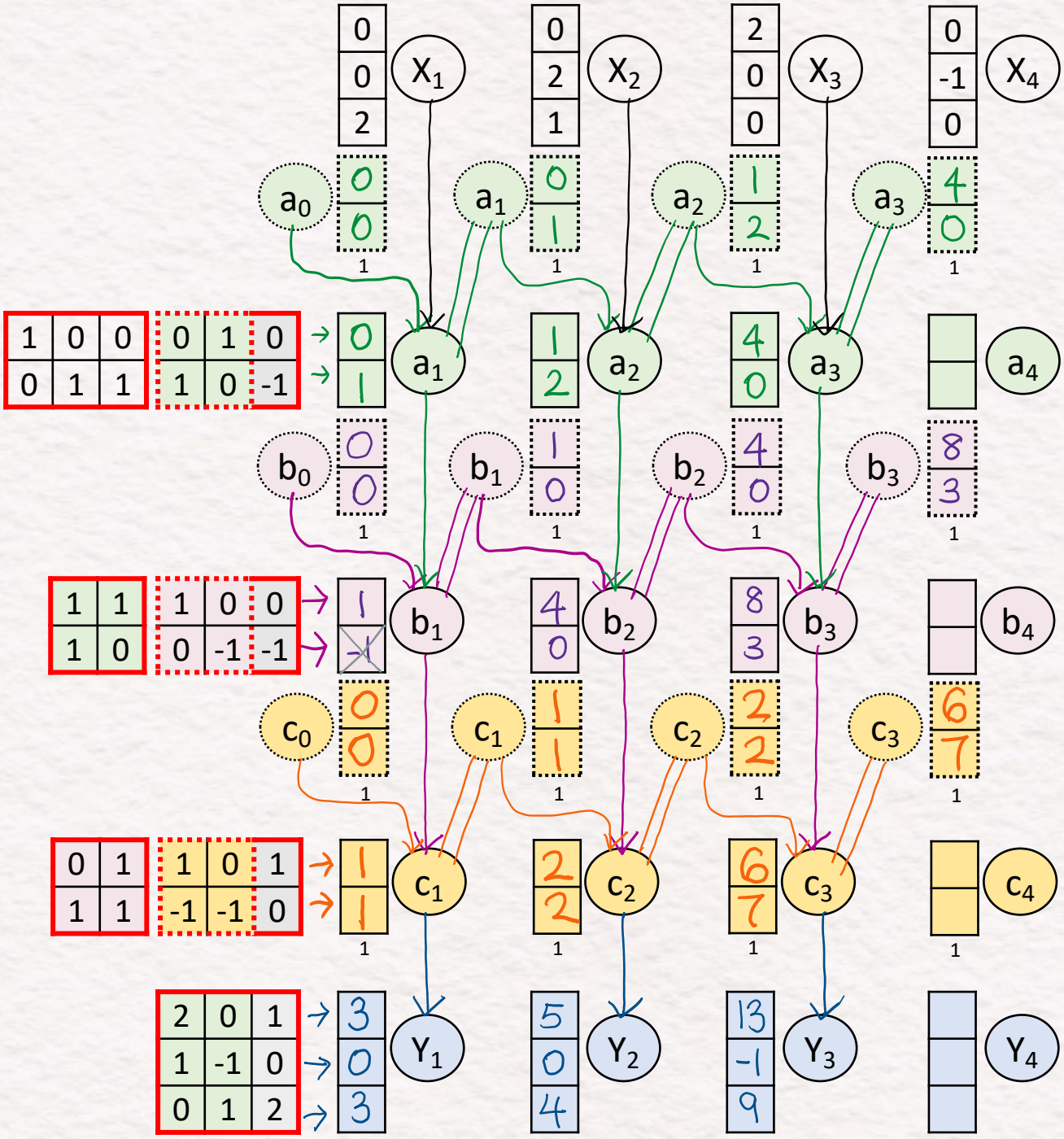
Deep RNN


X1
X2
X3
X4
a0
a1
a2
a3
a4
b0
b1
b2
b3
b4
c0
c1
c2
c3
c4
Y1
Y2
Y3
Y4
1 0 0
0 1 1
0 1 0
1 0 -1
1 1
1 0
1 0 0
0 -1 -1
0 1
1 1
1 0 1
-1 -1 0
2 0 1
1 -1 0
0 1 2

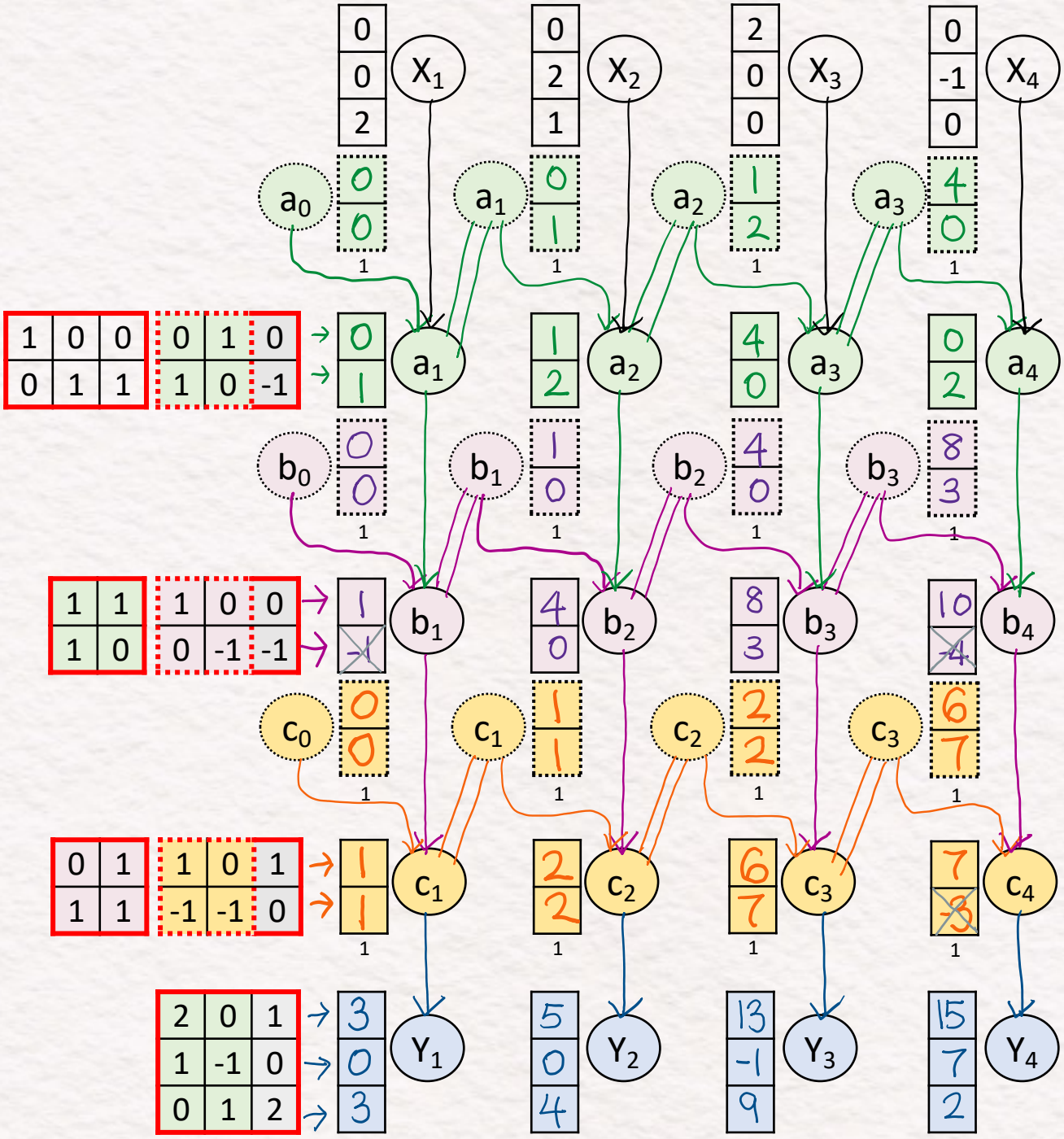
Deep RNN


X1
X2
X3
X4
a0
a1
a2
a3
a4
b0
b1
b2
b3
b4
c0
c1
c2
c3
c4
Y1
Y2
Y3
Y4
1 0 0
0 1 1
0 1 0
1 0 -1
1 1
1 0
1 0 0
0 -1 -1
0 1
1 1
1 0 1
-1 -1 0
2 0 1
1 -1 0
0 1 2

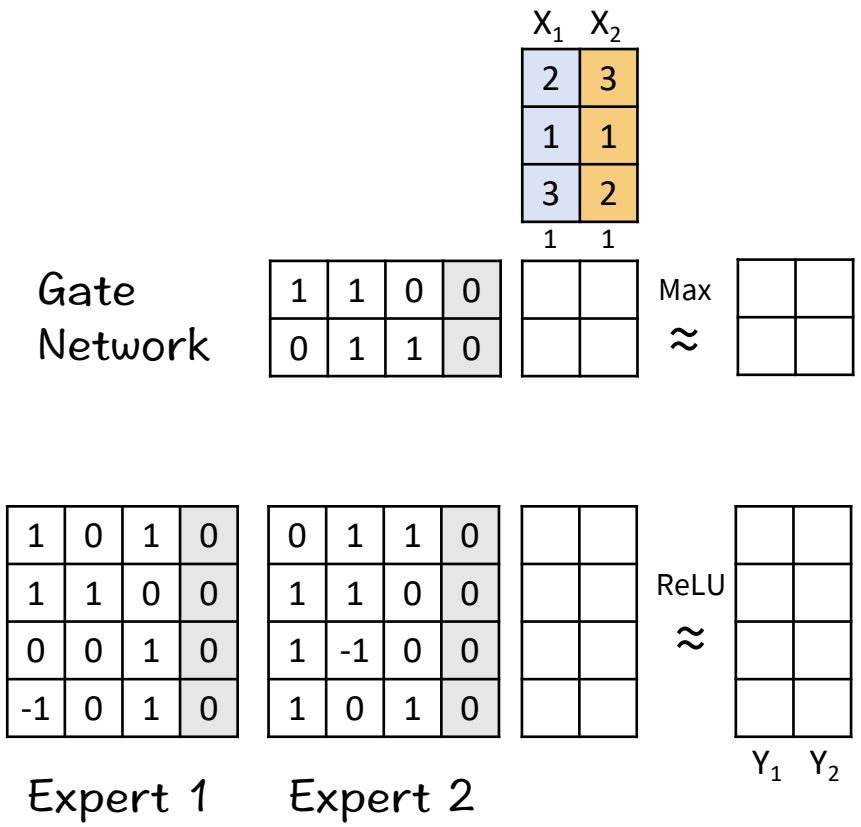
Deep RNN



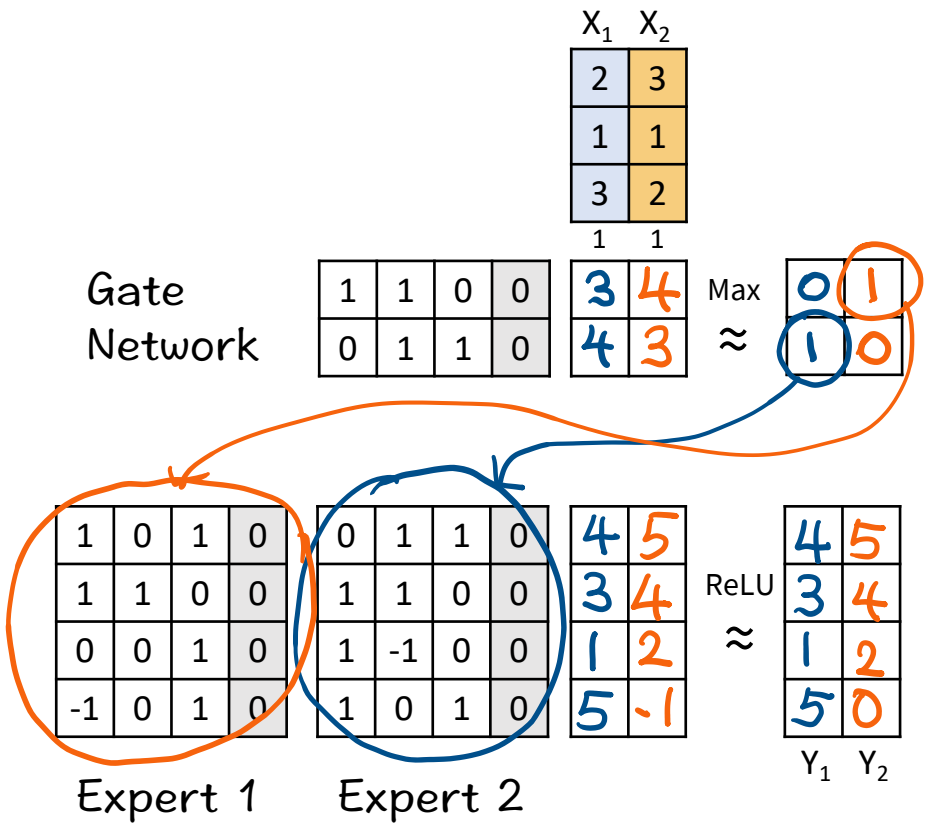
Deep RNN



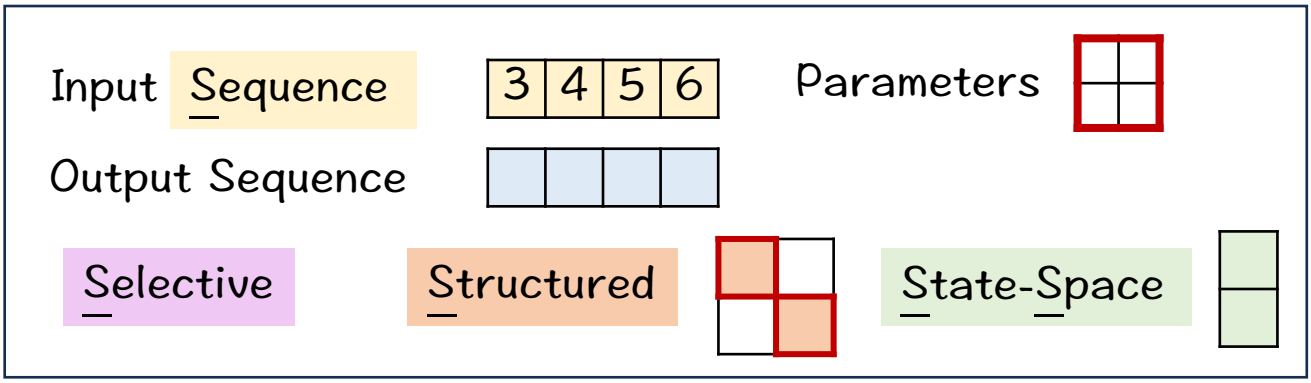
Mixture of Experts

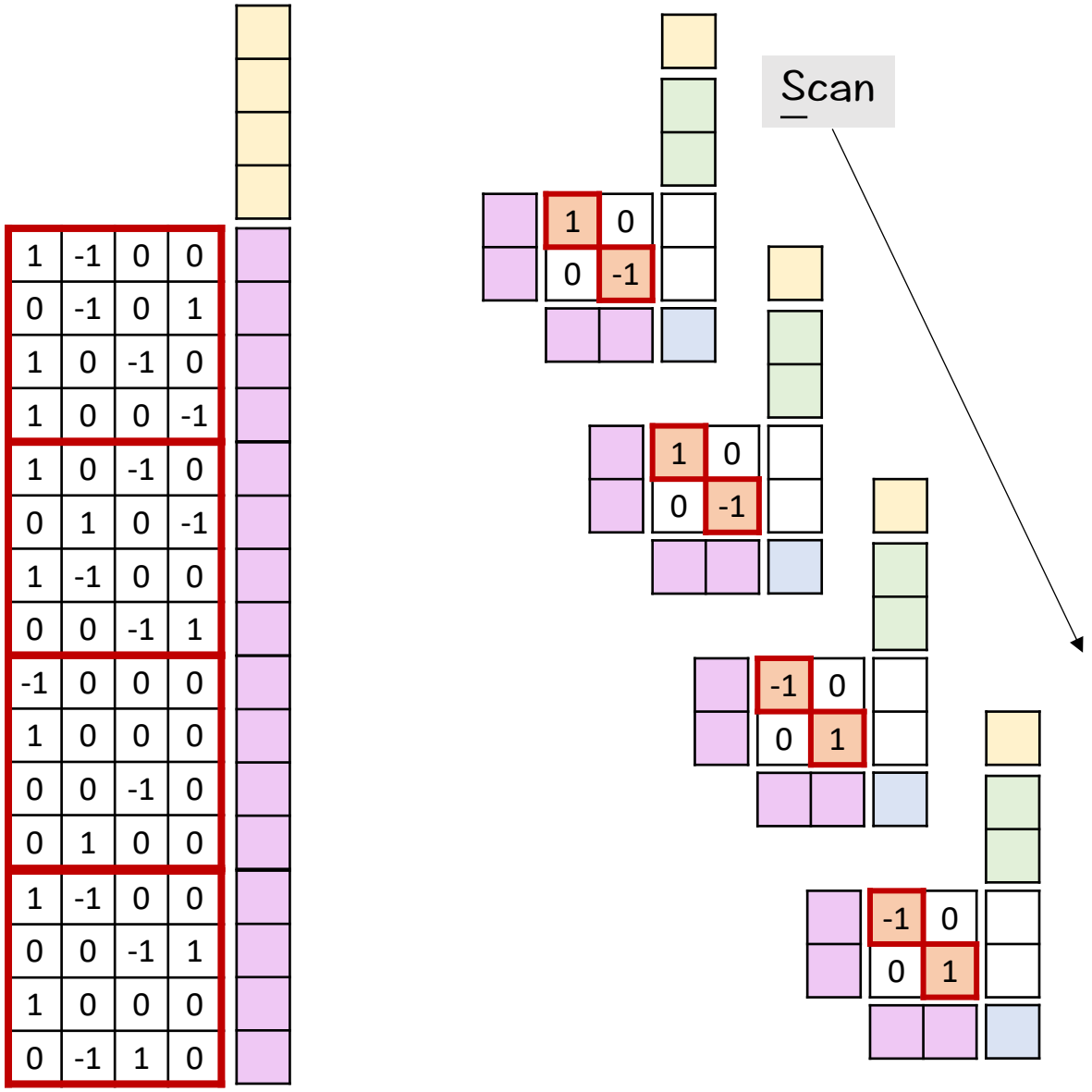


Mixture of Experts



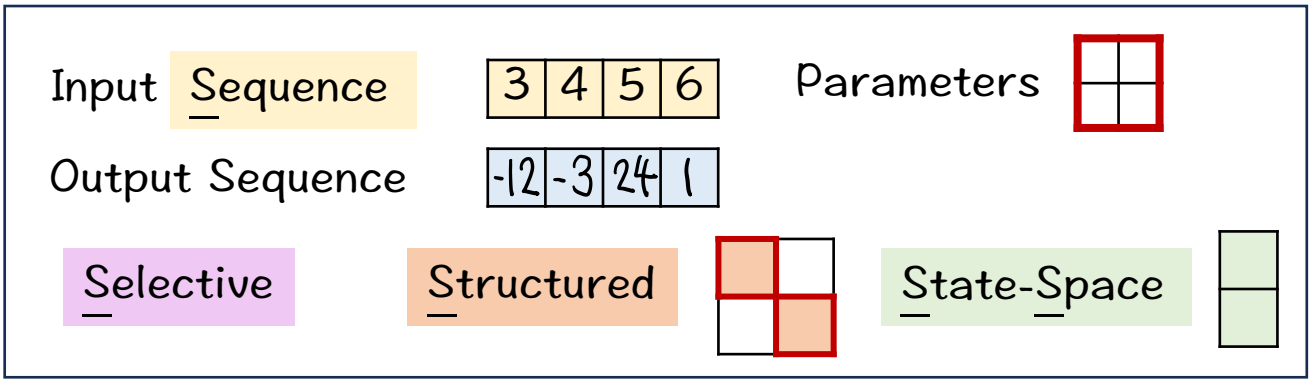
Mamba's S6 Model

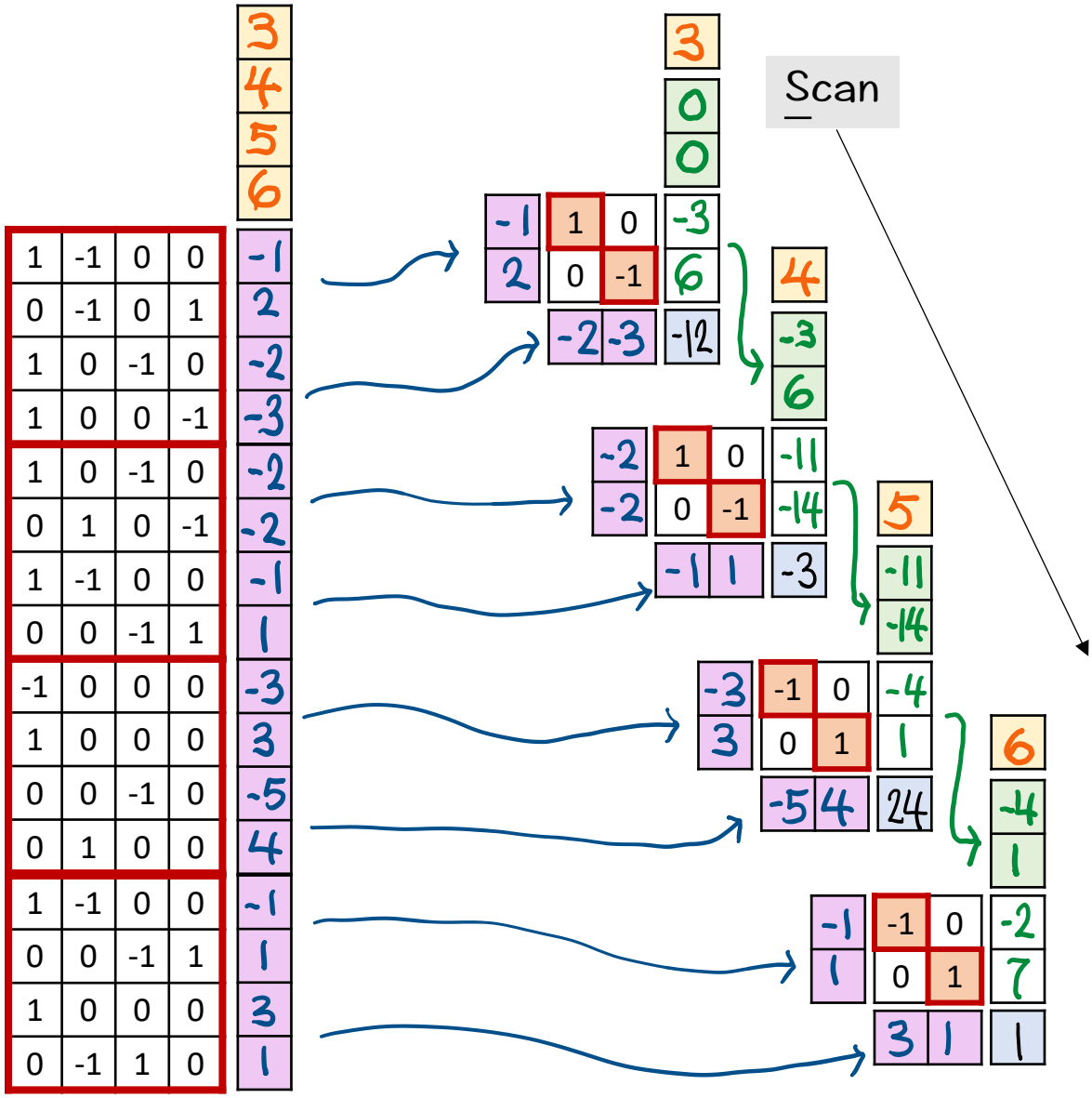




12.19.23

Mamba's S6 Model





12.19.23

Beginner's Guide to RAG

AI by Hand ✍️

Prof. Tom Yeh

University of Colorado Boulder

Hosted by:

SingleStore

Download Slides

https://by-hand.ai/rag

Roadmap

RAG Equation

Prompt

Index

Load

Split

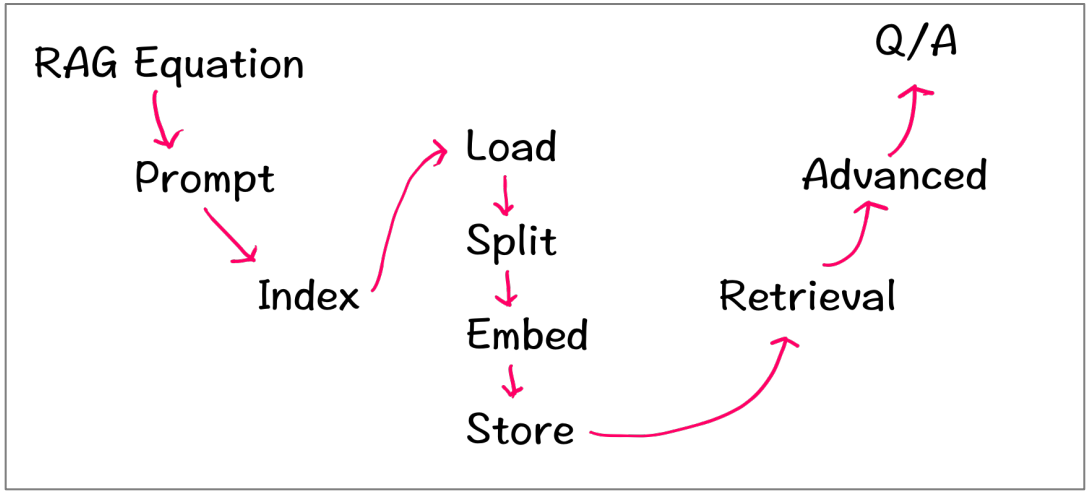
Embed

Store

Retrieval

Advanced

Q/A



RAG Equation

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado
Boulder

Download: https://by-hand.ai/rag

RAG = Query + Prompt + Context + LLM

How much does Tom make?

Search

Database

Amy, 100, CEO
Tom, 50, Prof
Sam, 120, CTO

```
[INST] <<SYS>> You are a _____, respectful, _____
assistant. Always answer as helpfully as possible,
while being safe. Your answers should ___ include
any harmful, _____, racist, sexist, toxic,
dangerous, or illegal content. Please ensure that
your responses are socially _____ and positive
in nature. <</SYS>>

Please answer the following question by using
information from the provided context information!
CONTEXT_INFORMATION:



QUESTION:

[/INST]
```

LLM

Download: https://by-hand.ai/rag

RAG = Query + ~~Prompt~~ + Context + LLM

How much does Tom make?

Search

```
[INST]
Please answer the following question by using
information from the provided context information!
CONTEXT_INFORMATION:


QUESTION:

[/INST]
```

Database

Amy, 100, CEO
Tom, 50, Prof
Sam, 120, CTO

LLM

Download: https://by-hand.ai/rag

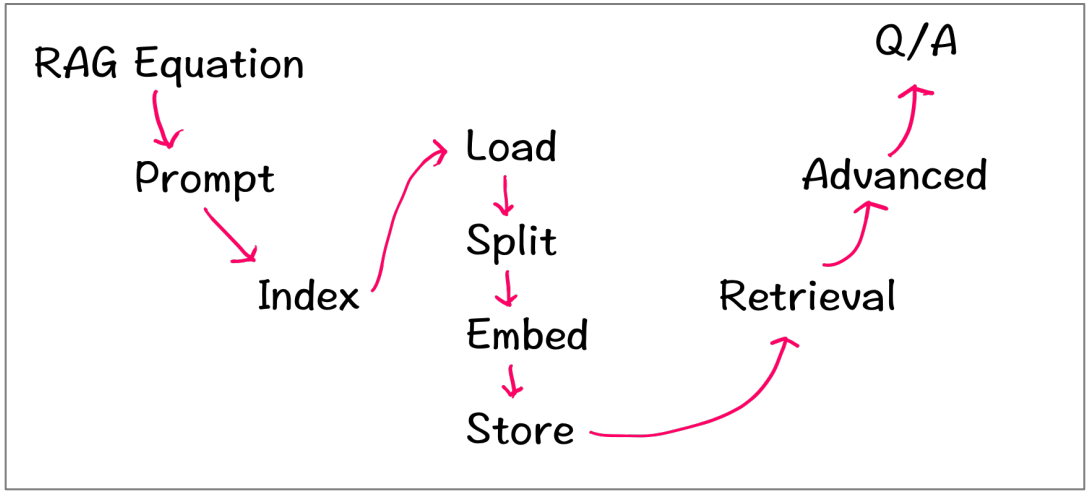
~~RAG~~ = Query + ~~Prompt~~ + ~~Context~~ + LLM

How much does Tom make?

Search

```
[INST]
QUESTION:

[/INST]
```

LLM



Prompt

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

Summarize

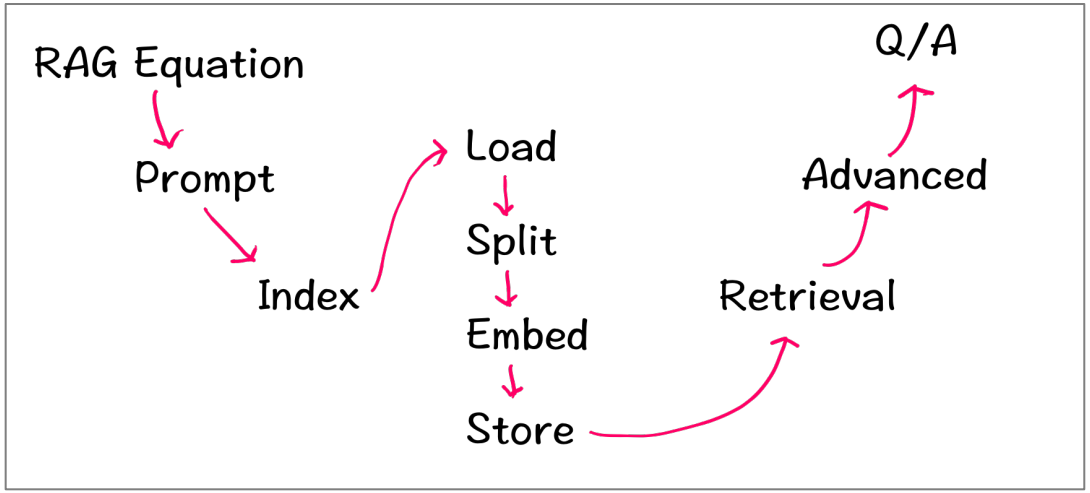
```
Context information from _____ sources is below.
---------------------

{context_str}

---------------------

Given the information from multiple sources and not
_____, answer the query.

Query: {query_str}

Answer:
```

Single Choice

```
Some choices are given below. It is provided in a numbered list
(1 to {num_chunks}),
where each item in the list corresponds to a summary.
---------------------

{context_list}

---------------------
Using only the choices above and not prior knowledge, return the
_________ that is most relevant to the question:

{query_str}

Provide choice in the following format: 'ANSWER: <number>' and
_____________ why this choice was selected in relation to the
question.
```

Refine

```
The original query is as follows: {query_str}
We have provided an existing answer: {existing_answer}
We have the opportunity to _________ the existing answer
with some more context below.
------------

{context_str}

------------
Given the new context, refine the original answer to better answer
the query. If the context isn't useful, return the _________ answer.

Refined Answer:
```



Index

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

Load

id: 1 date: 5/10	id: 2 date: 5/15	id: 2 date: 5/17
大人吃大包子，小孩吃小包子	每天都吃一個蘋果，可以保持健康	高山，大海，森林，動物，大自然

Split

	每天都	高山大
	吃一個	大海森
	蘋果可	林動物
	以保持	大自然

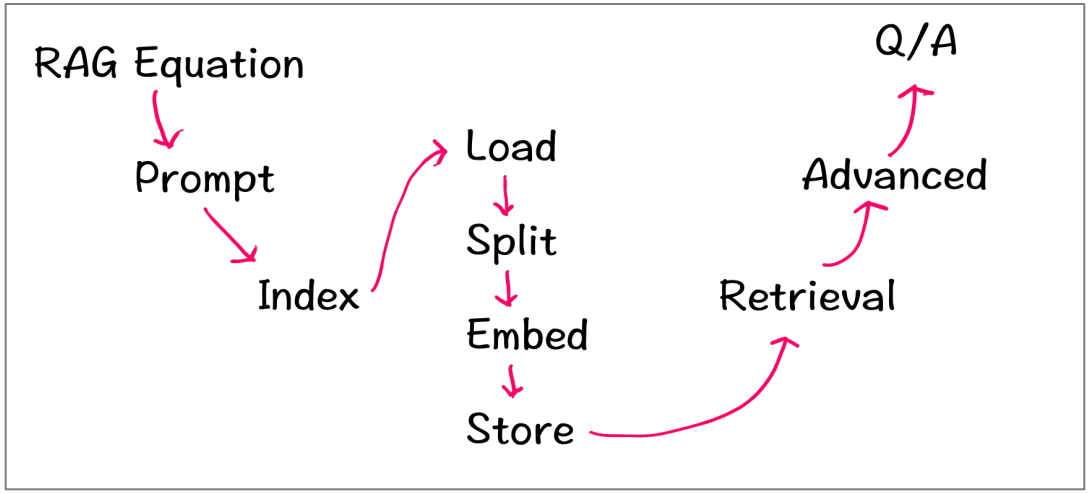
Embed

1	1	0		1	1	-2			
1	0	1		1	1	1			
0	0	1		0	0	1			
0	1	0		0	1	1			

Store

Chunk	Embed	Meta
		{id:_, date: _____}
		{id:_, date: _____}

Download: https://by-hand.ai/rag



Load

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

LlamaHub Data Loaders

LlamaHub
Powered by LlamaIndex

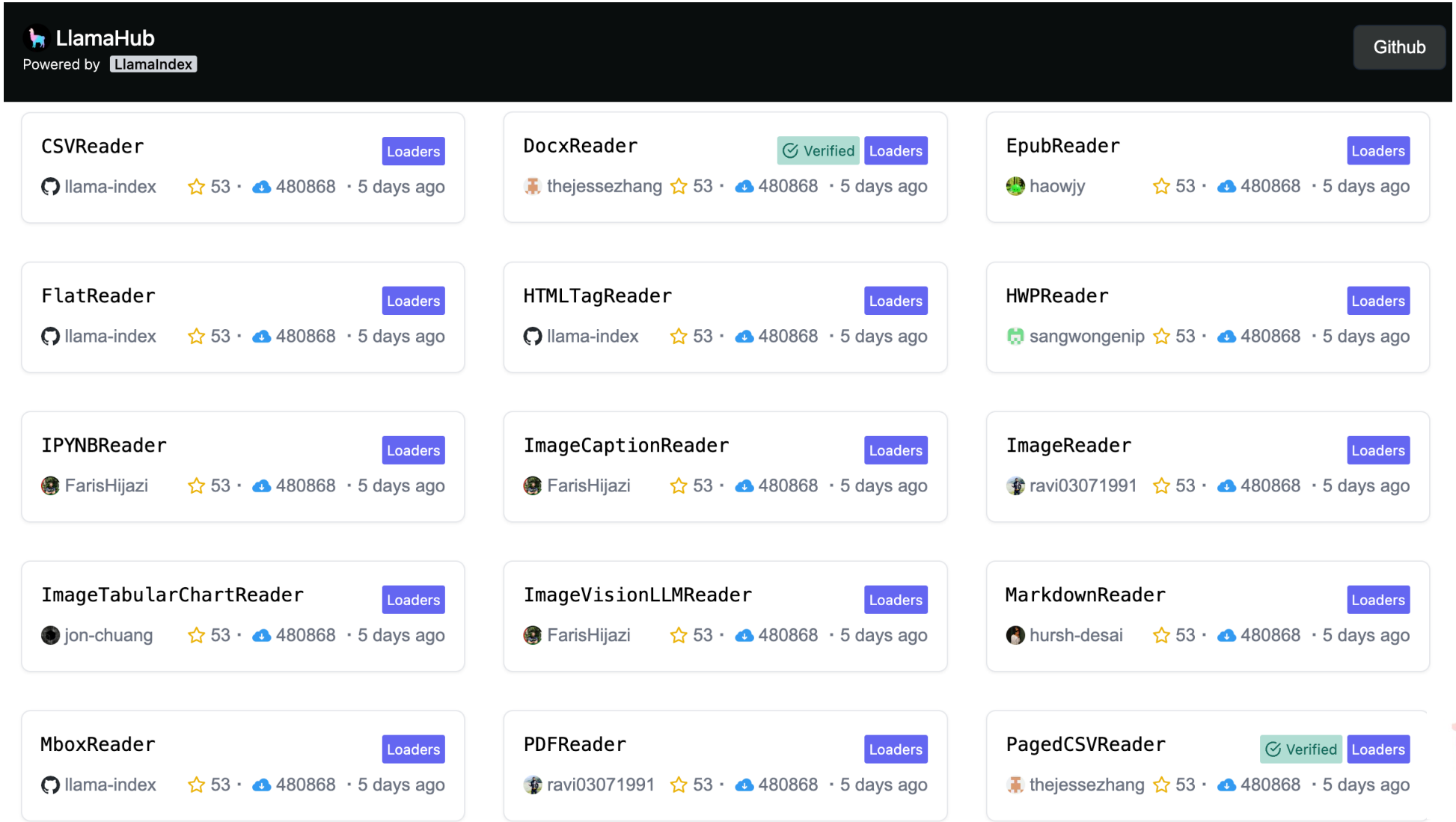
Github

CSVReader Loaders llama-index ☆ 53 · 480868 · 5 days ago	DocxReader Verified Loaders thejessezhang ☆ 53 · 480868 · 5 days ago	EpubReader Loaders haowjy ☆ 53 · 480868 · 5 days ago
FlatReader Loaders llama-index ☆ 53 · 480868 · 5 days ago	HTMLTagReader Loaders llama-index ☆ 53 · 480868 · 5 days ago	HWPReader Loaders sangwongenip ☆ 53 · 480868 · 5 days ago
IPYNBReader Loaders FarisHijazi ☆ 53 · 480868 · 5 days ago	ImageCaptionReader Loaders FarisHijazi ☆ 53 · 480868 · 5 days ago	ImageReader Loaders ravi03071991 ☆ 53 · 480868 · 5 days ago
ImageTabularChartReader Loaders jon-chuang ☆ 53 · 480868 · 5 days ago	ImageVisionLLMReader Loaders FarisHijazi ☆ 53 · 480868 · 5 days ago	MarkdownReader Loaders hursh-desai ☆ 53 · 480868 · 5 days ago
MboxReader Loaders llama-index ☆ 53 · 480868 · 5 days ago	PDFReader Loaders ravi03071991 ☆ 53 · 480868 · 5 days ago	PagedCSVReader Verified Loaders thejessezhang ☆ 53 · 480868 · 5 days ago

LlamaIndex

Download: https://by-hand.ai/rag

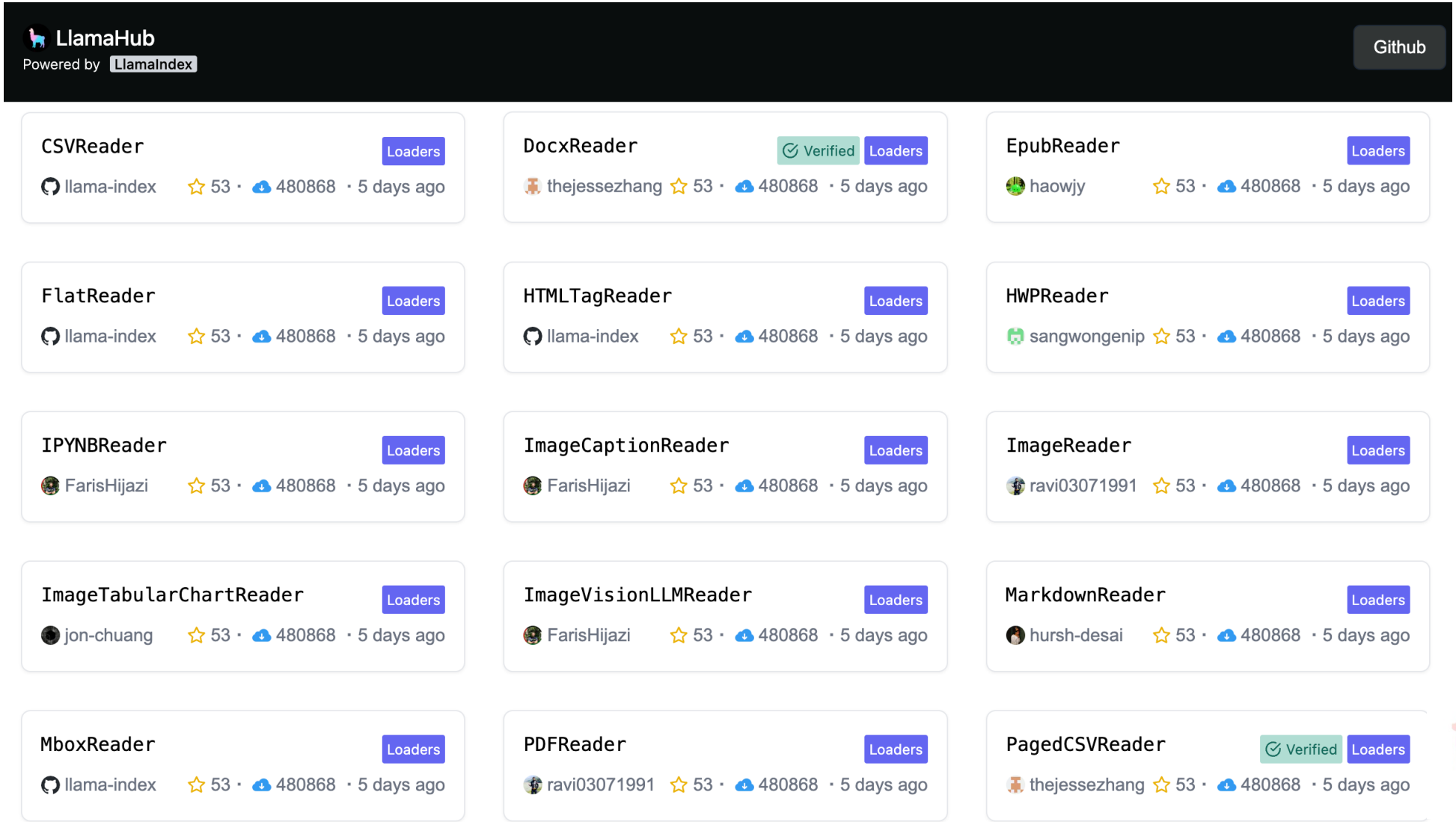
Which loader to use to read CSV data?



LlamaIndex

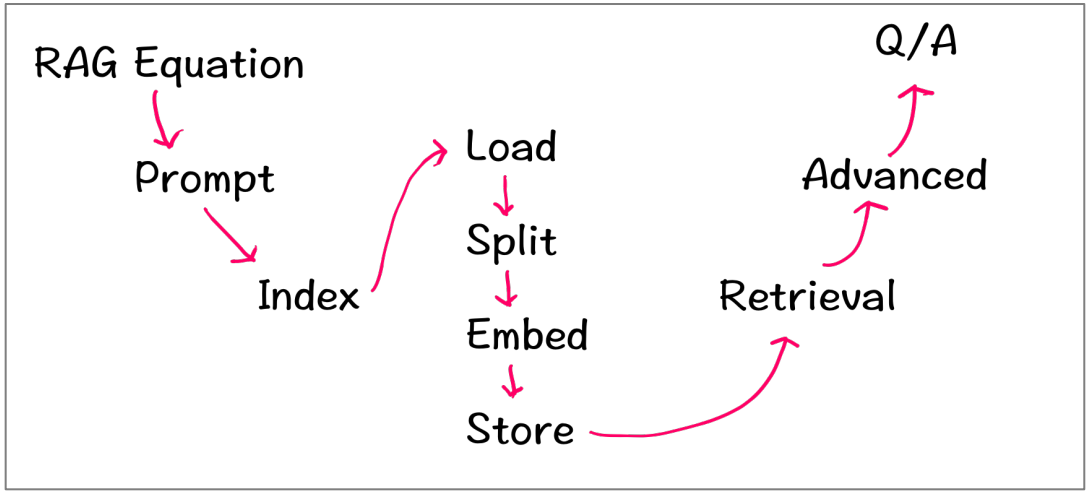
Download: https://by-hand.ai/rag

Which loader to use to read Word documents?



LlamaIndex

Download: https://by-hand.ai/rag



Split

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

Split by Tokens

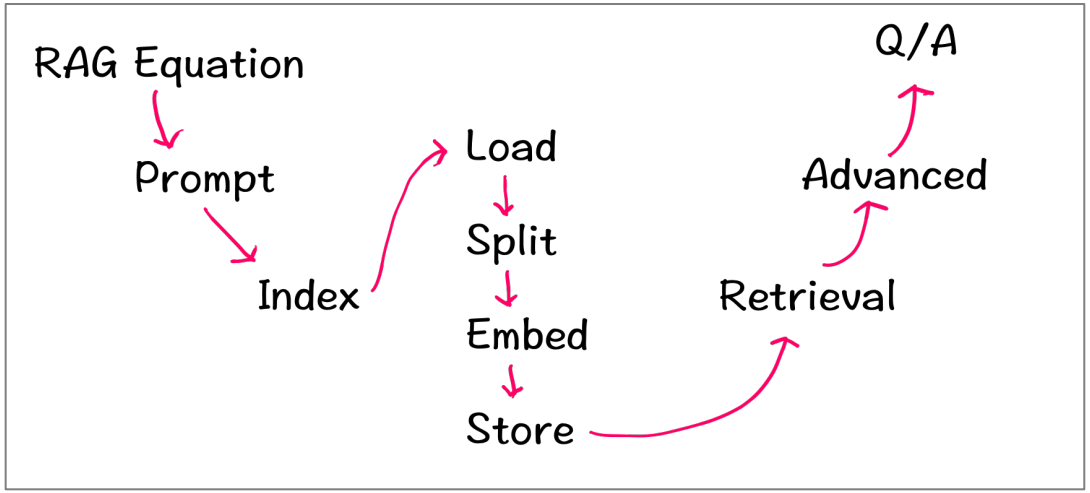
```
TokenTextSplitter(
    chunk_size=4,
    chunk_overlap=0
)
```

How are you doing today? I hope you're doing well. Whether your day is busy or calm, I hope you find a moment of joy. Life moves fast, with many ups and downs. It's important to stop and enjoy the little things. Maybe today brought a good chat, a tasty meal, or a beautiful sunset. Cherish these happy moments. Each day is a chance to make good memories and connect with others. Take time to appreciate your surroundings and the people in your life. Even on challenging days, there are opportunities to learn and grow. Let's embrace each experience and look forward to tomorrow with optimism.

Split by Tokens

```
TokenTextSplitter(
    chunk_size=4,
    chunk_overlap=1
)
```

How are you doing today? I hope you're doing well. Whether your day is busy or calm, I hope you find a moment of joy. Life moves fast, with many ups and downs. It's important to stop and enjoy the little things. Maybe today brought a good chat, a tasty meal, or a beautiful sunset. Cherish these happy moments. Each day is a chance to make good memories and connect with others. Take time to appreciate your surroundings and the people in your life. Even on challenging days, there are opportunities to learn and grow. Let's embrace each experience and look forward to tomorrow with optimism.



Embed

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

Massive Text Embedding Benchmark

https://huggingface.co/spaces/mteb/leaderboard

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

Which models have 7B parameters?

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

Download: https://by-hand.ai/rag

Which model has the most parameters?

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

Download: https://by-hand.ai/rag

Which model supports the longest context window?

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

Download: https://by-hand.ai/rag

Which model has the best classification performance?

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

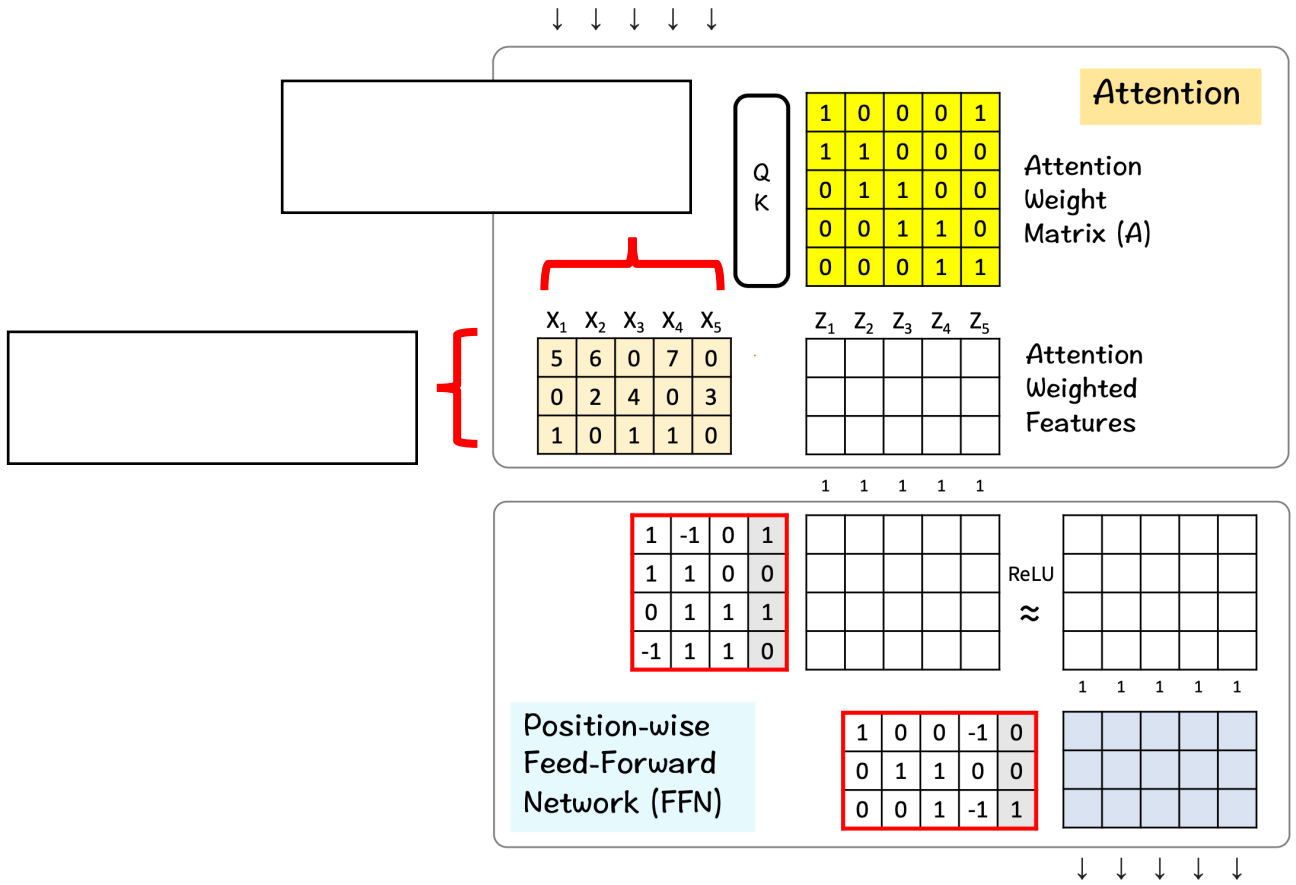
Download: https://by-hand.ai/rag

Which model has the best clustering performance?

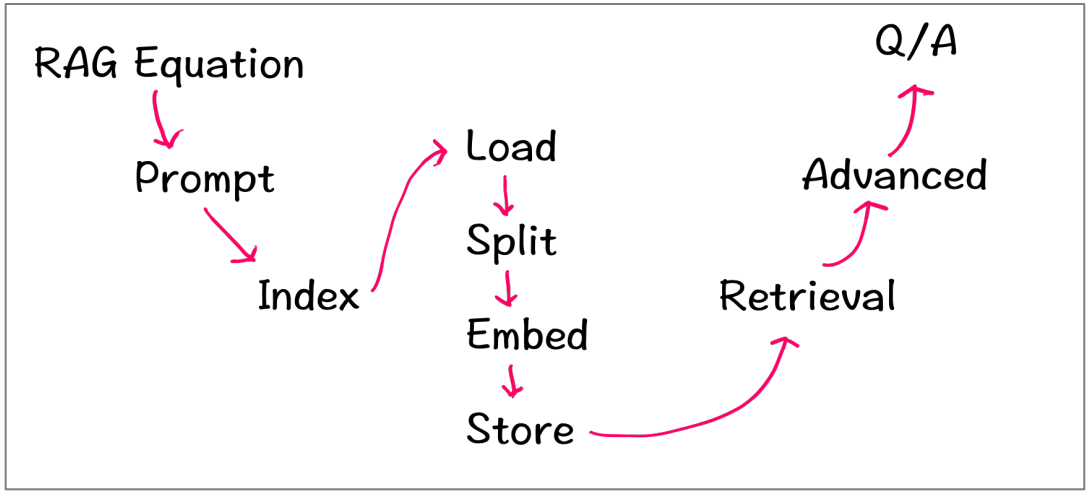
Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35
2	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67
3	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83
4	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42
5	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61
6	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26
7	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48
8	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14
9	gte-large-en-v1.5	434	1.62	1024	8192	65.39	77.75	47.96
10	LLM2Vec-Meta-Llama-3-supervis	7505	27.96	4096	8192	65.01	75.92	46.45

Download: https://by-hand.ai/rag

	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clusterin Average (11 datasets)
	text-embedding-3-large			3072	8191	64.59	75.45	49.01



Download: https://by-hand.ai/rag



Store

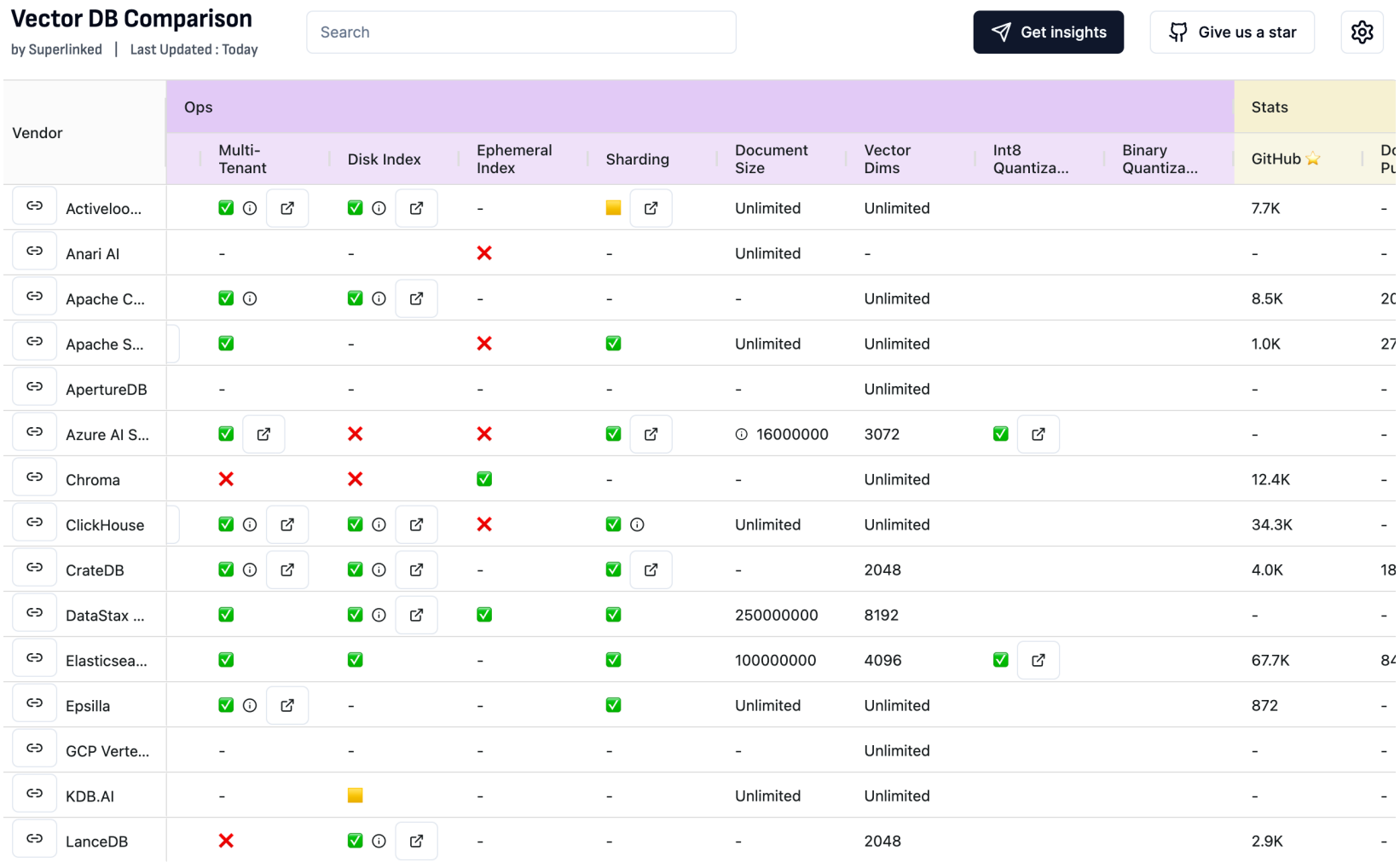
Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

Choose a Vector Database

https://superlinked.com/vector-db-comparison



Vector DB Comparison

by Superlinked | Last Updated : Today

Search

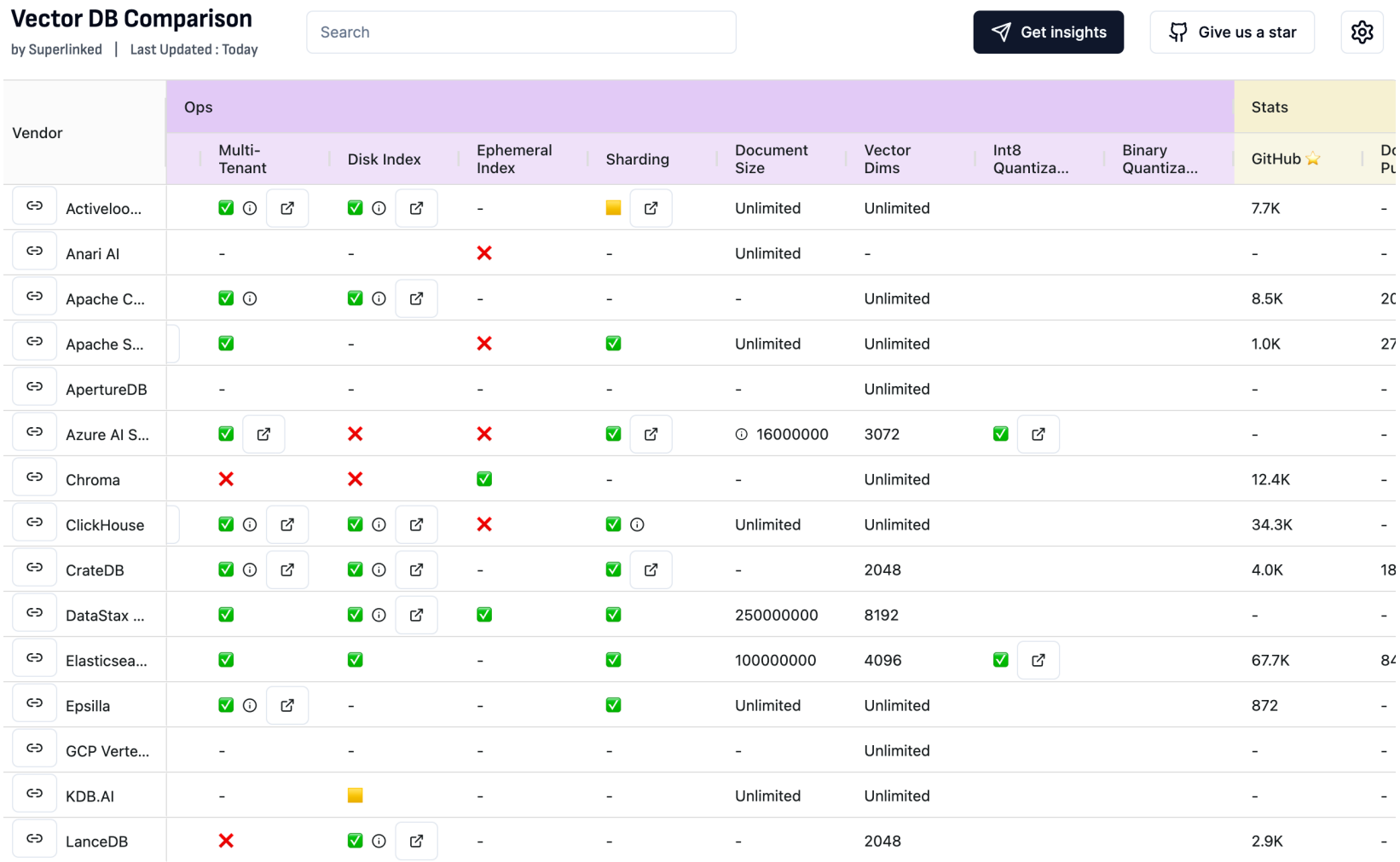
Get insights | Give us a star

Vendor	Ops								Stats	
	Multi-Tenant	Disk Index	Ephemeral Index	Sharding	Document Size	Vector Dims	Int8 Quantiza...	Binary Quantiza...	GitHub ⭐	Do Pu
Activeloo...	✅	✅	-	🟨	Unlimited	Unlimited			7.7K	-
Anari AI	-	-	❌	-	Unlimited	-			-	-
Apache C...	✅	✅	-	-	-	Unlimited			8.5K	20
Apache S...	✅	-	❌	✅	Unlimited	Unlimited			1.0K	27
ApertureDB	-	-	-	-	-	Unlimited			-	-
Azure AI S...	✅	❌	❌	✅	16000000	3072	✅		-	-
Chroma	❌	❌	✅	-	-	Unlimited			12.4K	-
ClickHouse	✅	✅	❌	✅	Unlimited	Unlimited			34.3K	-
CrateDB	✅	✅	-	✅	-	2048			4.0K	18
DataStax ...	✅	✅	✅	✅	250000000	8192			-	-
Elasticsea...	✅	✅	-	✅	100000000	4096	✅		67.7K	84
Epsilla	✅	-	-	✅	Unlimited	Unlimited			872	-
GCP Verte...	-	-	-	-	-	Unlimited			-	-
KDB.AI	-	🟨	-	-	Unlimited	Unlimited			-	-
LanceDB	❌	✅	-	-	-	2048			2.9K	-

Superlinked

Download: https://by-hand.ai/rag

Which database can fit 3072 dimensions?



Vector DB Comparison

by Superlinked | Last Updated : Today

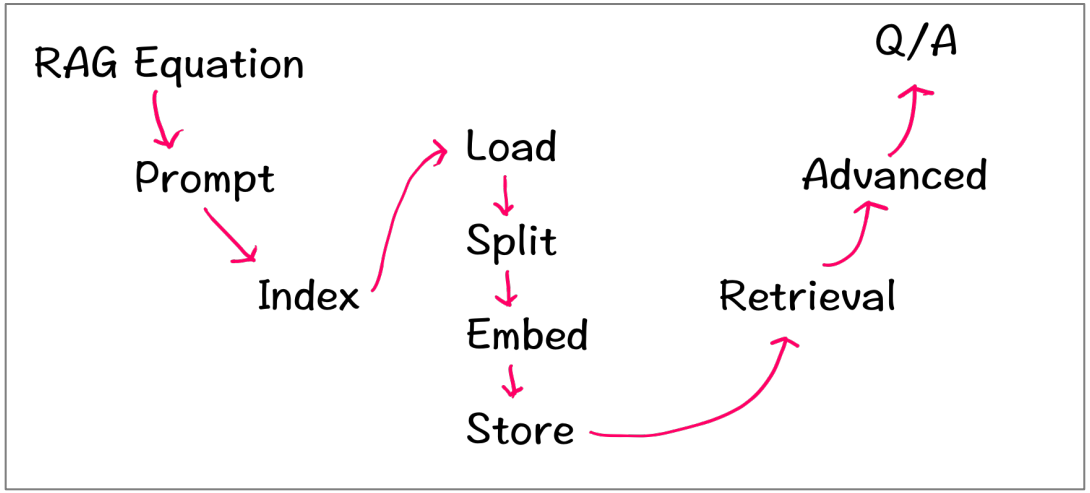
Search

Get insights | Give us a star

Vendor	Ops								Stats	
	Multi-Tenant	Disk Index	Ephemeral Index	Sharding	Document Size	Vector Dims	Int8 Quantiza...	Binary Quantiza...	GitHub ⭐	Do Pu
Activeloo...	✅	✅	-	🟨	Unlimited	Unlimited			7.7K	-
Anari AI	-	-	❌	-	Unlimited	-			-	-
Apache C...	✅	✅	-	-	-	Unlimited			8.5K	20
Apache S...	✅	-	❌	✅	Unlimited	Unlimited			1.0K	27
ApertureDB	-	-	-	-	-	Unlimited			-	-
Azure AI S...	✅	❌	❌	✅	16000000	3072	✅		-	-
Chroma	❌	❌	✅	-	-	Unlimited			12.4K	-
ClickHouse	✅	✅	❌	✅	Unlimited	Unlimited			34.3K	-
CrateDB	✅	✅	-	✅	-	2048			4.0K	18
DataStax ...	✅	✅	✅	✅	250000000	8192			-	-
Elasticsea...	✅	✅	-	✅	100000000	4096	✅		67.7K	84
Epsilla	✅	-	-	✅	Unlimited	Unlimited			872	-
GCP Verte...	-	-	-	-	-	Unlimited			-	-
KDB.AI	-	🟨	-	-	Unlimited	Unlimited			-	-
LanceDB	❌	✅	-	-	-	2048			2.9K	-

Superlinked

Download: https://by-hand.ai/rag



Retrieval

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

SQL — Retrieve by Similarity

```
_________ comment, emb<___>[0, 2, 1, 1] AS score
FROM posts
ORDER BY ______  ASC | DESC ;
```

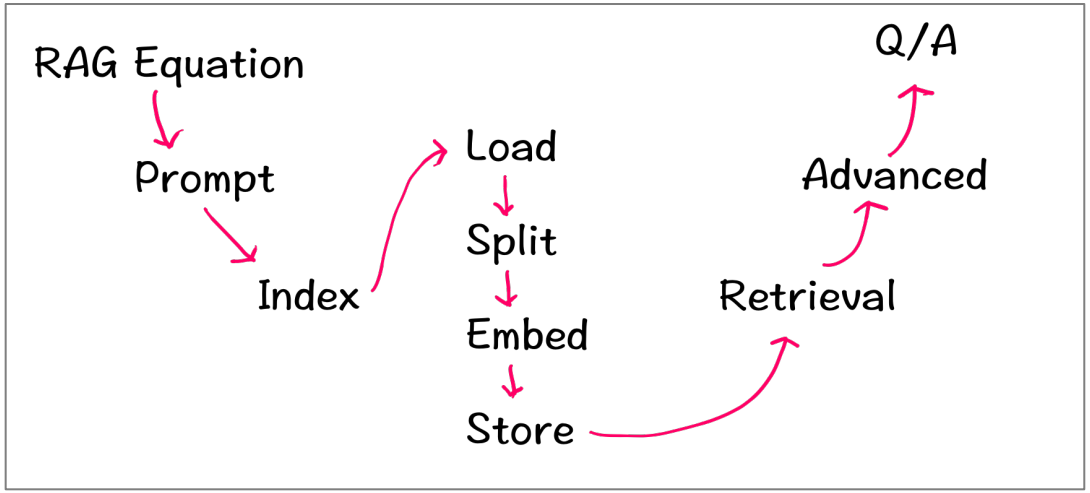
SingleStore

High-level API: Superlinked 

```
query = Query(post_index)
    .__________(post)
    .__________(relevance_space.text, Param("_________"))

app.query(query,  _________ = "how are you?" )
```

Superlinked

Download: https://by-hand.ai/rag



Advanced RAGs

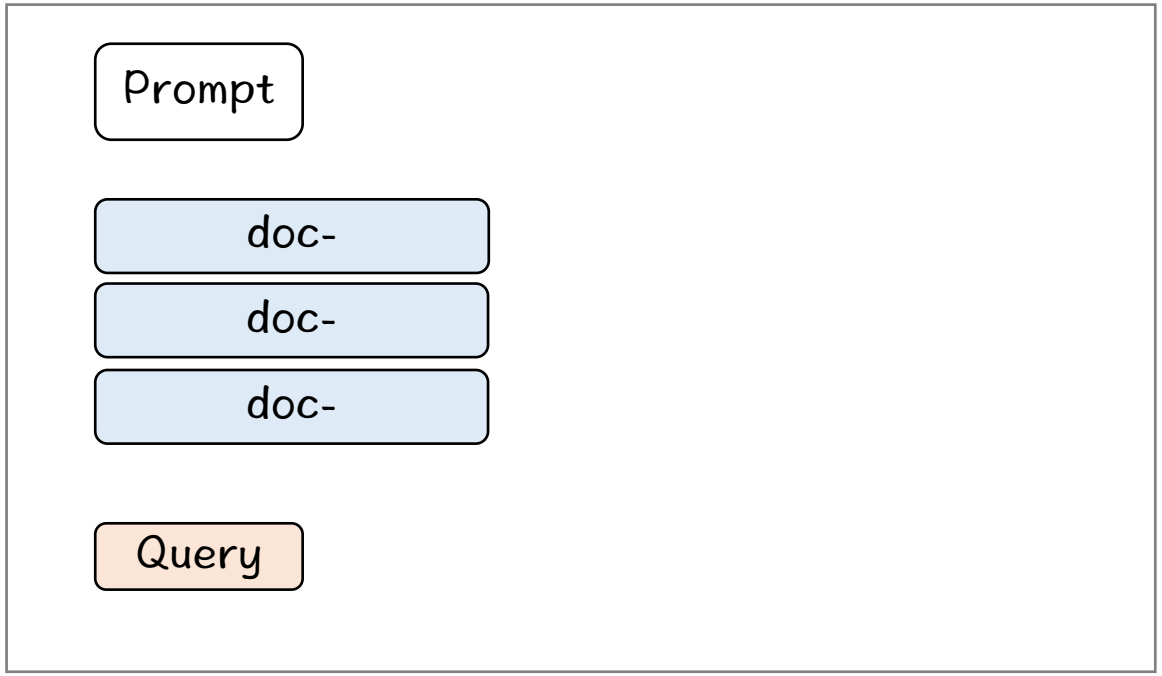
Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado
Boulder

Download: https://by-hand.ai/rag

RAG

DB

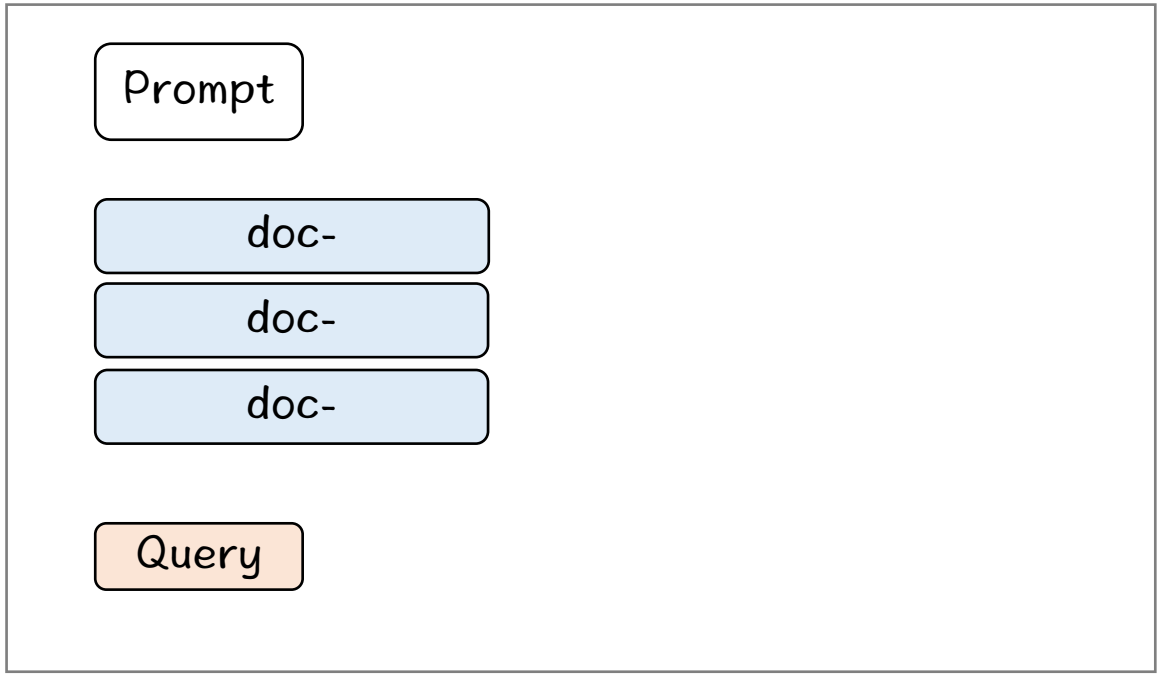


LLM

Download: https://by-hand.ai/rag

Multiple Embedding Spaces

DB



LLM

Download: https://by-hand.ai/rag

Superlinked Vector Compute

Search

Output

df = result._____()

_____(df)

```
result = app.query(
    simple_query,
    query_text=[redacted]
    description_weight=1,
    headline_weight=1,
    recency_weight=0,
)
```

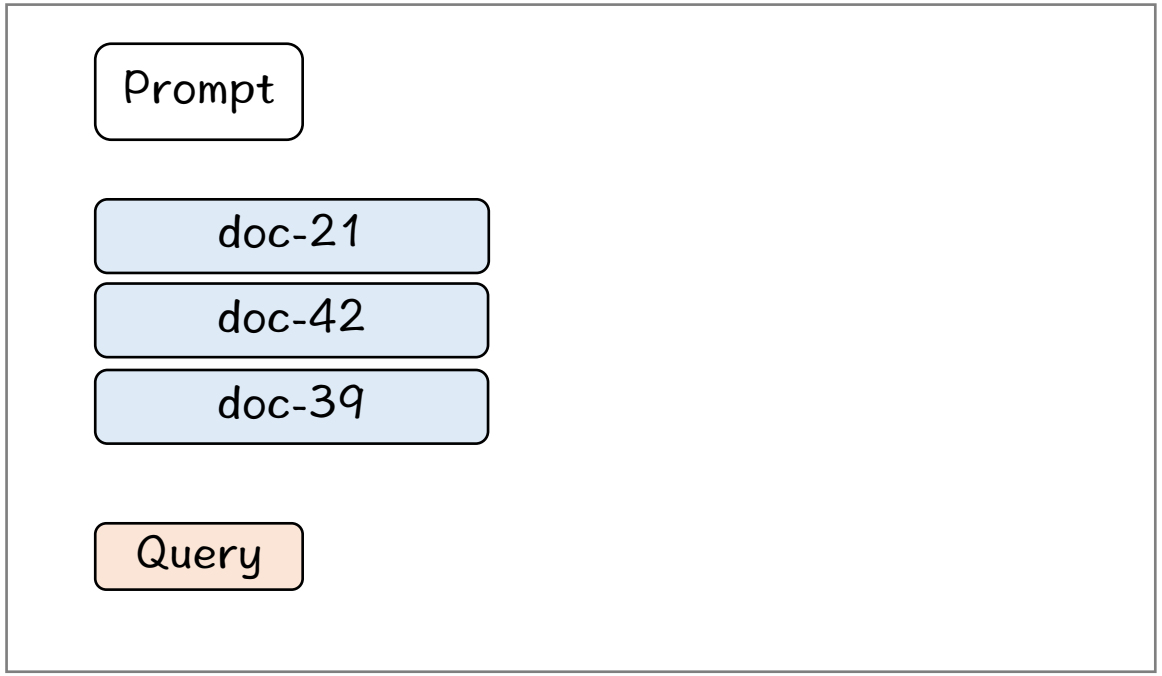
Memory

ID	Description	Headline	Recency
1	0 1 1 0	0 1 0	1 0
2	1 1 0 0	1 0 0	1 0
3	0 1 1 1	0 0 1	0 1

Download: https://by-hand.ai/rag

RankGPT

DB



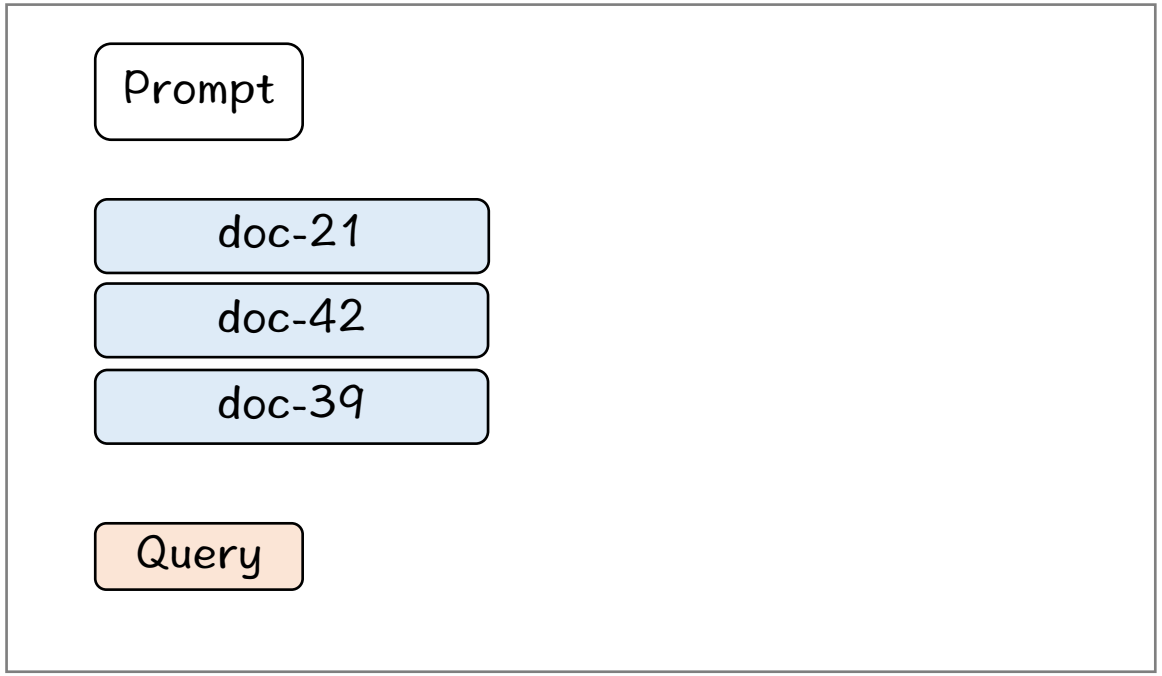
LLM

RankGPT

```
Search Query: {query}.

Rank the {num} passages above based on their ____________ to the
search query. The passages should be listed in ____________ order
using identifiers.

The most relevant passages should be listed first.

The output format should be [] > [], e.g., [1] > [2].

Only response the ranking results,

do not ____________.
```

Multi-Query Retrieval

DB



LLM

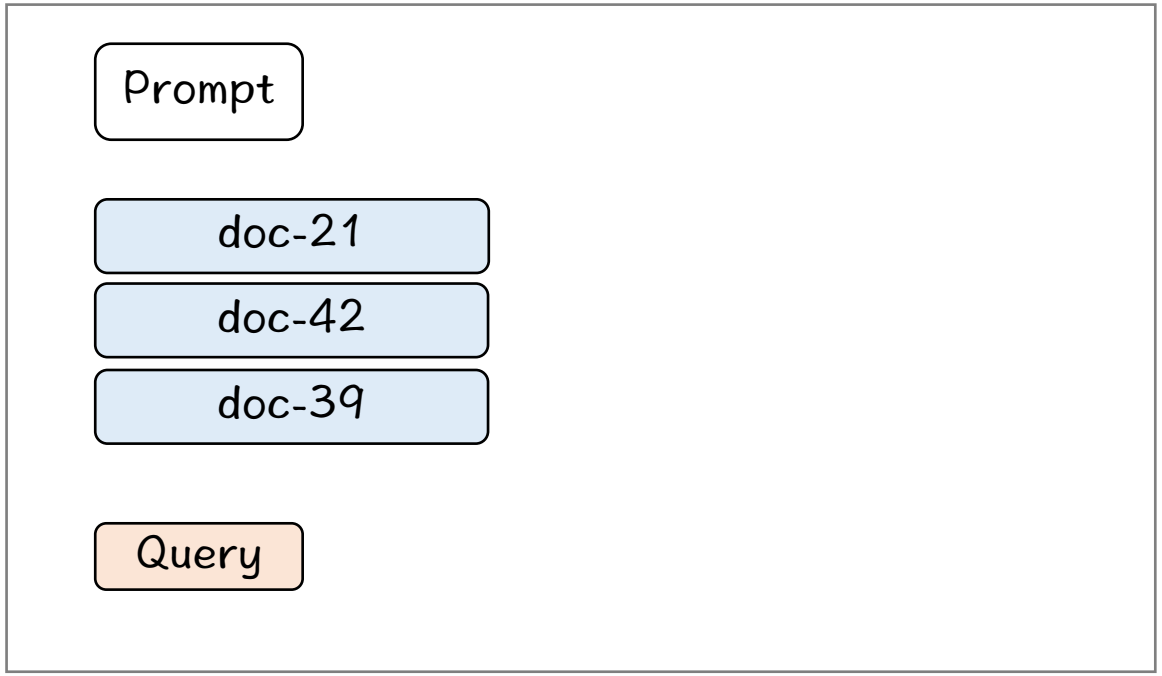
Download: https://by-hand.ai/rag

Multi-Query Retrieval

```
You are an AI language model assistant. Your task is to generate ____
different versions of the given user question to retrieve relevant
documents from a vector database.

By generating multiple _____ on the user question, your goal
is to help the user overcome some of the limitations of _____
search.

Provide these alternative questions seperated by newlines.

Original question: {question}
```

Contextual Compression

DB

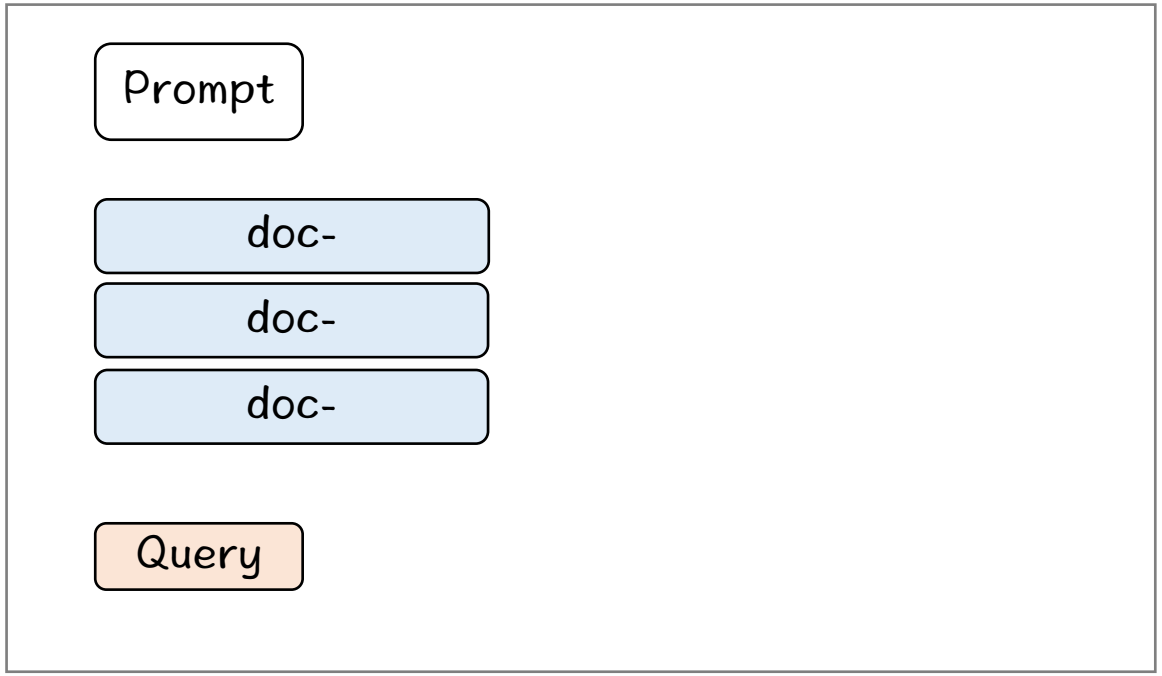


LLM

Download: https://by-hand.ai/rag

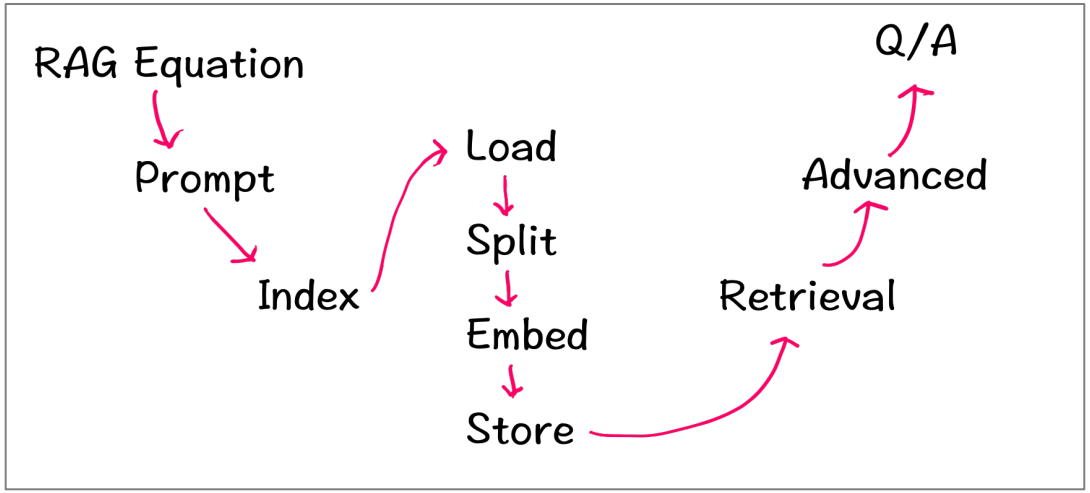
Hypothetical Document Embeddings (HyDE)

DB



LLM

Download: https://by-hand.ai/rag



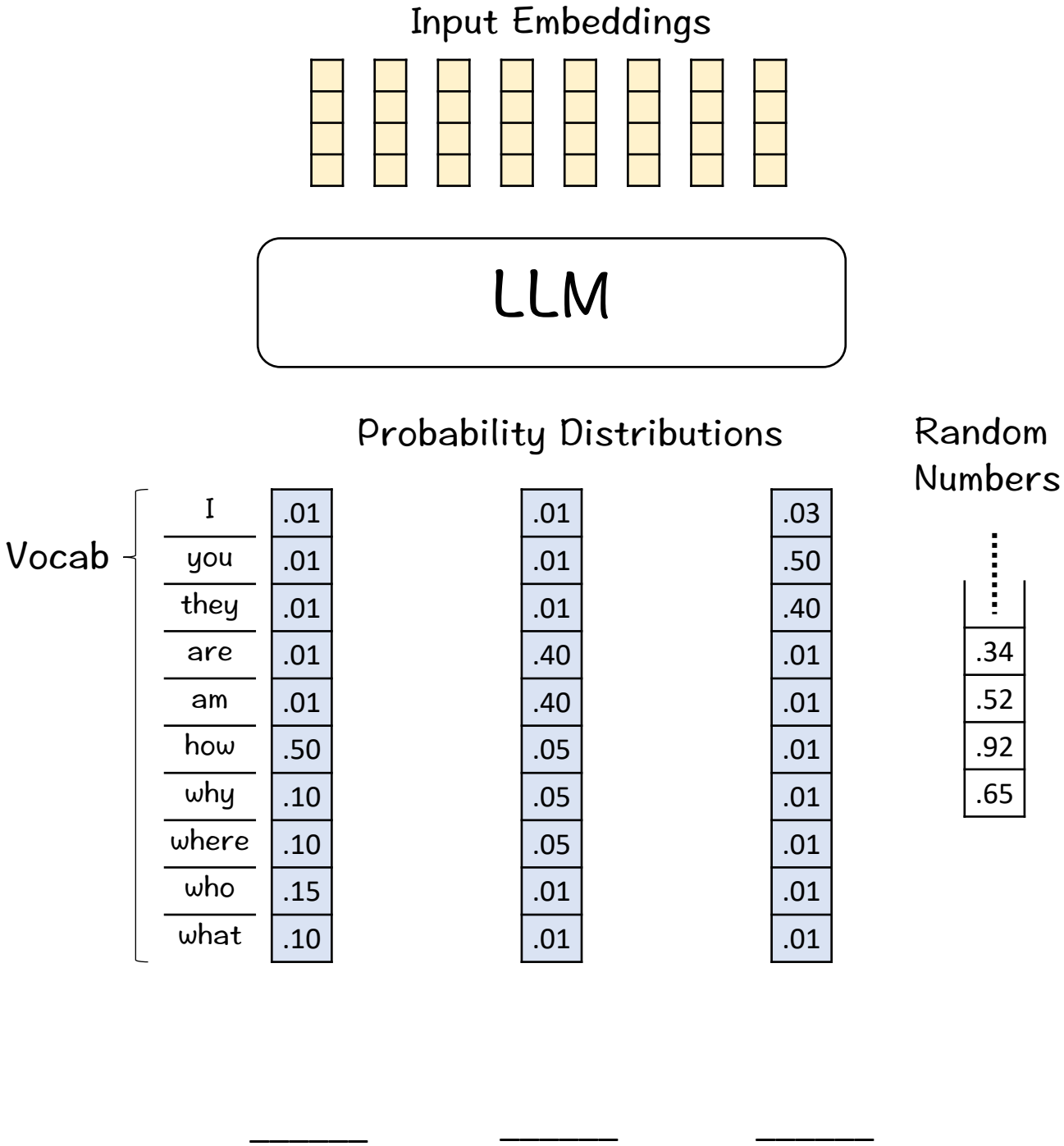
Q/A

Beginner's Guide to RAG - AI by Hand ✍️

University of Colorado Boulder

Download: https://by-hand.ai/rag

How does an LLM sample a sentence?



How does an LLM sample a sentence?