DATA 442: Neural Networks & Deep Learning

Dan Runfola – danr@wm.edu

icss.wm.edu/data442/

Summary

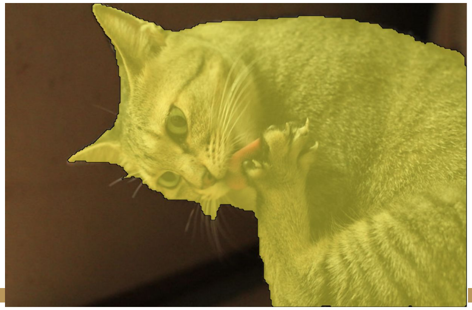
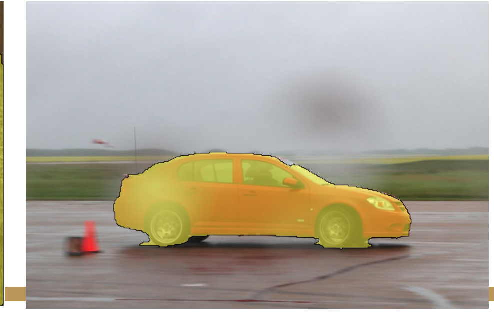
Total Loss=

$$\{(x_i, y_i)\}_{i=1}^{N=3}$$

$$\frac{1}{N}\sum_i^N Loss_i(f(x_i, W), y_i) + \lambda R(W)$$

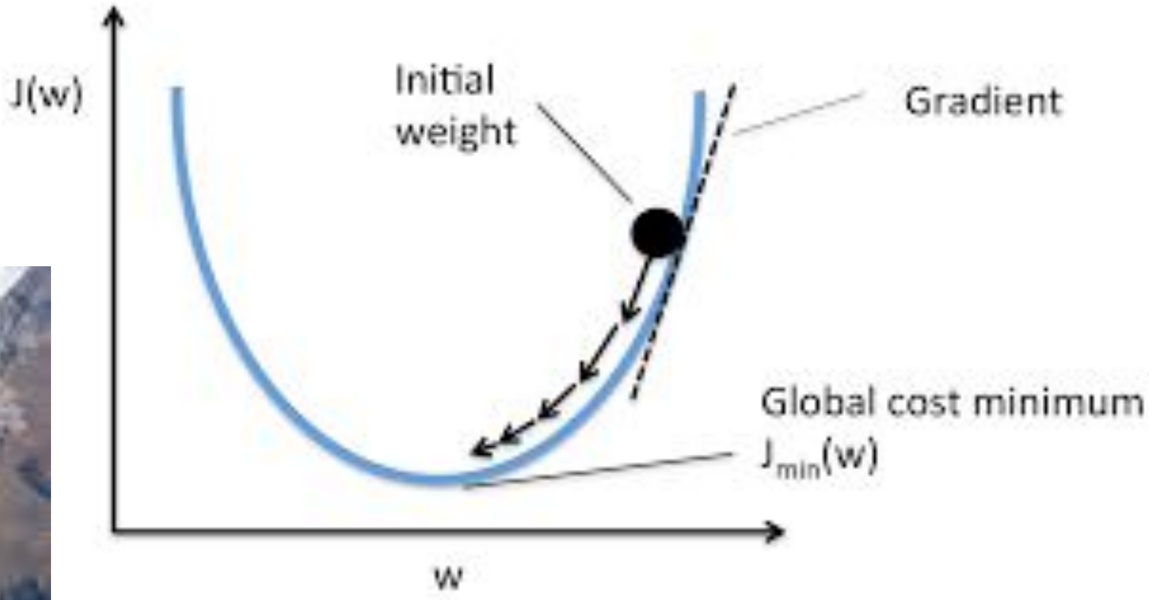
```
def predict(image, W):
    return(W*image)
```

Cat	3.2	1.3	2.2
Car	5.1	4.9	2.5
Frog	-1.7	2.0	-3.1


$$L_i = -log(\frac{e^s_k}{\sum_{j=1}^J e^s_j})$$

Optimization



Analytic Gradient

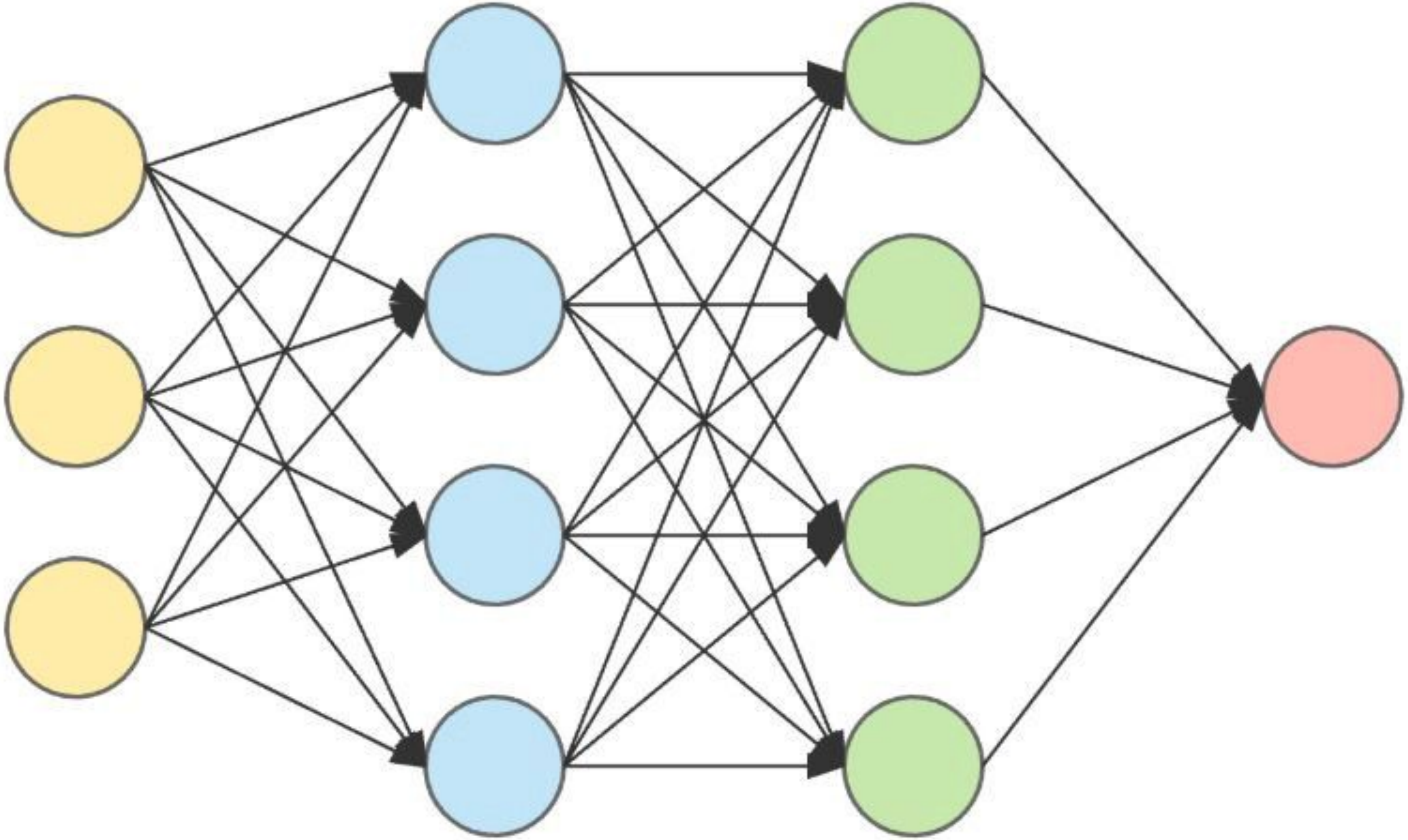
W = [0.34, -1.11, 0.78, 0.12 … 0.3, 0.77]

$$dw = f(X, W)$$

$$\nabla f(X, W) = [\ldots]$$

Gradient

dW: [-2.5, 0.6, 4.3, 0.5 … 0, 0.3]



$$f(X, W)$$

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

$$f(X, W)$$

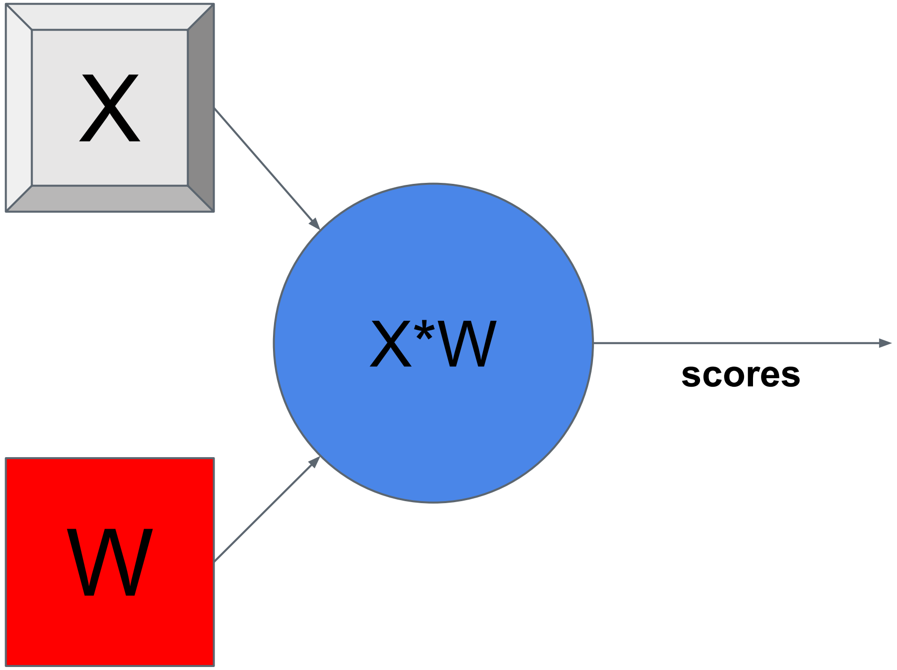
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

X

W

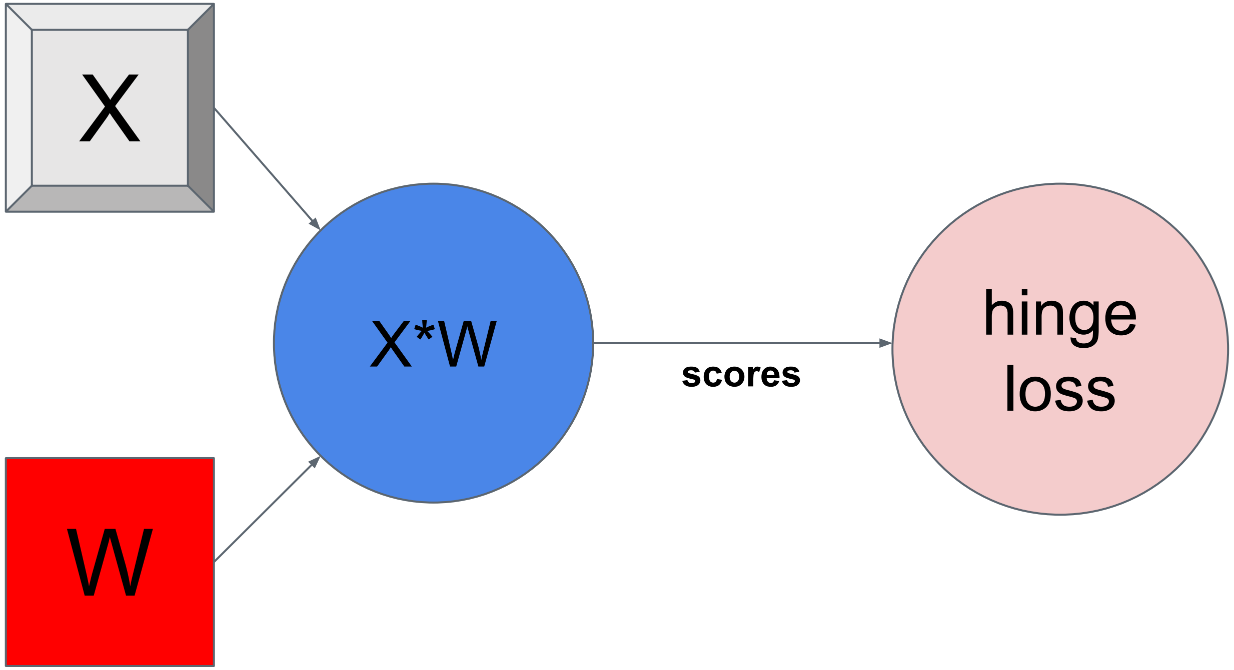
$$f(X, W)$$

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



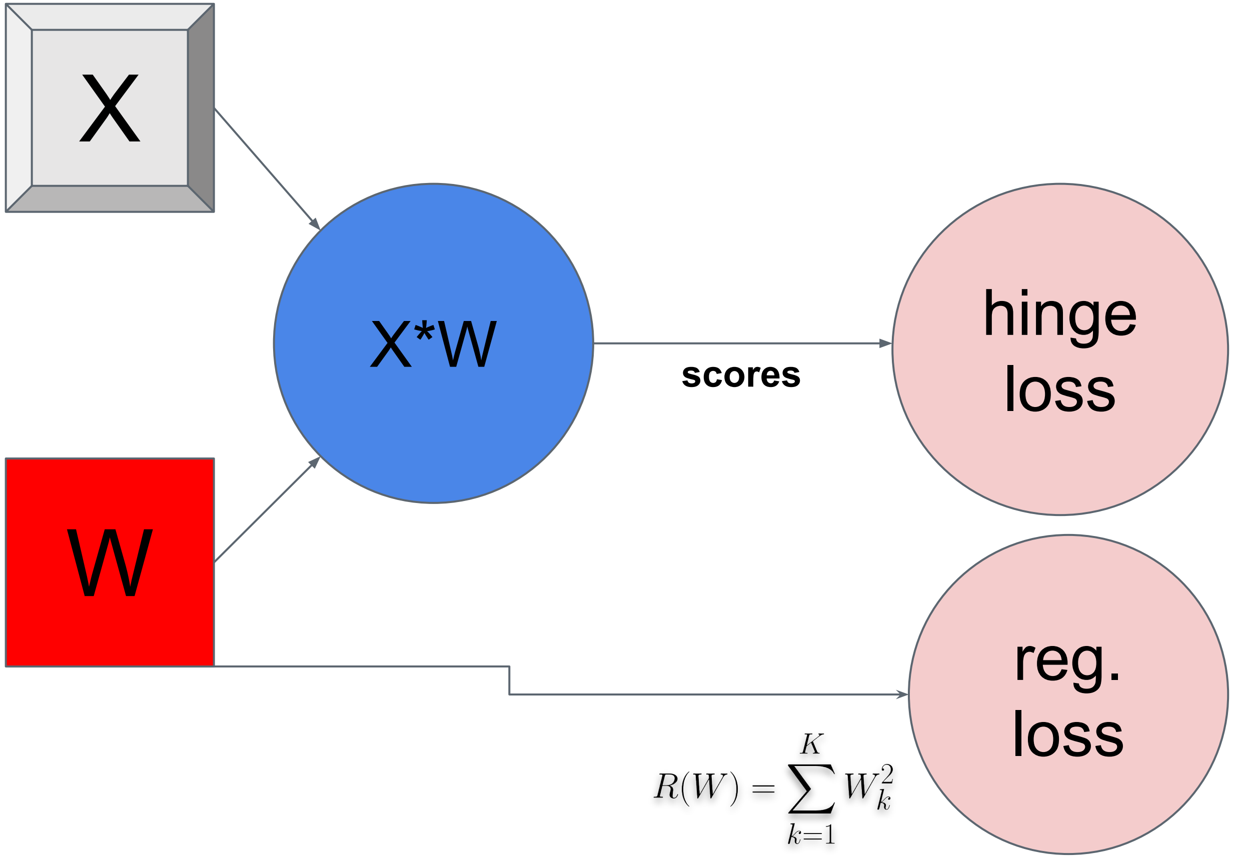
$$f(X, W)$$

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



$$f(X, W)$$

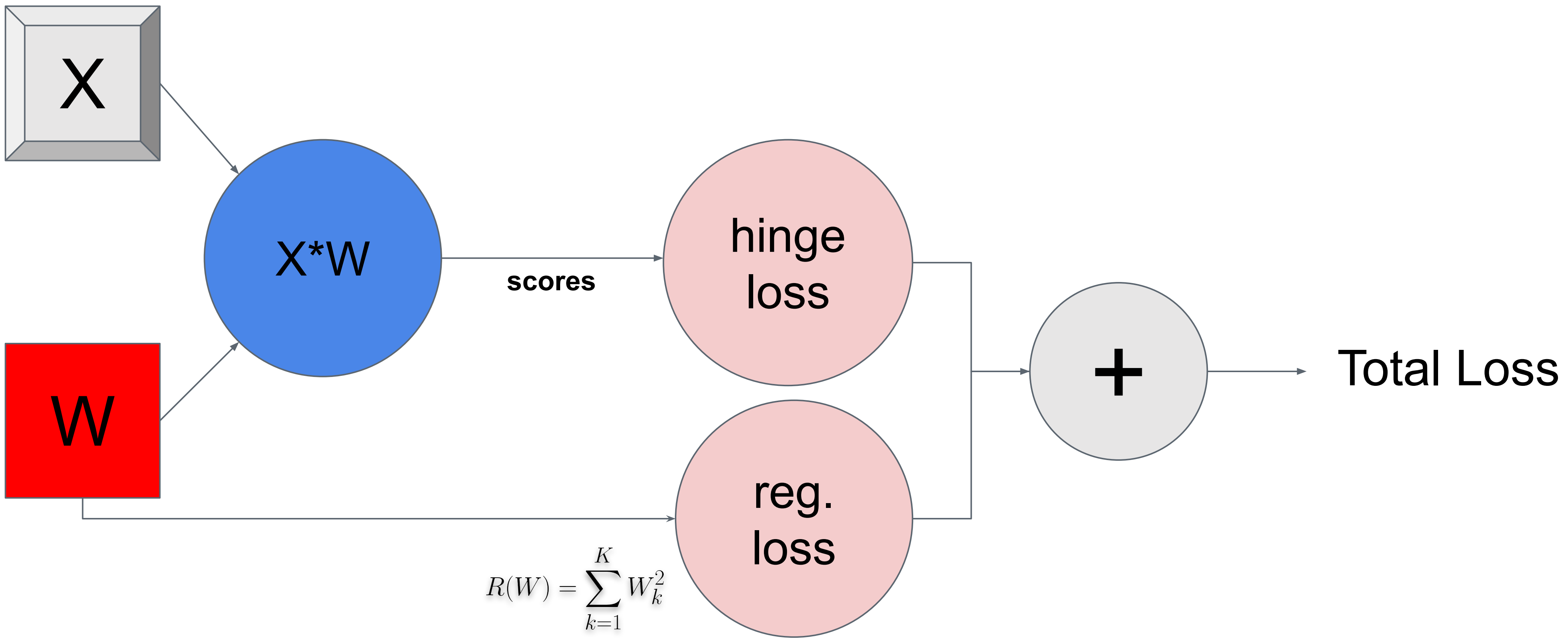
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



$$R(W) = \sum_{k=1}^{K} W_k^2$$

$$f(X, W)$$

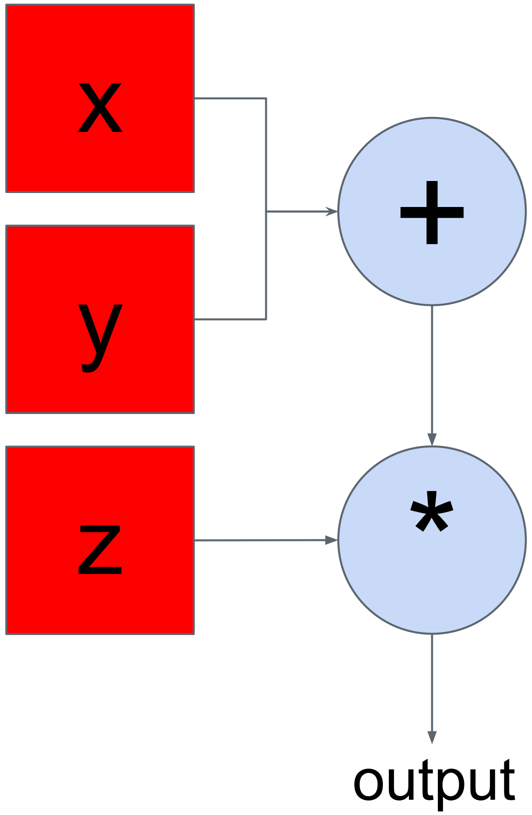
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



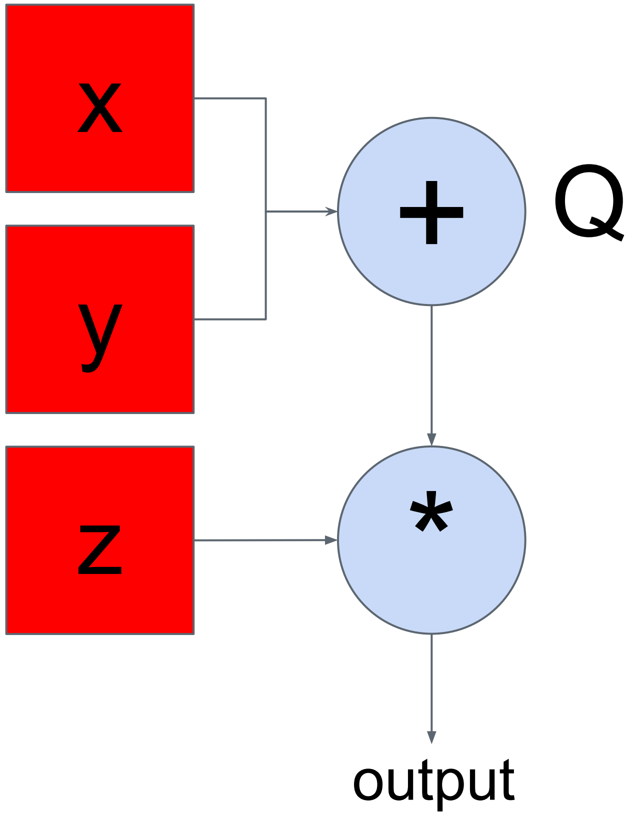
Backpropogation

$$f(x, y, z) = (x + y) * z$$

$$f(x, y, z) = (x + y) * z$$

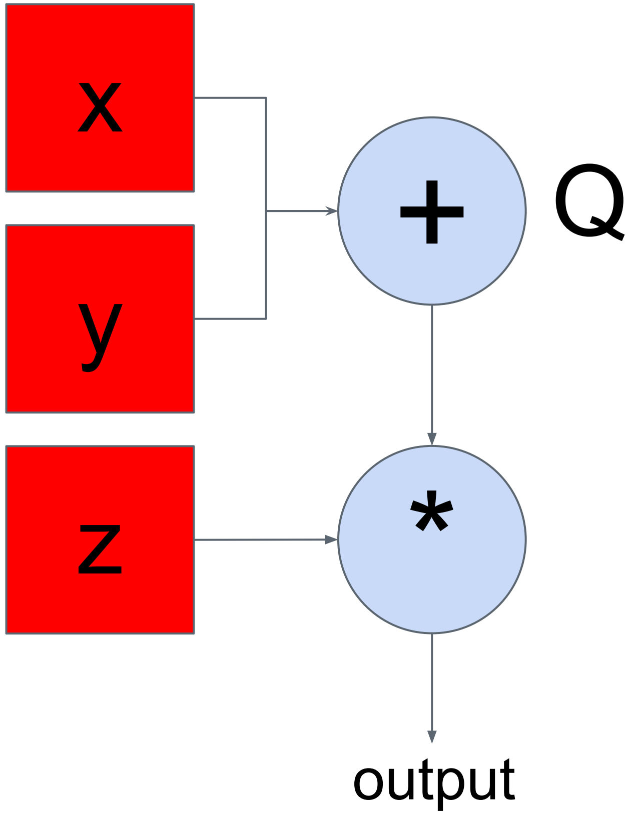


$$f(x, y, z) = (x + y) * z$$



$$Q = x + y$$

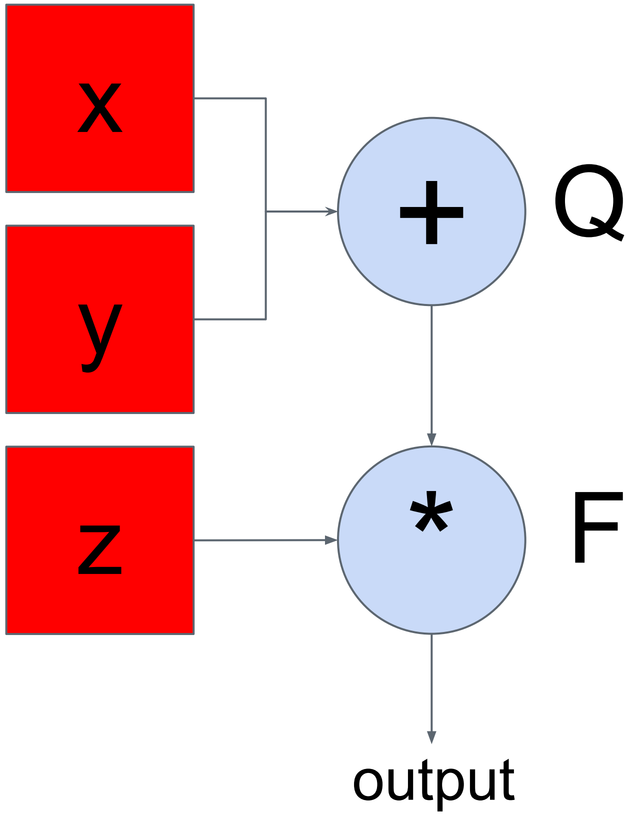
$$f(x,y,z) = (x+y)*z$$



$$Q = x + y$$

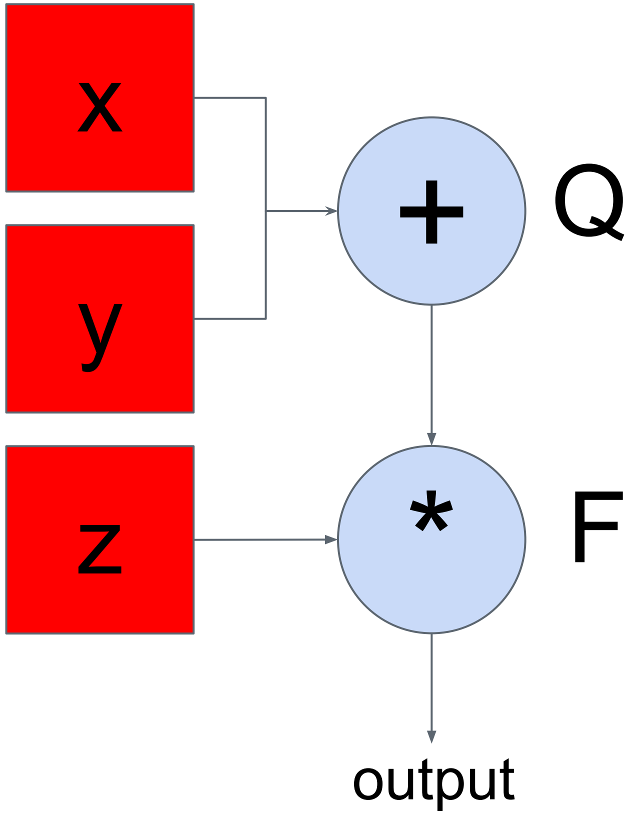
$$\frac{\partial q}{\partial x} = 1 \quad \frac{\partial q}{\partial y} = 1$$

$$f(x, y, z) = (x + y) * z$$



$$F = qz$$

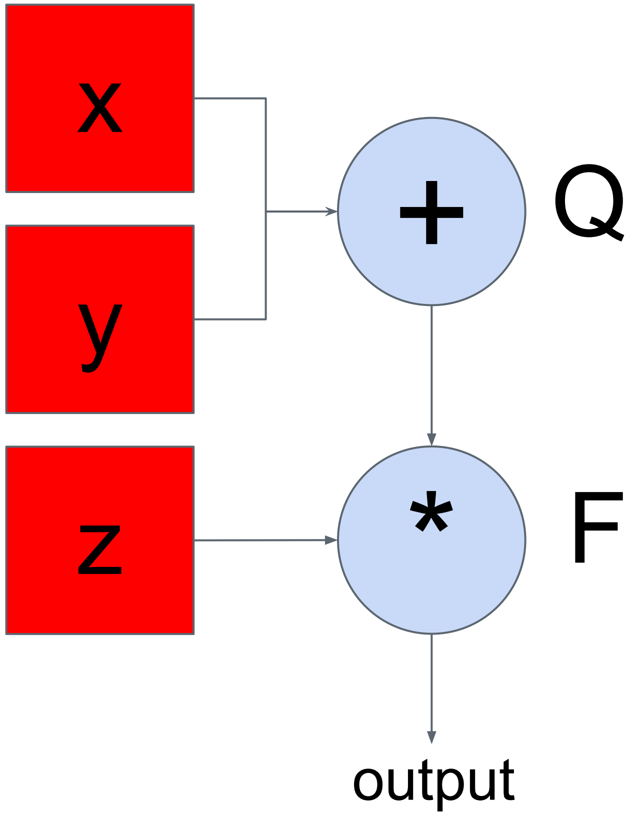
$$f(x, y, z) = (x + y) * z$$



$$F = qz$$

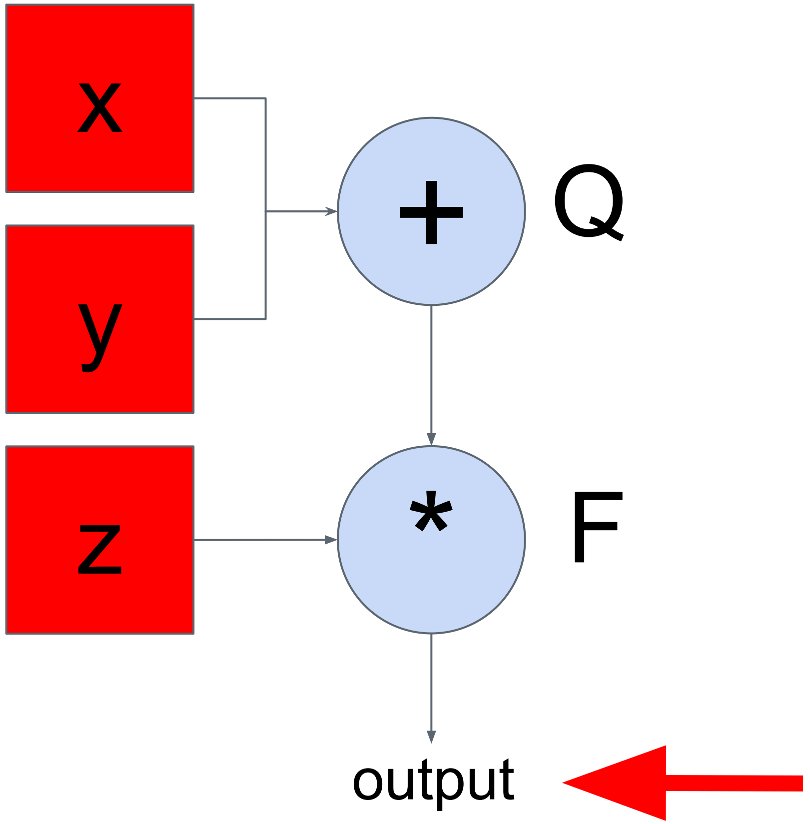
$$\frac{\partial f}{\partial Q} = z \quad \frac{\partial f}{\partial z} = Q$$

$$f(x, y, z) = (x + y) * z$$



$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$

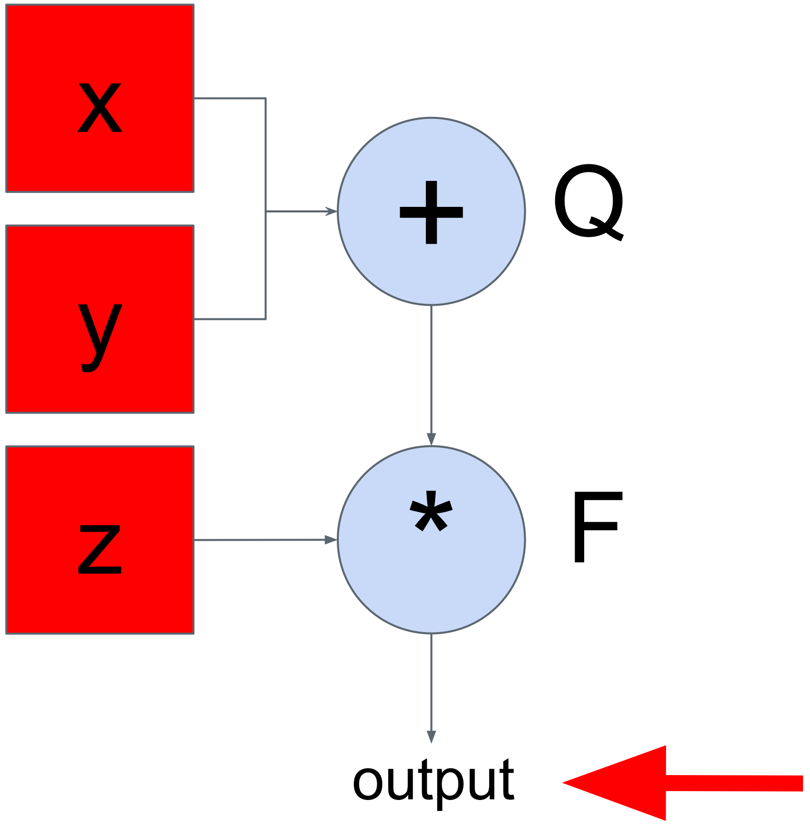


$$\frac{\partial F}{\partial F}$$

The Goal

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$

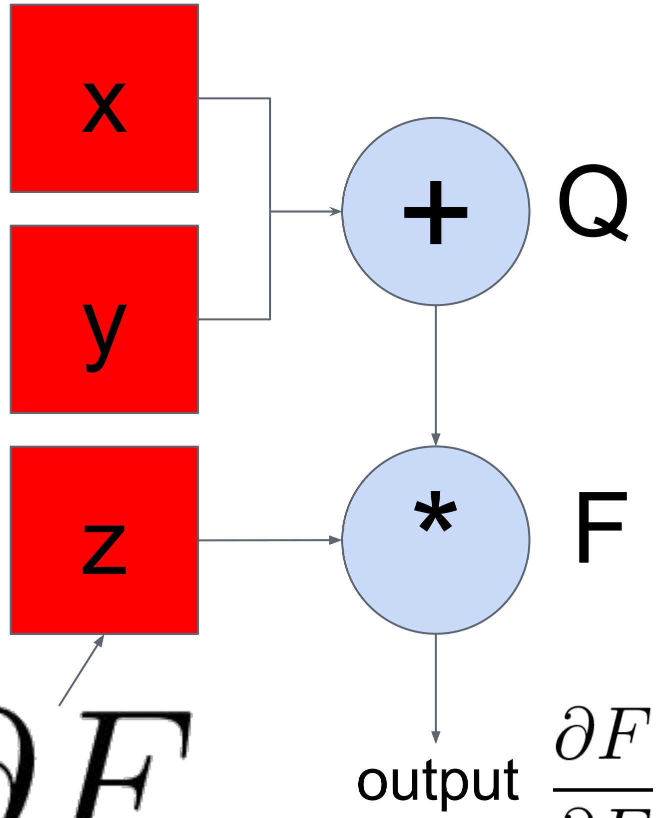


$$\frac{\partial F}{\partial F} = 1$$

The Goal

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$



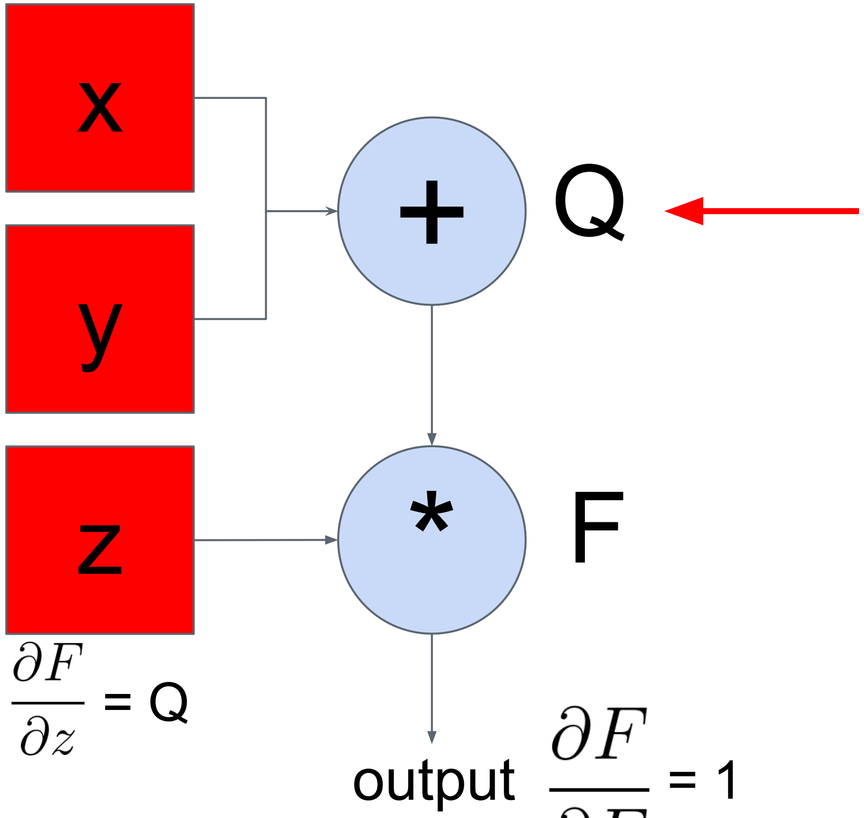
$$\frac{\partial F}{\partial F} = 1$$

$$\frac{\partial F}{\partial z} = Q$$

The Goal

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x,y,z) = (x+y) * z$$



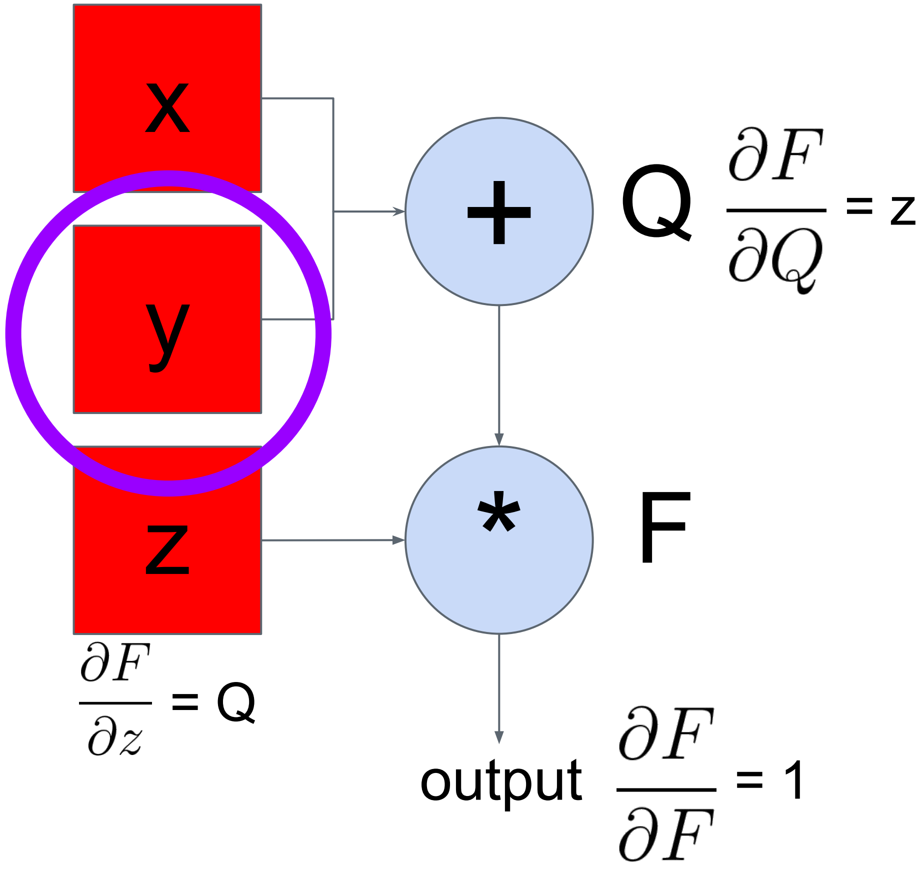
$$\frac{\partial F}{\partial z} = Q \qquad \frac{\partial F}{\partial F} = 1$$

$$\frac{\partial F}{\partial Q} = z$$

<u>The Goal</u>

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$

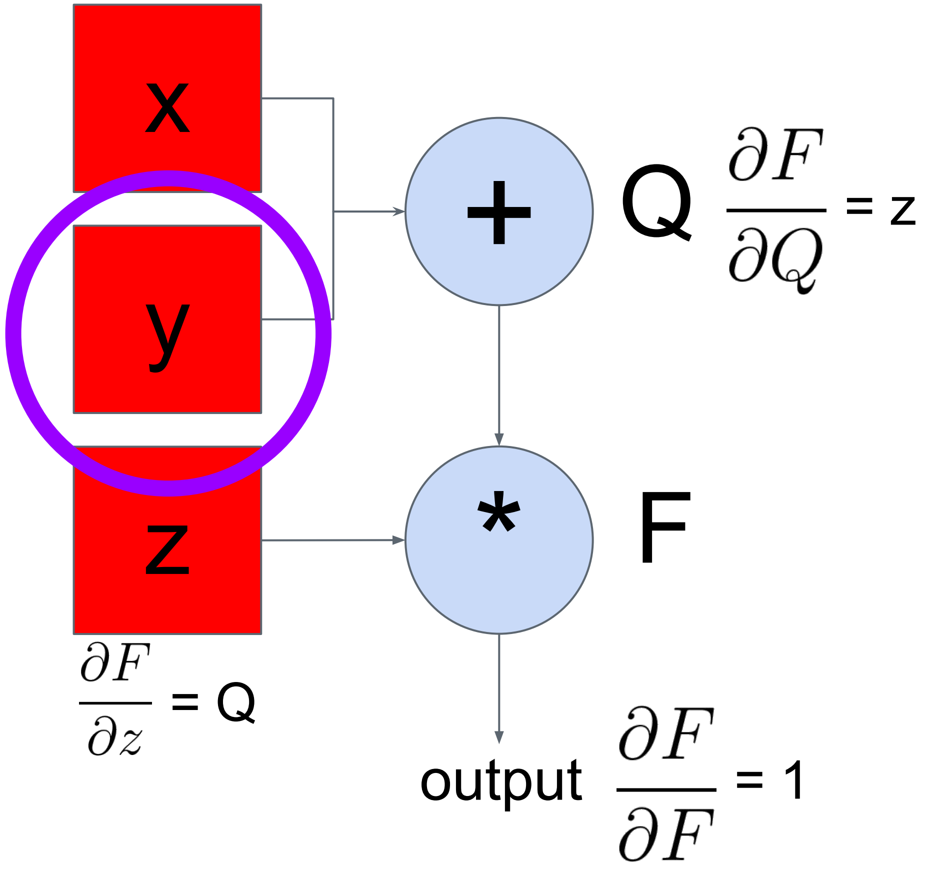


$$\frac{\partial F}{\partial y}$$

<u>The Goal</u>

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$

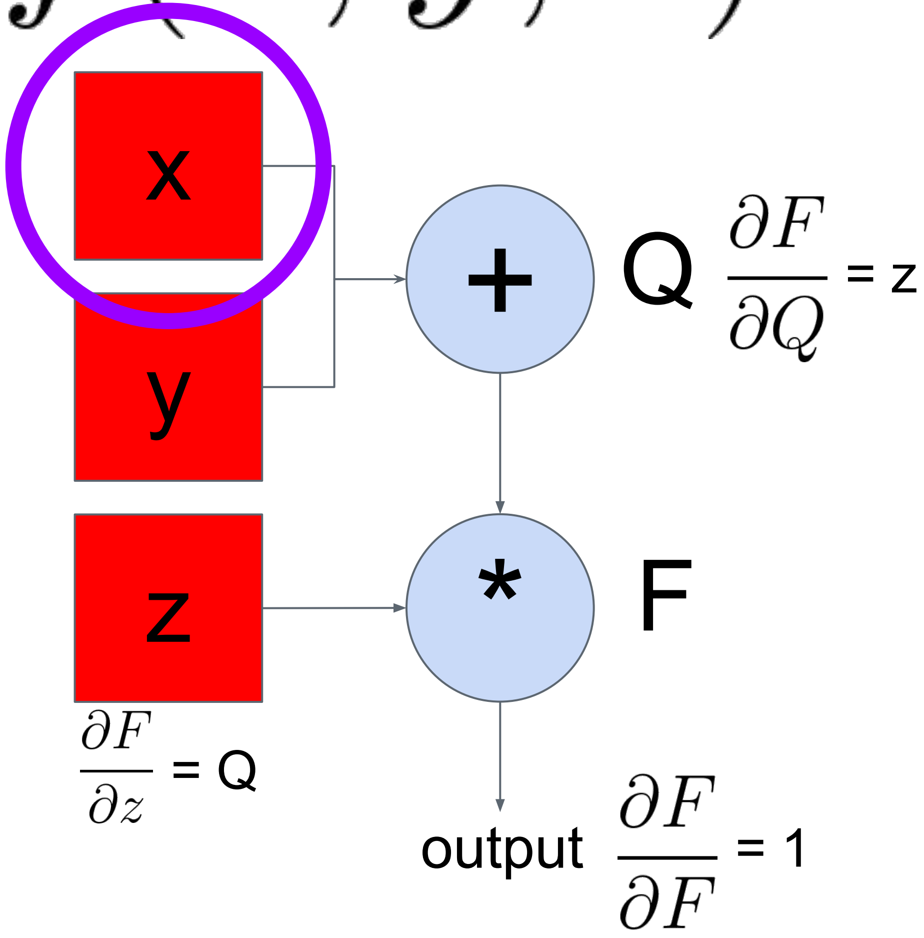


$$\frac{\partial F}{\partial y} = \frac{\partial F}{\partial q}\frac{\partial Q}{\partial y}$$

<u>The Goal</u>

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

$$f(x, y, z) = (x + y) * z$$

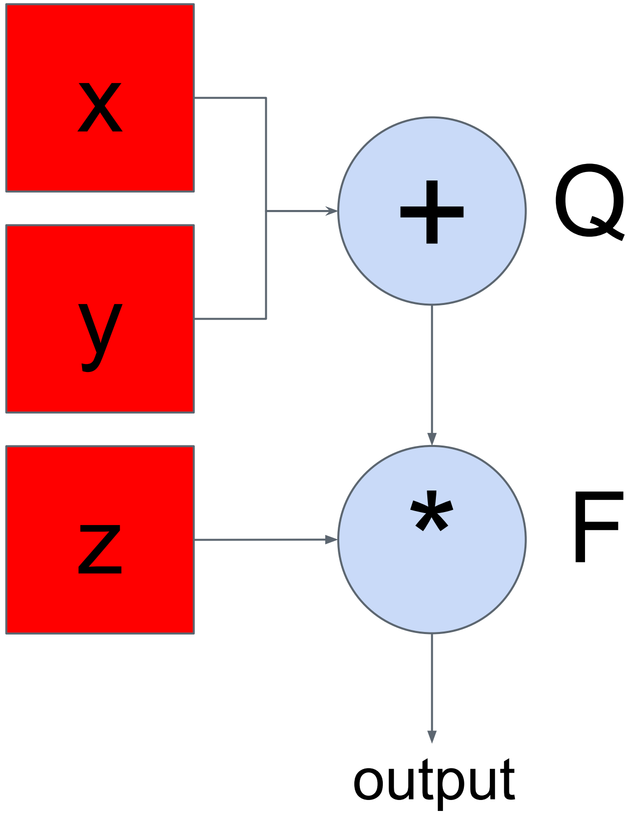


$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial q}\frac{\partial Q}{\partial x}$$

The Goal

$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

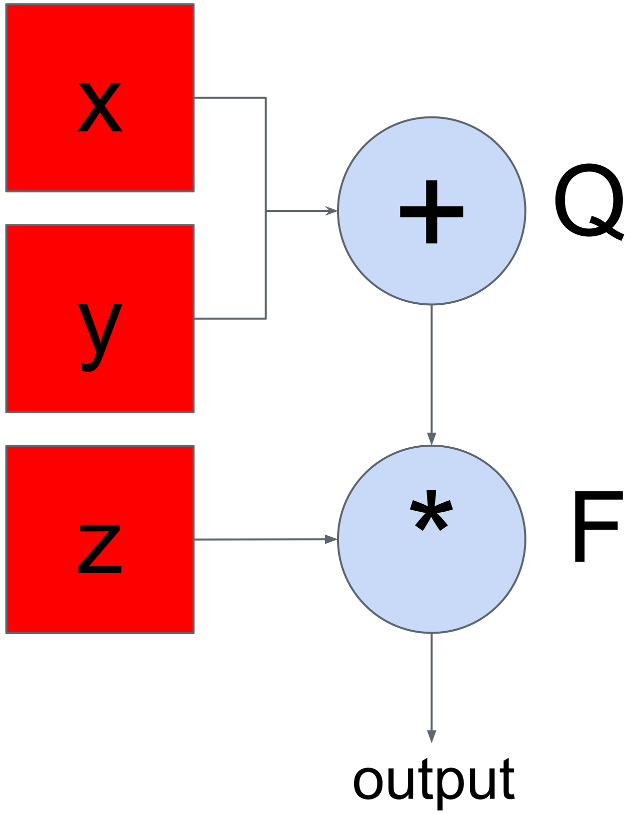
$$f(x,y,z) = (x+y) * z$$



$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial q}\frac{\partial Q}{\partial x}$$

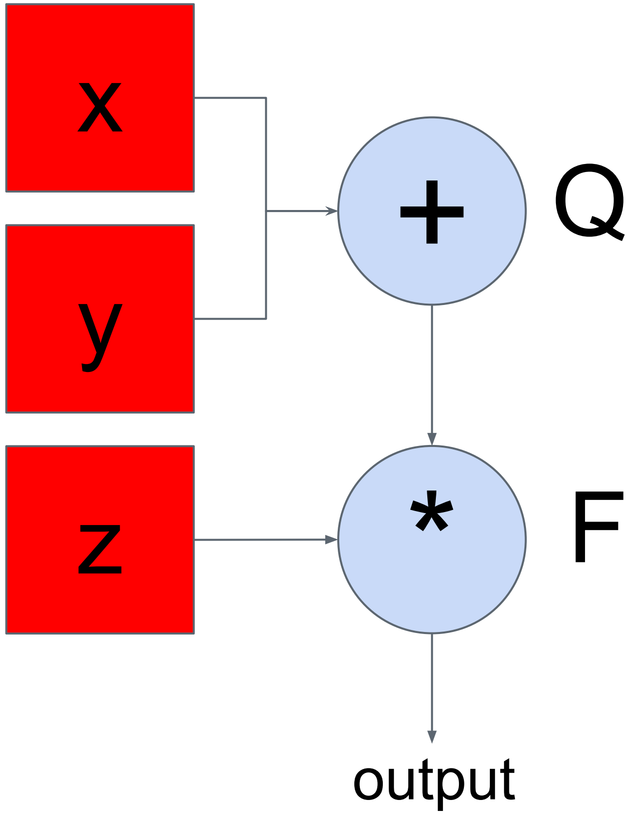
$$\frac{\partial F}{\partial Q} = z$$

$$f(x, y, z) = (x + y) * z$$



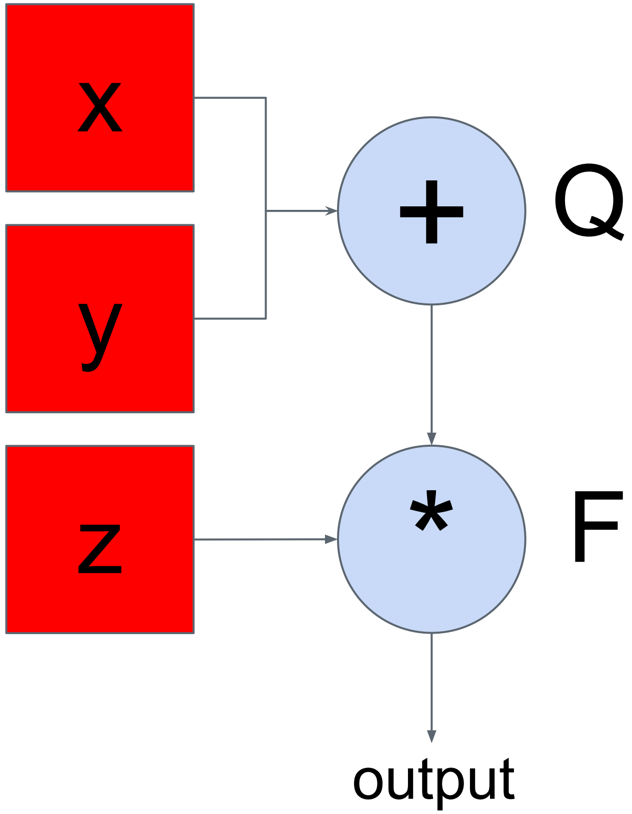
$$\frac{\partial F}{\partial x} = z \frac{\partial Q}{\partial x}$$

$$f(x, y, z) = (x + y) * z$$



$$\frac{\partial F}{\partial x} = z * 1$$

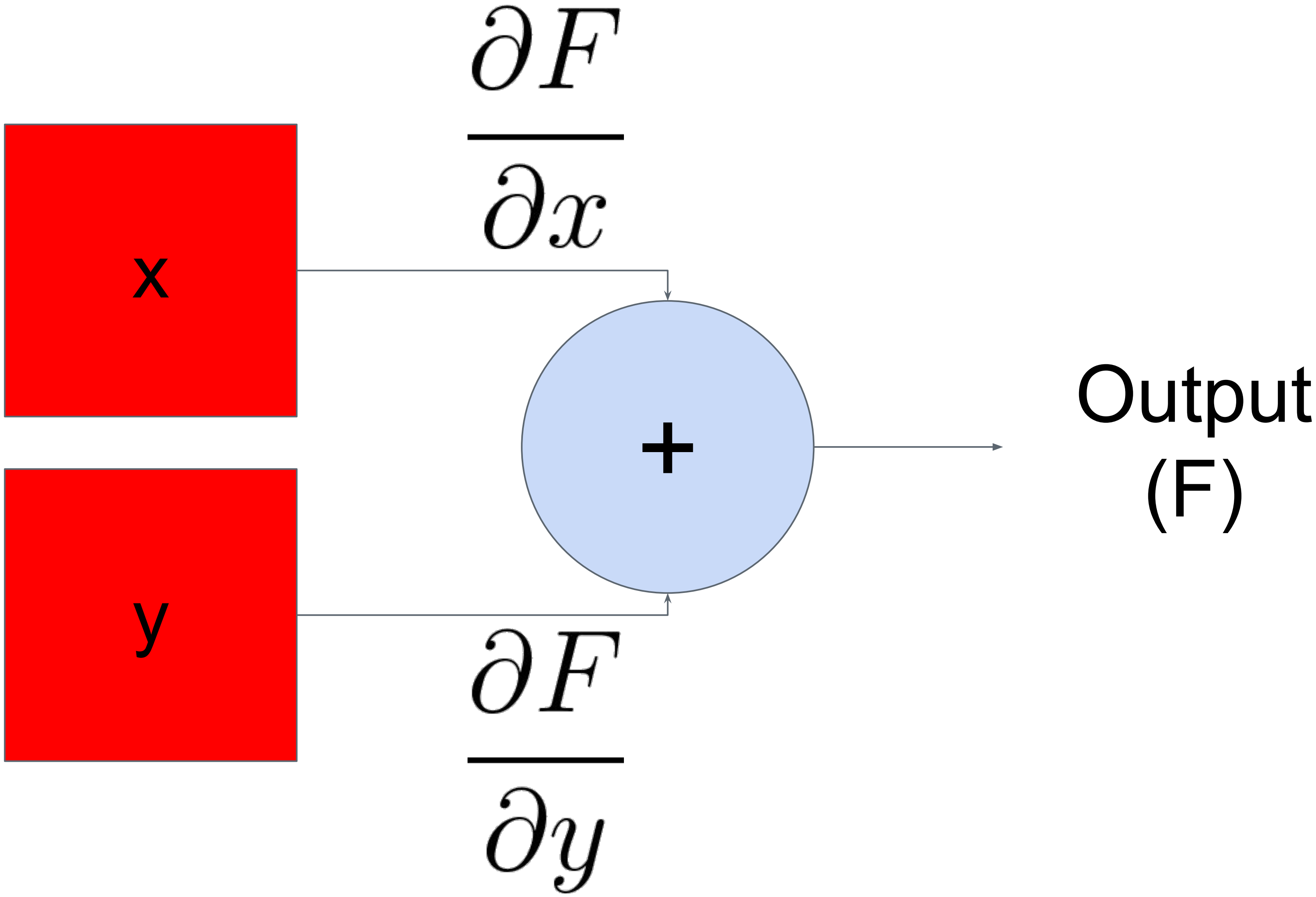
$$f(x,y,z) = (x+y)*z$$



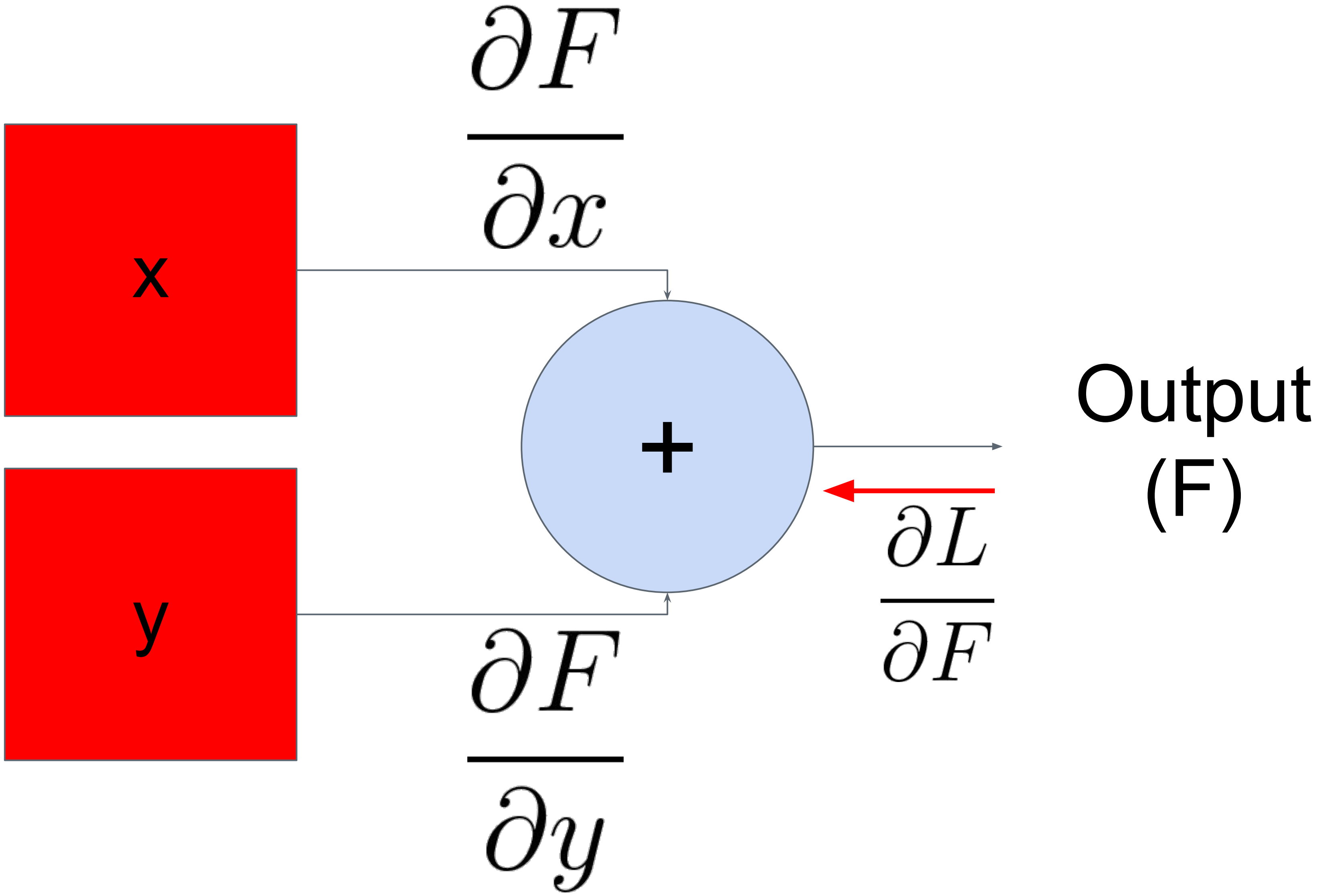
The Goal

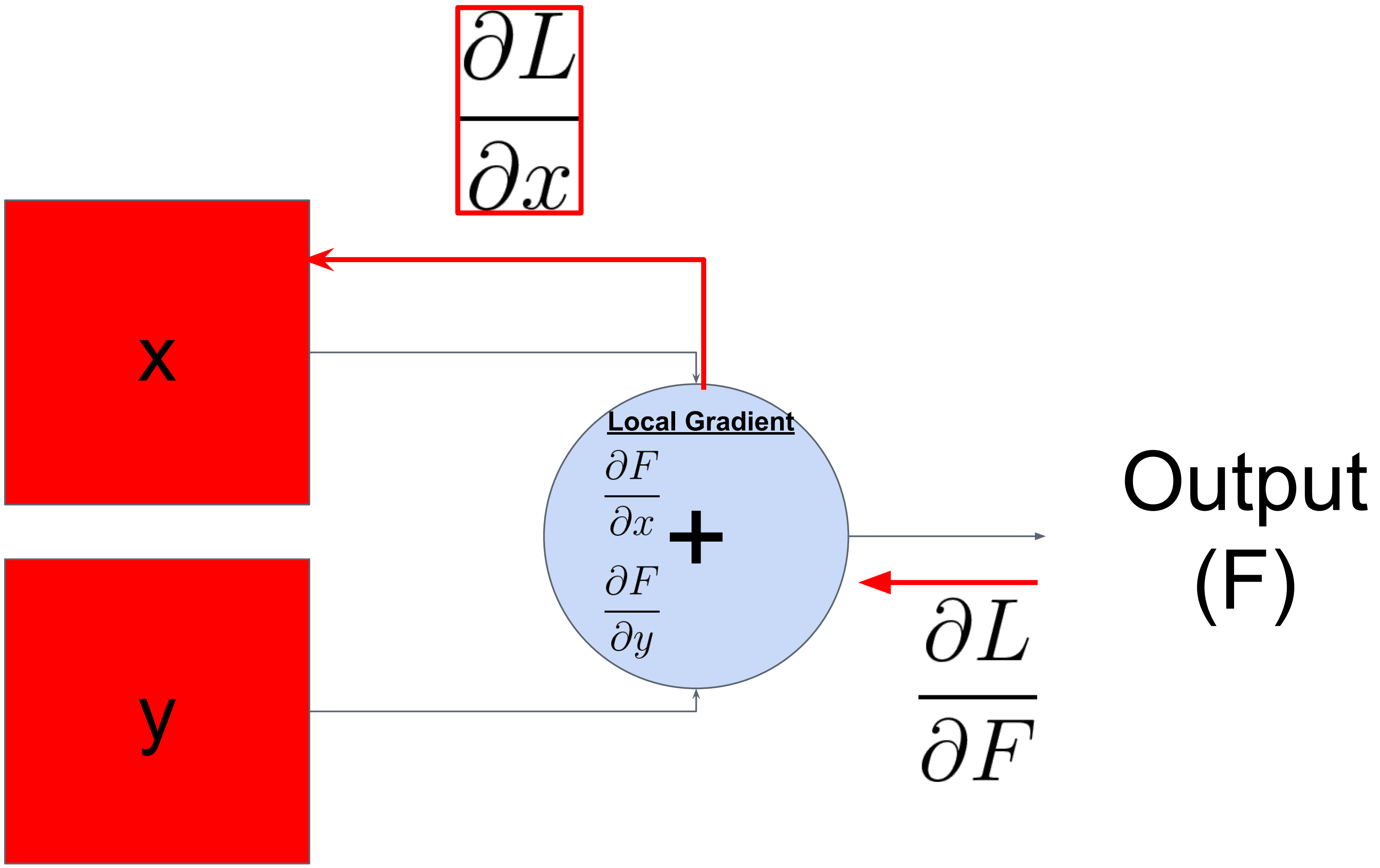
$$\frac{\partial F}{\partial x} \quad \frac{\partial F}{\partial y} \quad \frac{\partial F}{\partial z}$$

Local Gradient

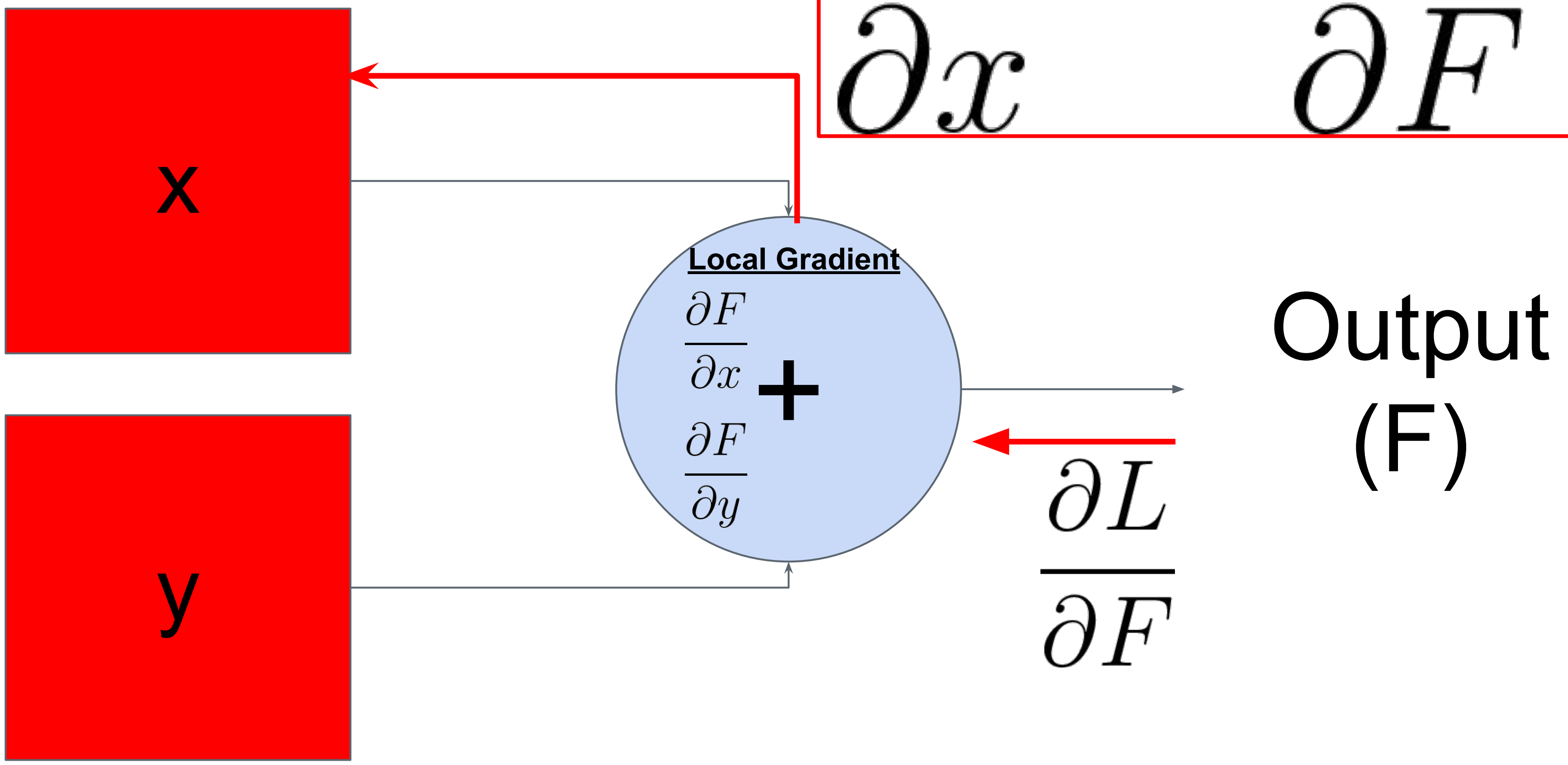


Local Gradient

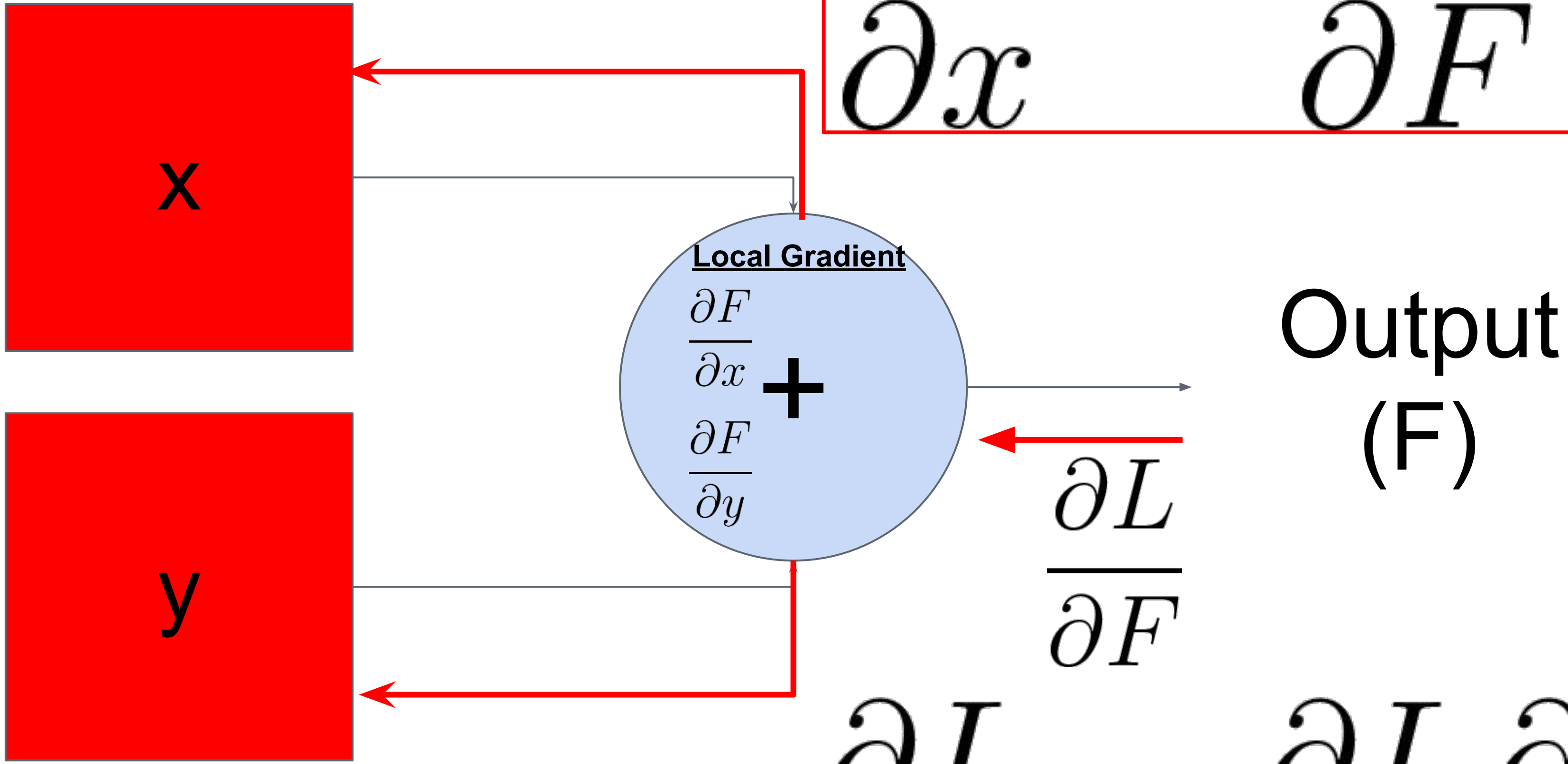



$\frac{\partial L}{\partial x}$
x
Local Gradient
$\frac{\partial F}{\partial x}$
+
$\frac{\partial F}{\partial y}$
y
Output
(F)
$\frac{\partial L}{\partial F}$

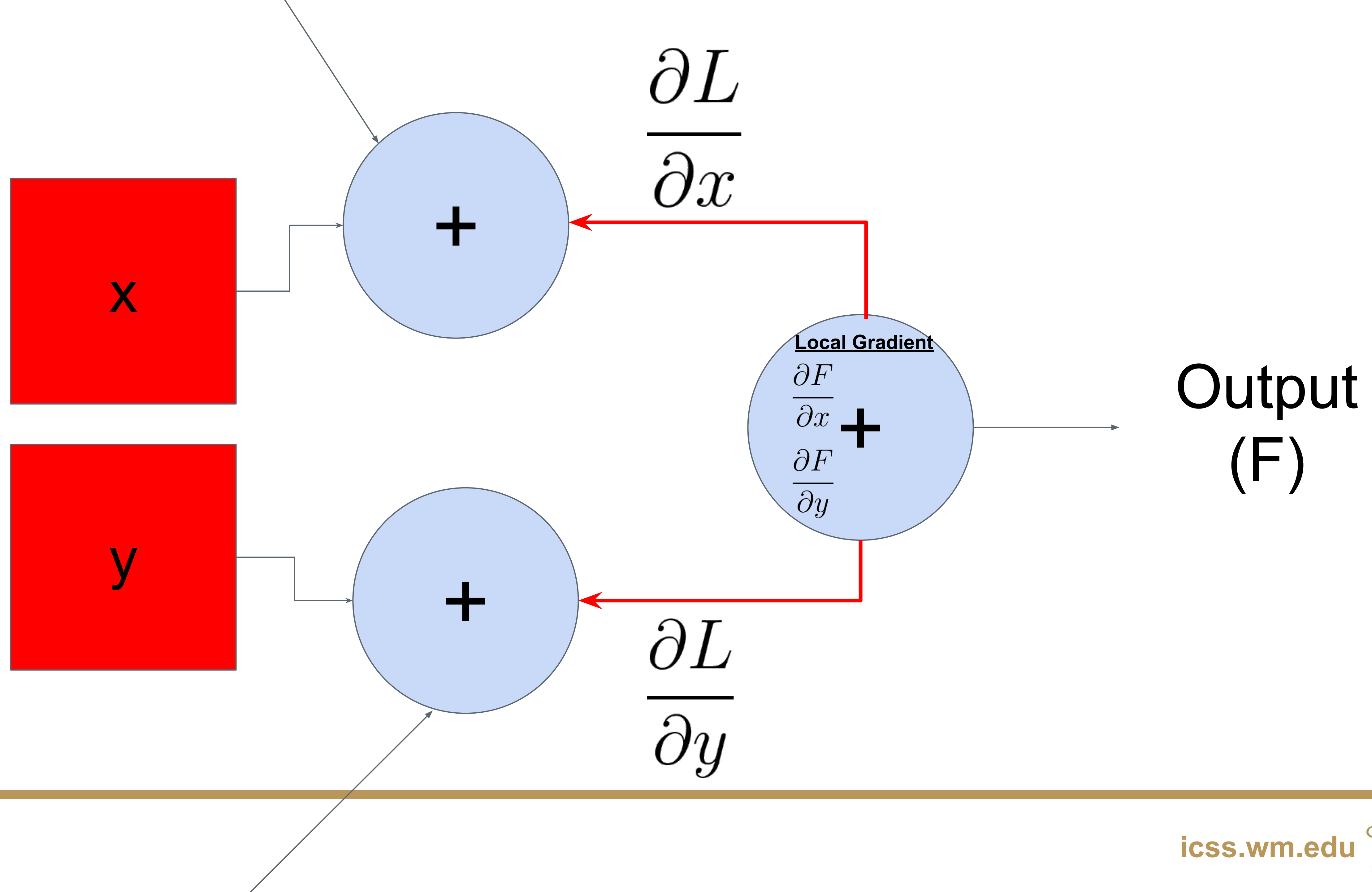
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial F}\frac{\partial F}{\partial x}$$

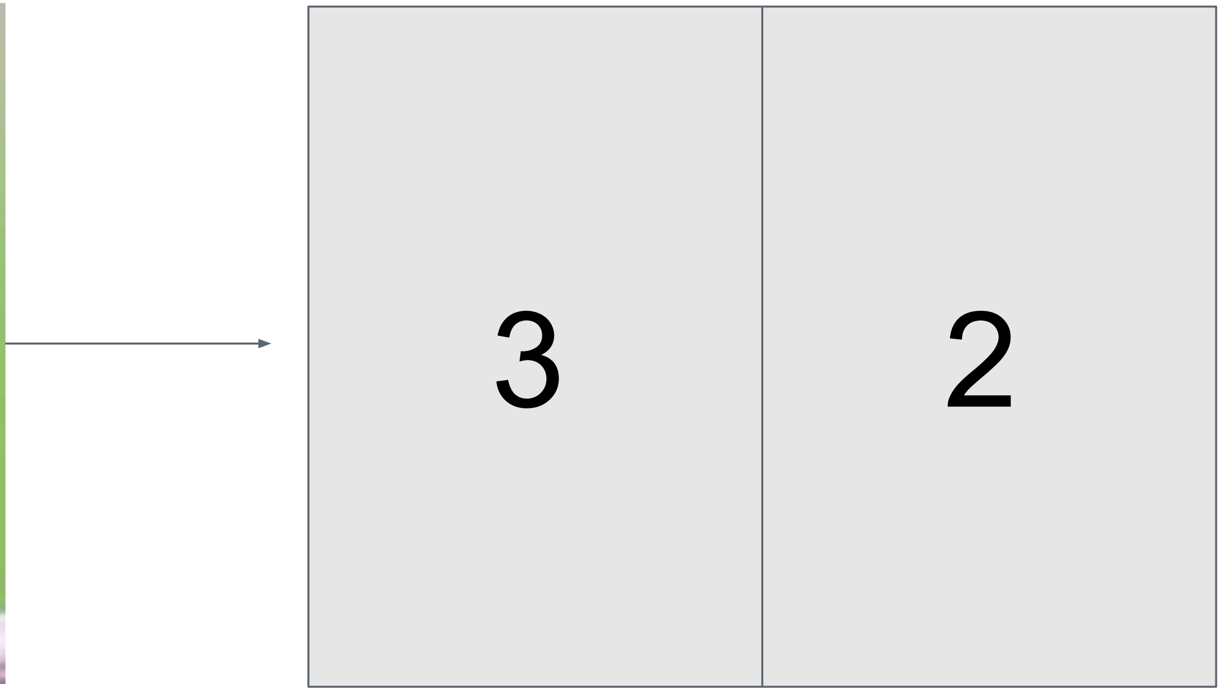


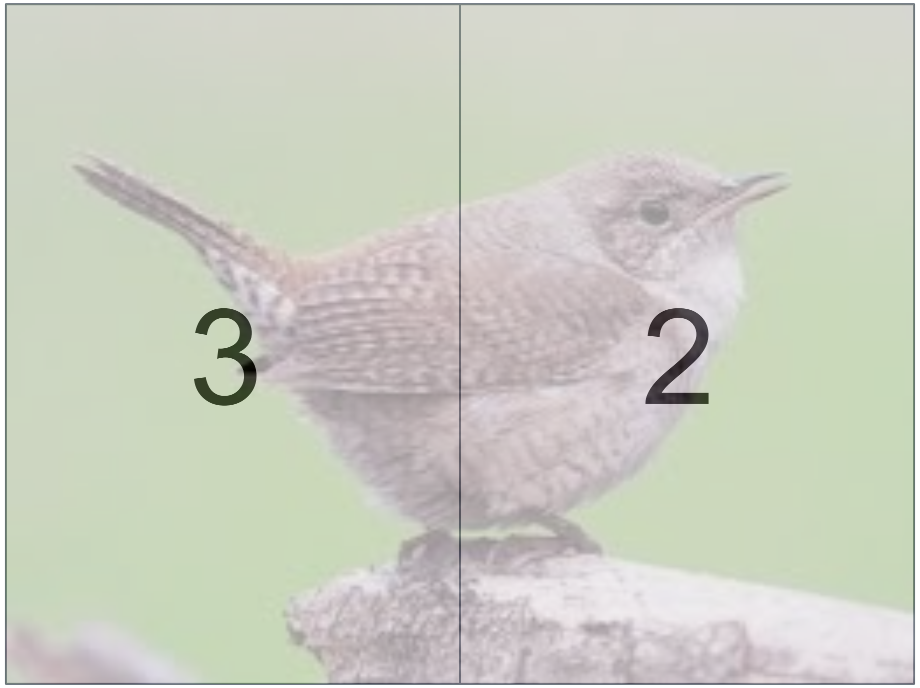
$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial F}\frac{\partial F}{\partial x}$$



$$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial F}\frac{\partial F}{\partial y}$$


x
y
+
+
$\frac{\partial L}{\partial x}$
$\frac{\partial L}{\partial y}$
Local Gradient
$\frac{\partial F}{\partial x}$
+
$\frac{\partial F}{\partial y}$
Output
(F)


3
2


3
2

W1_1 (Bird)
-2

W1_2 (Bird)
-1

$$f(X, W)$$

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

W2_1 (Car)
1

W2_2 (Car)
-5

W1_1 (Bird)
-2

W1_2 (Bird)
-1

Pixel_1
3

Pixel_2
2

W2_1 (Car)
1

W2_2 (Car)
-5

$$f(X, W)$$

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

W1_1 (Bird)
-2

Pixel_1
3

W2_1 (Car)
1

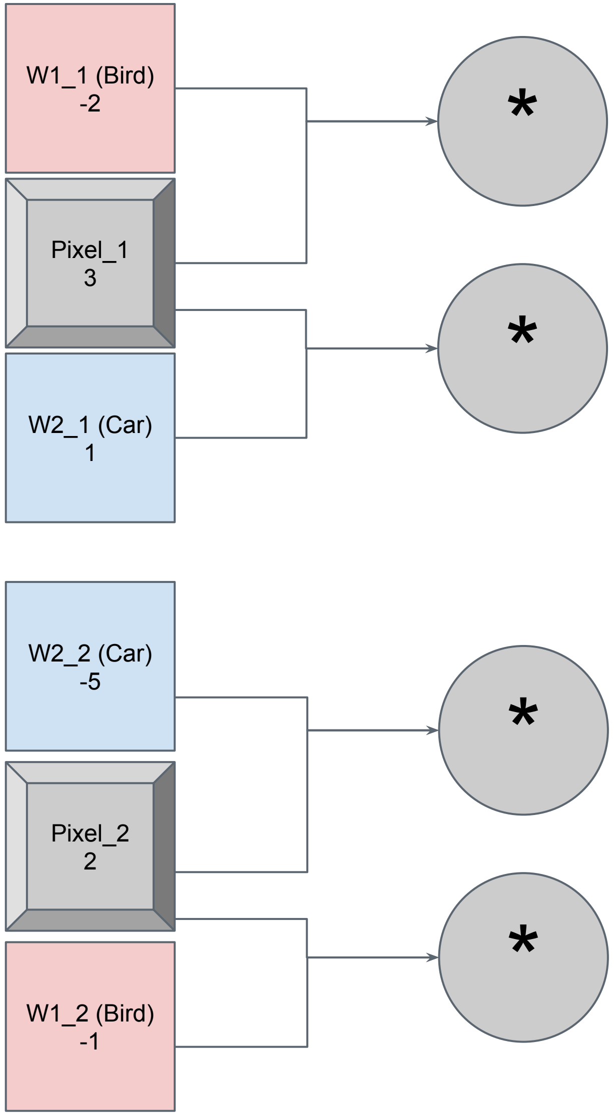
W2_2 (Car)
-5

Pixel_2
2

W1_2 (Bird)
-1

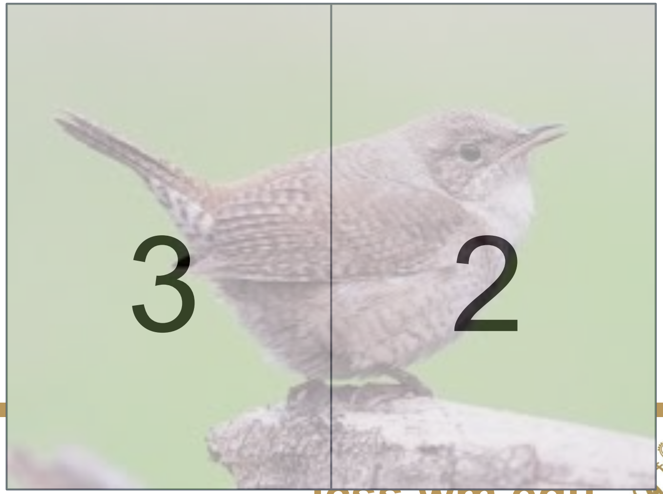
$$f(X, W)$$

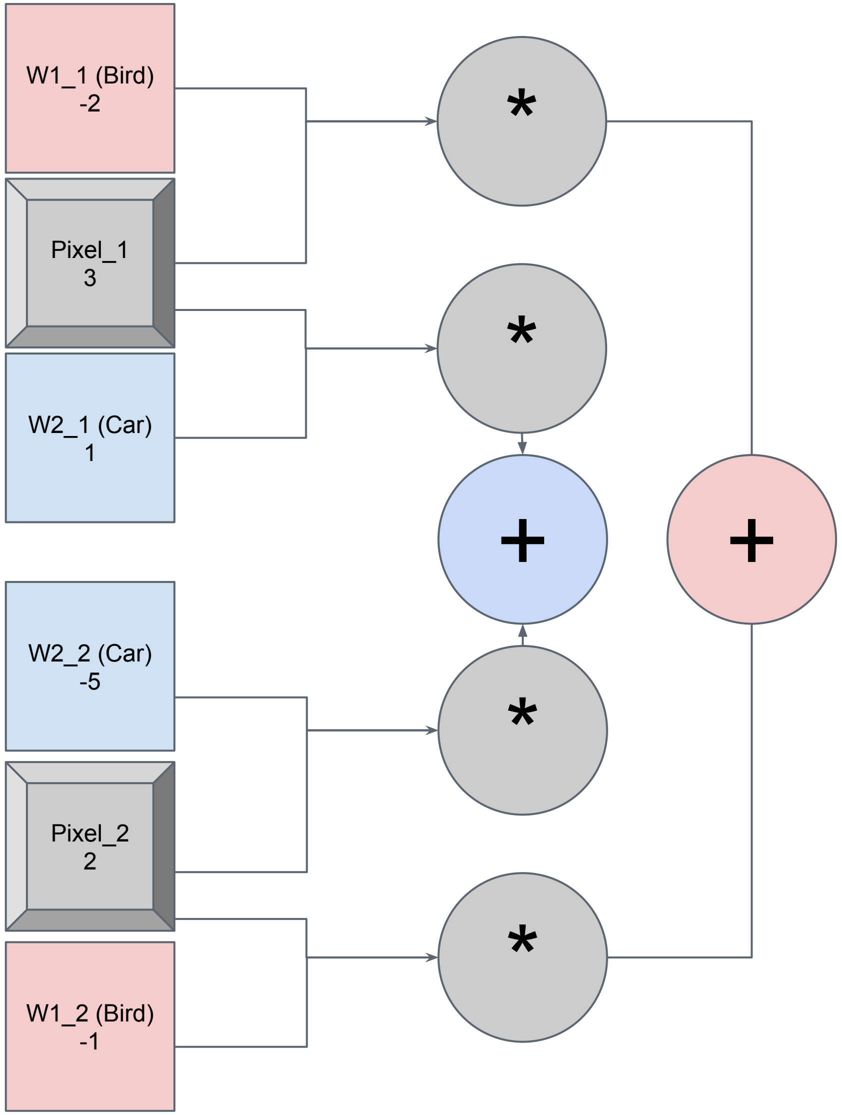
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



$$f(X, W)$$

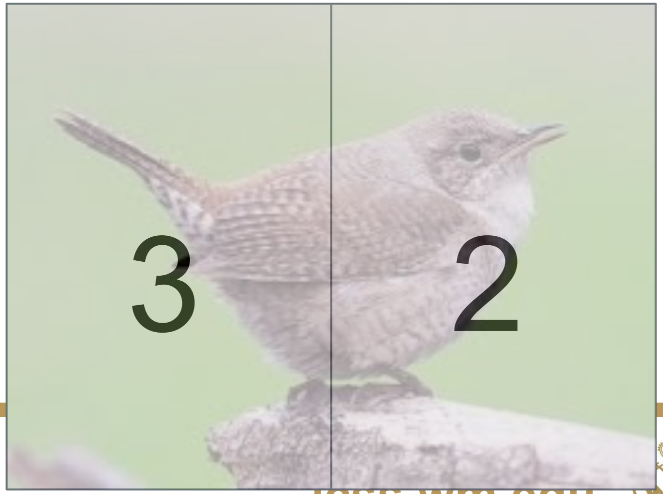
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



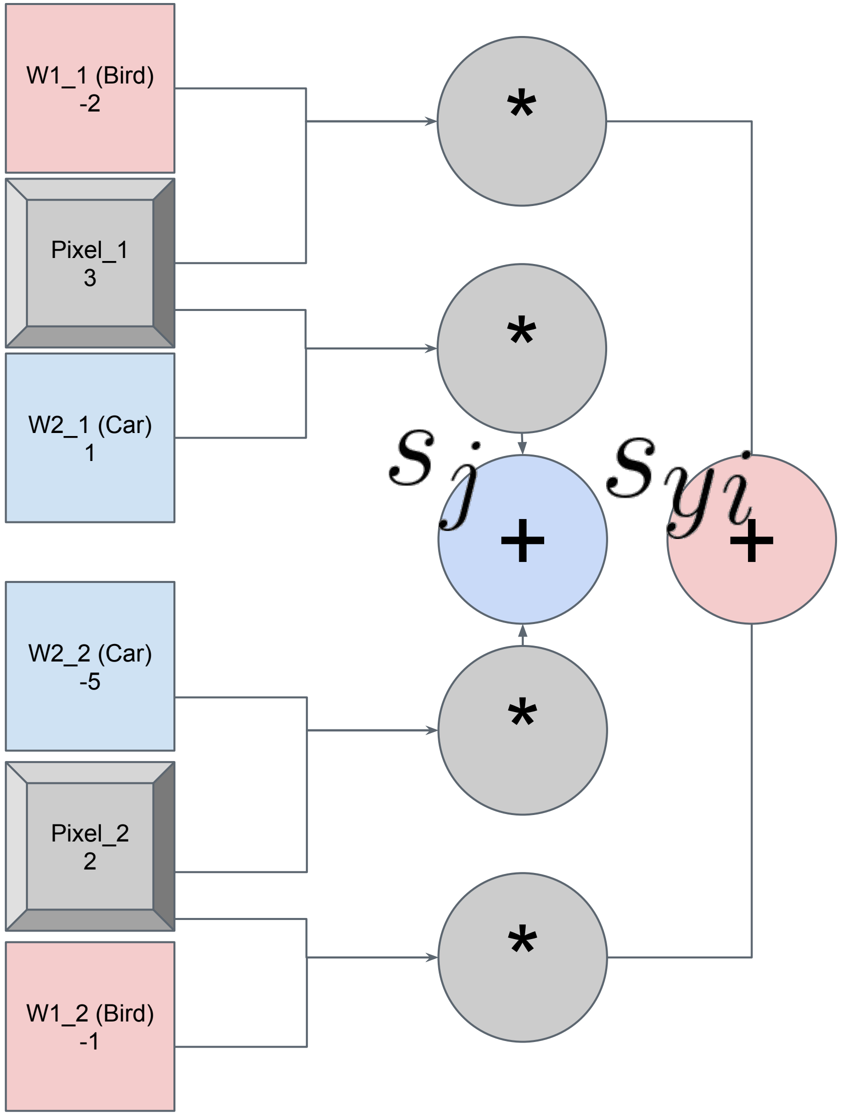


$$f(X, W)$$

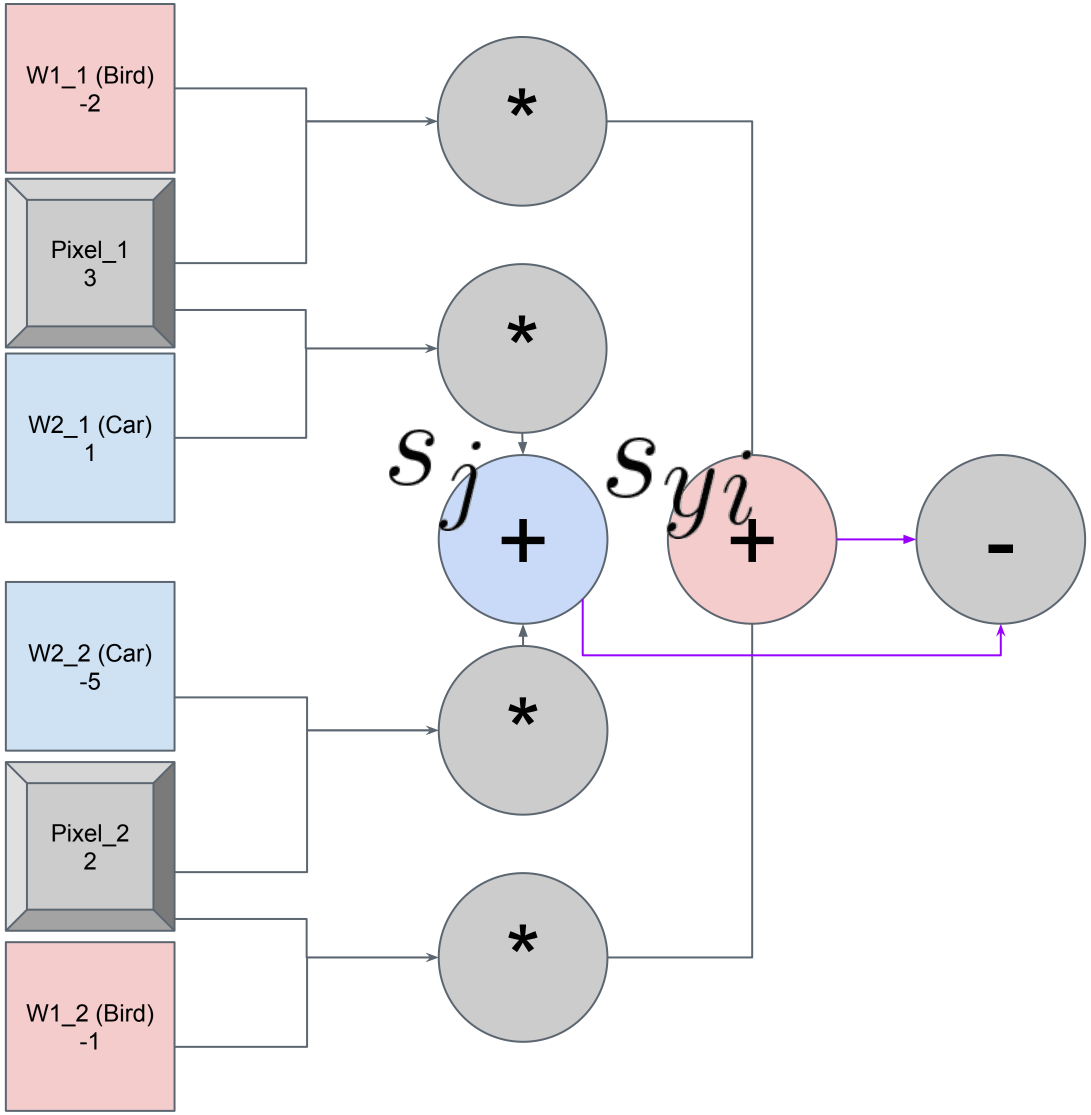
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



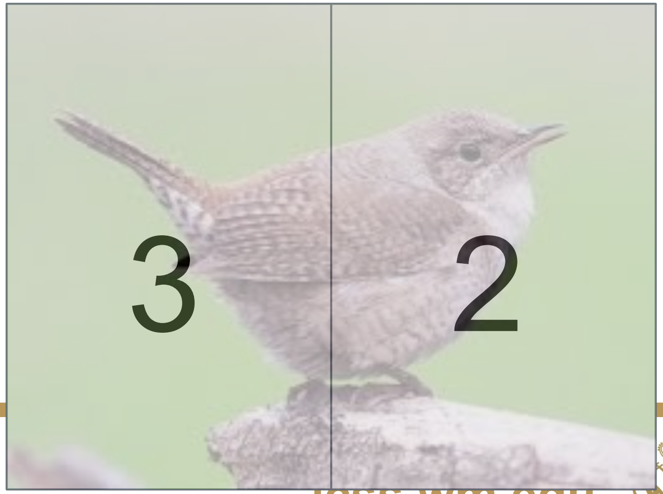
$$f(X, W)$$



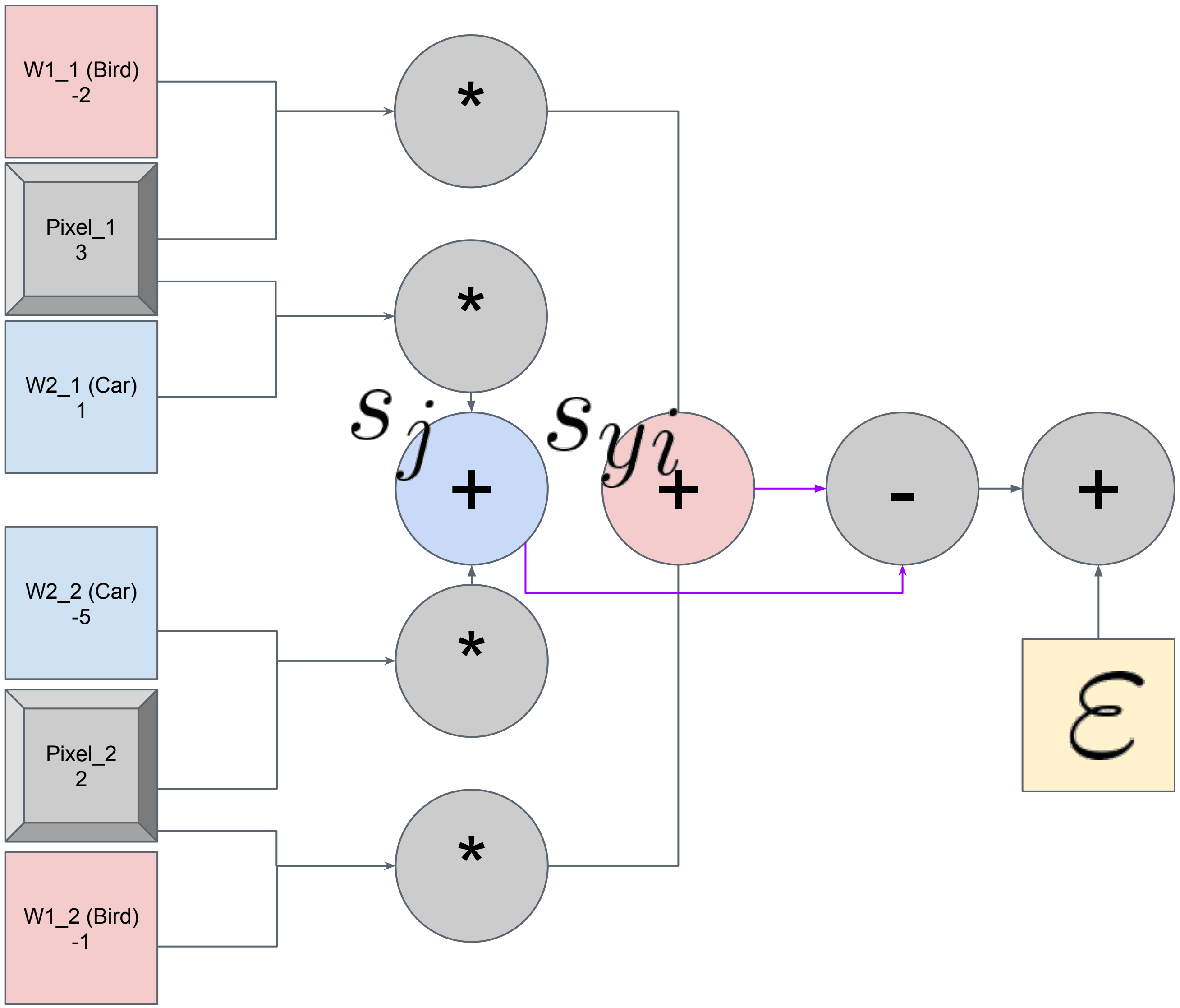
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

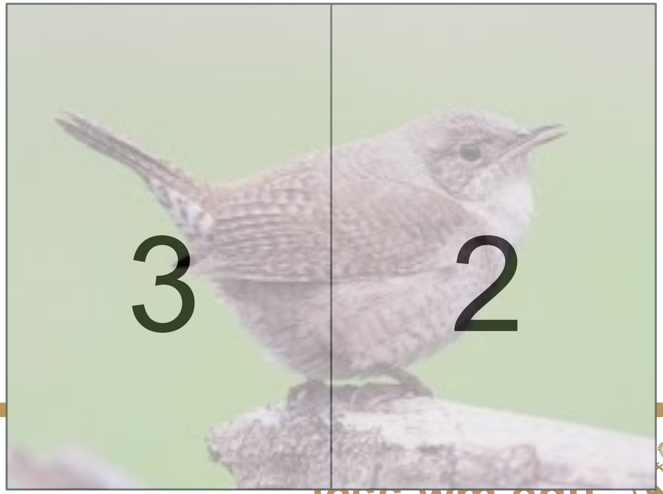


$$\sum_{j \neq y_i}^{J} max(0, \boxed{s_j - s_{y_i}} + \varepsilon)$$

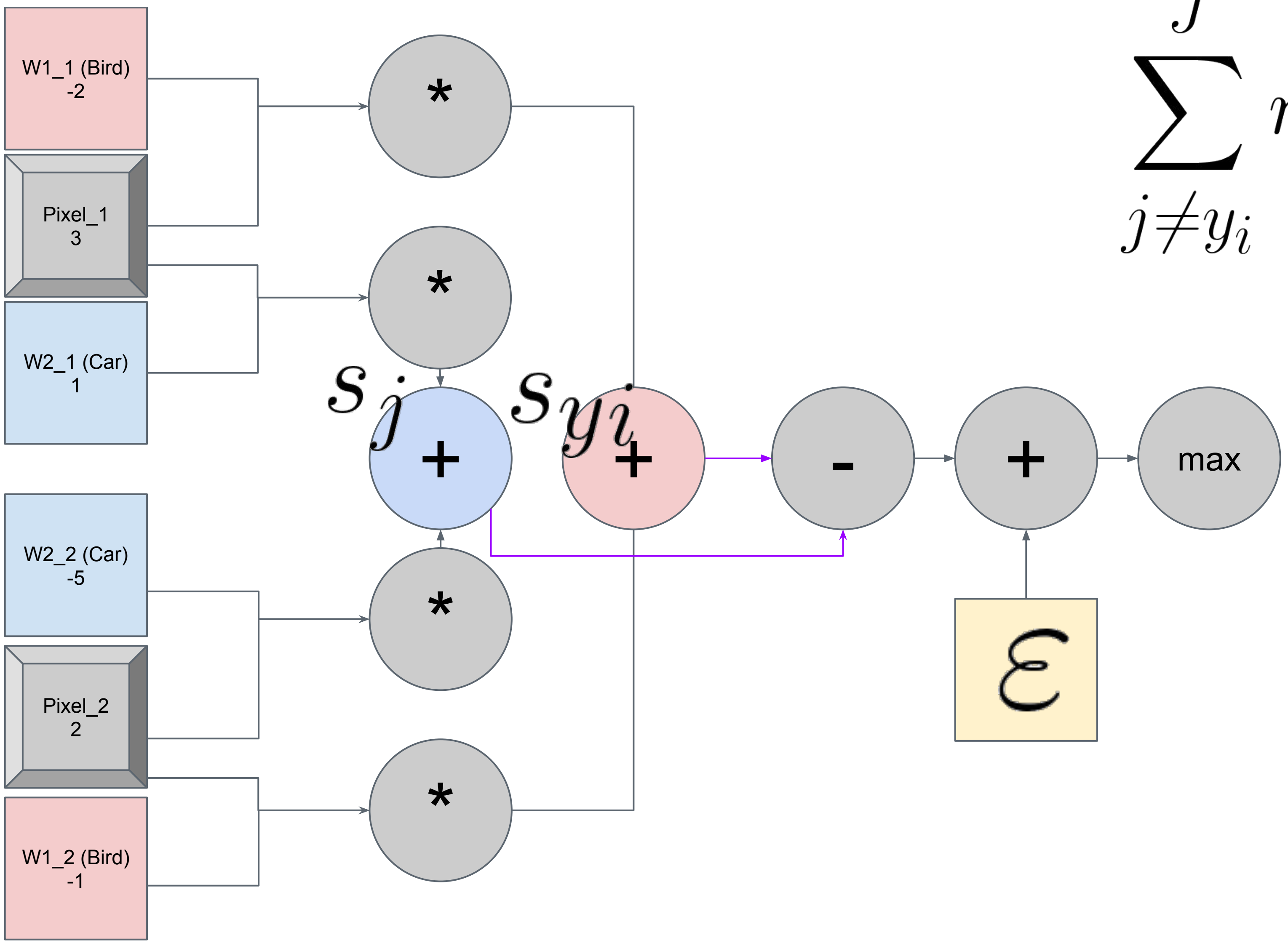


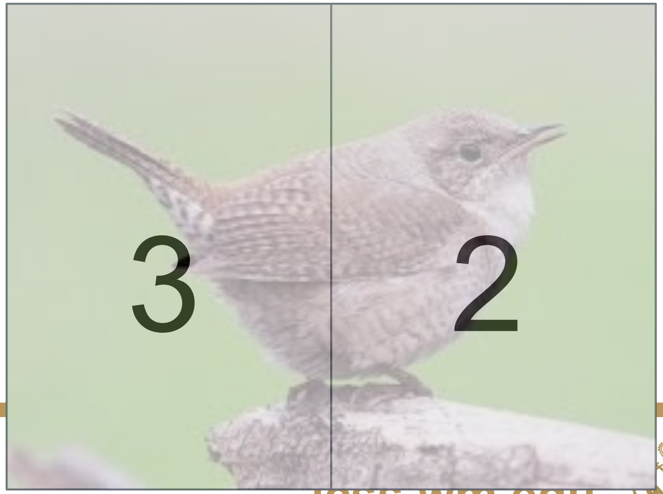
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



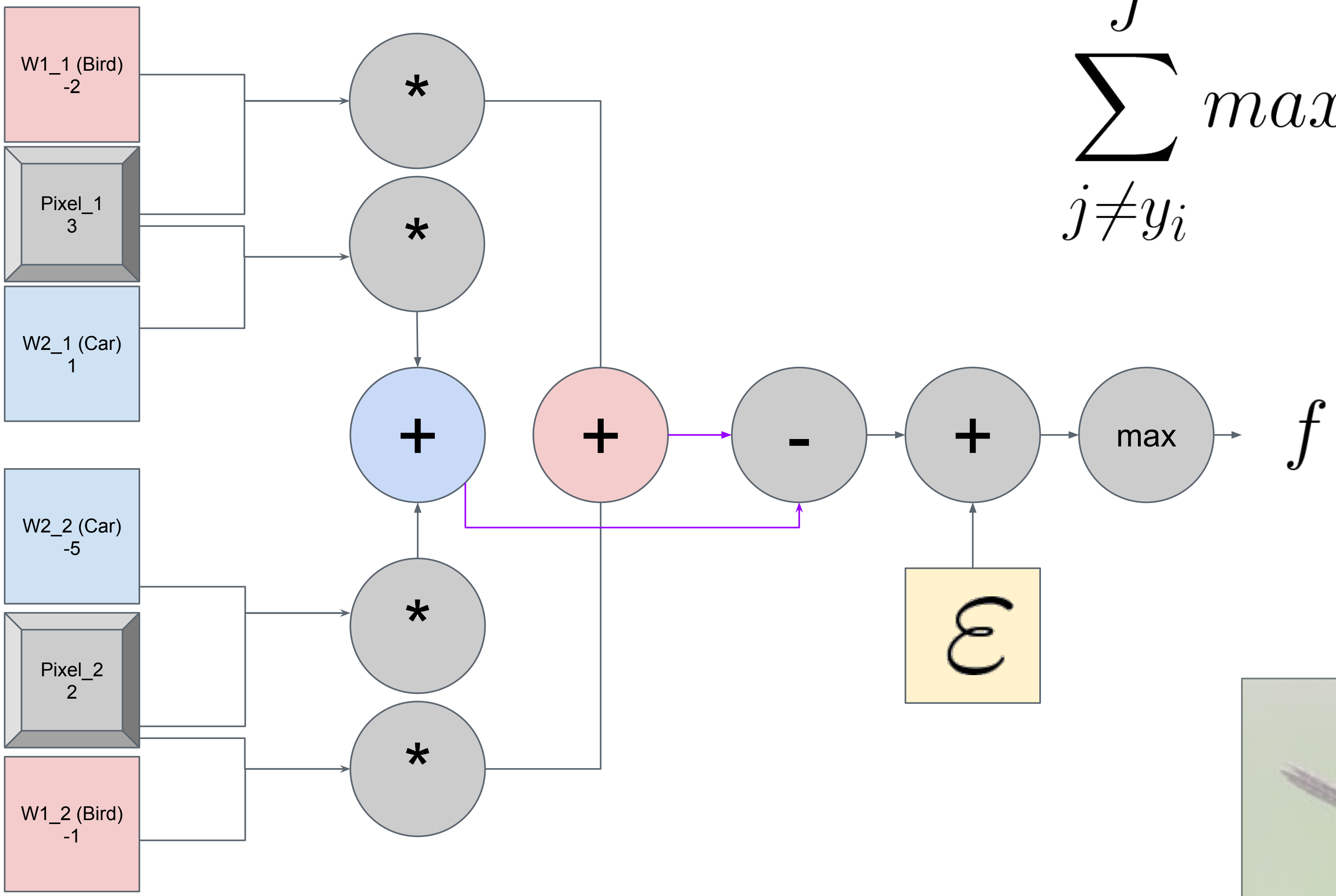


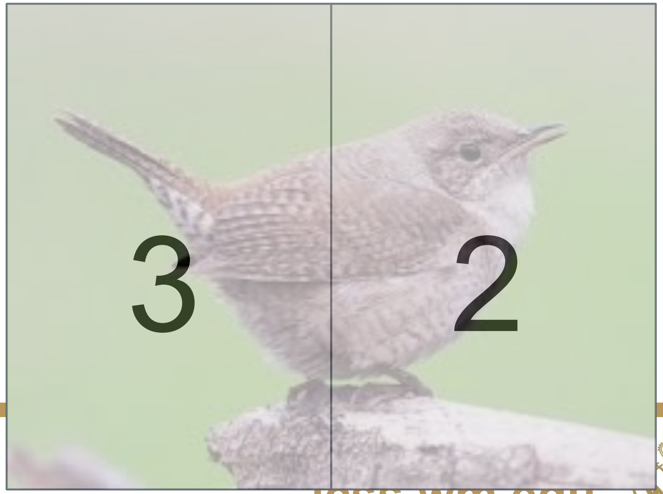
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

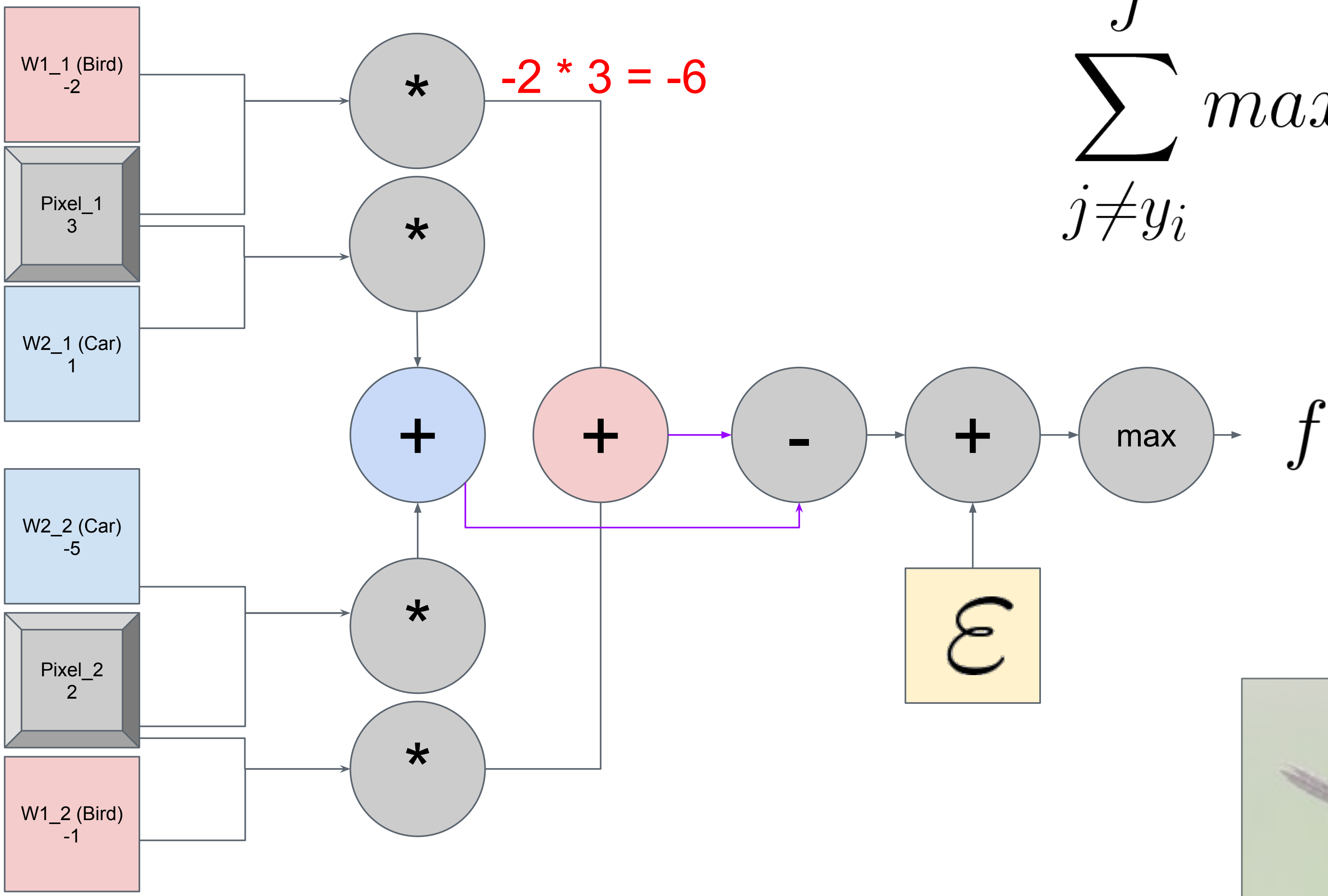




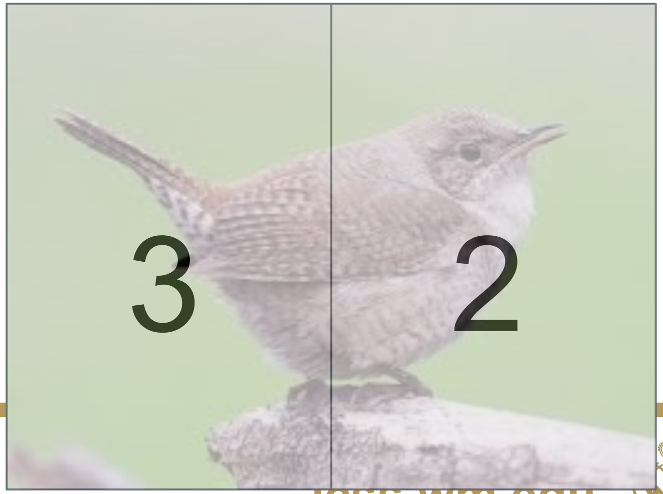
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



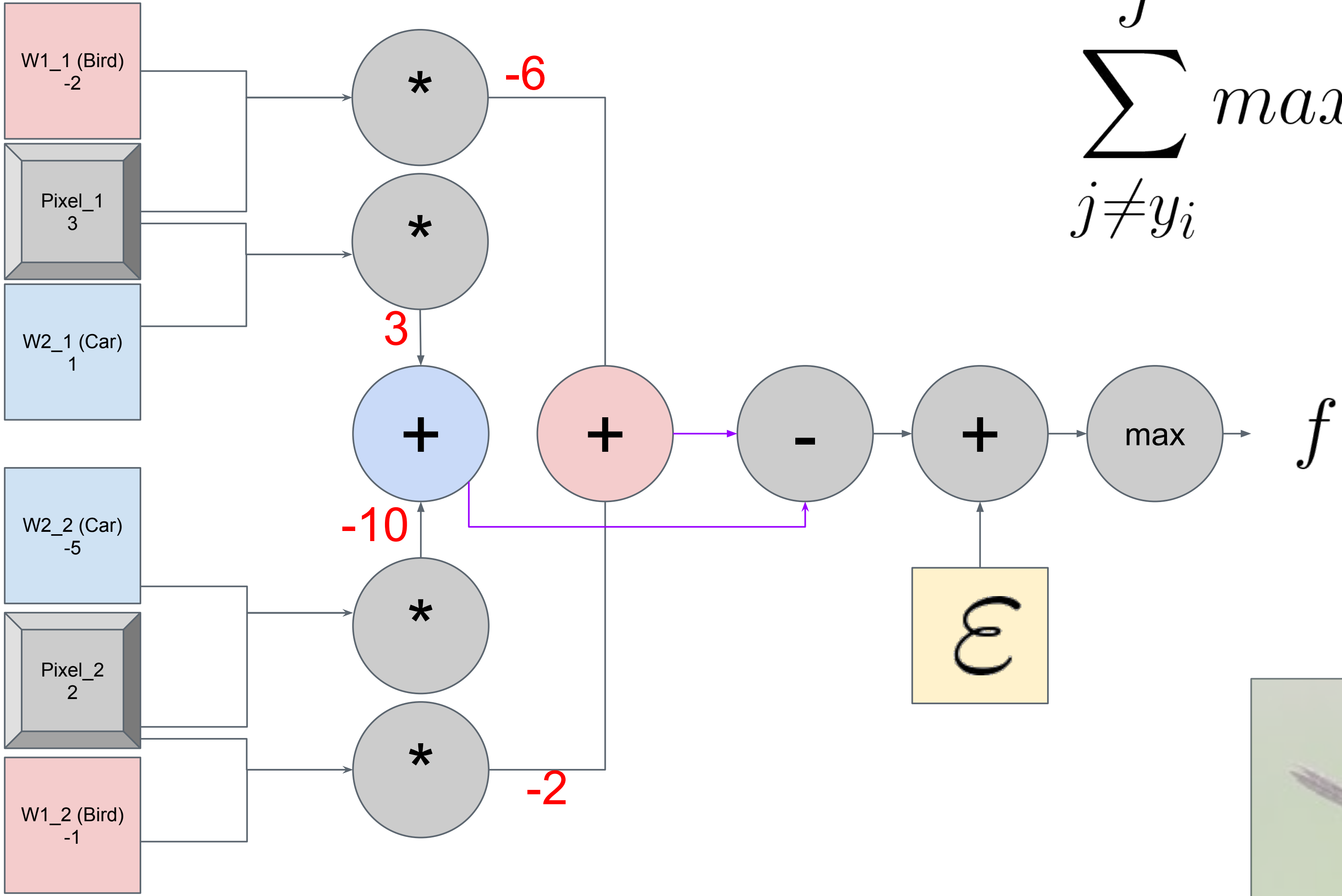




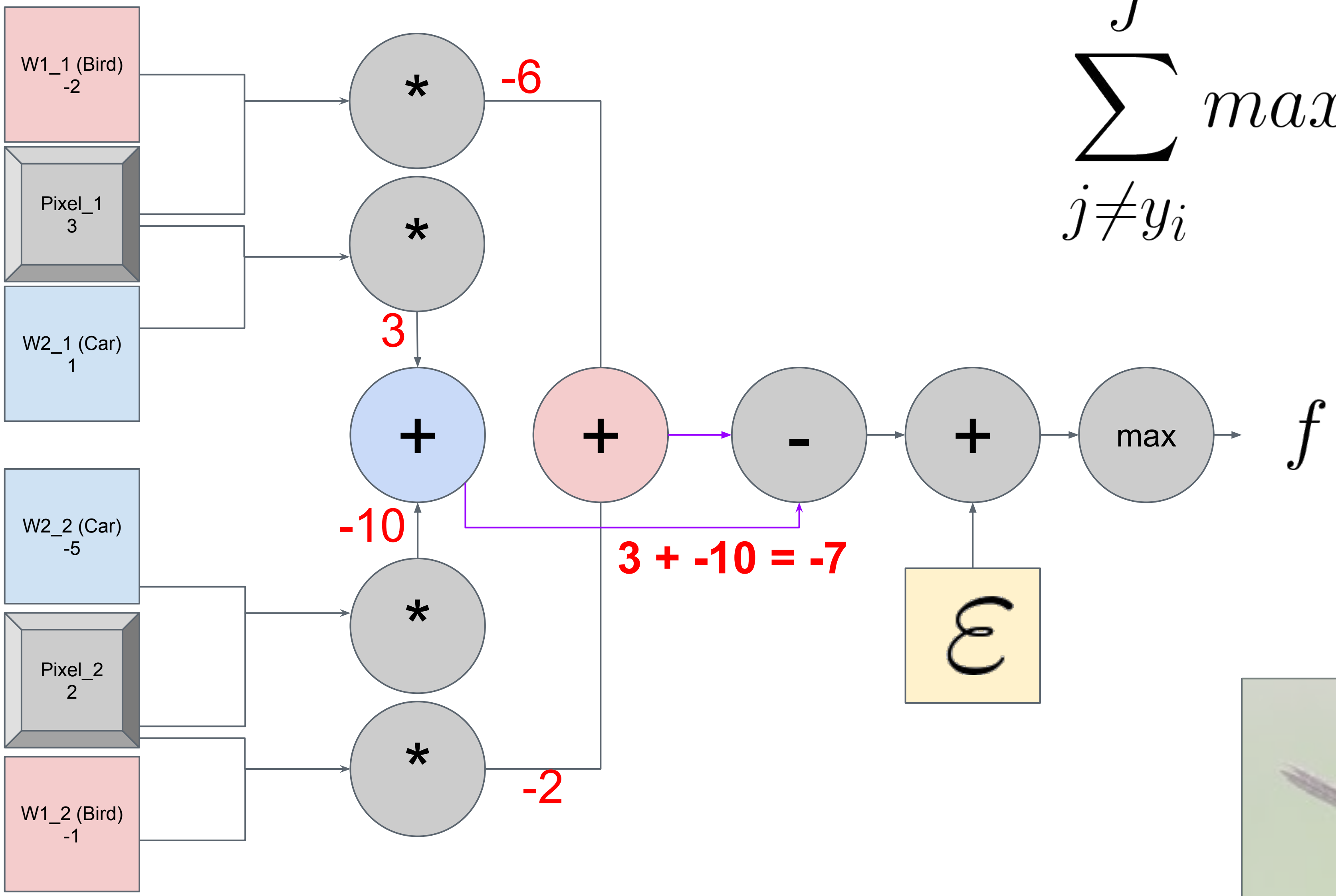
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

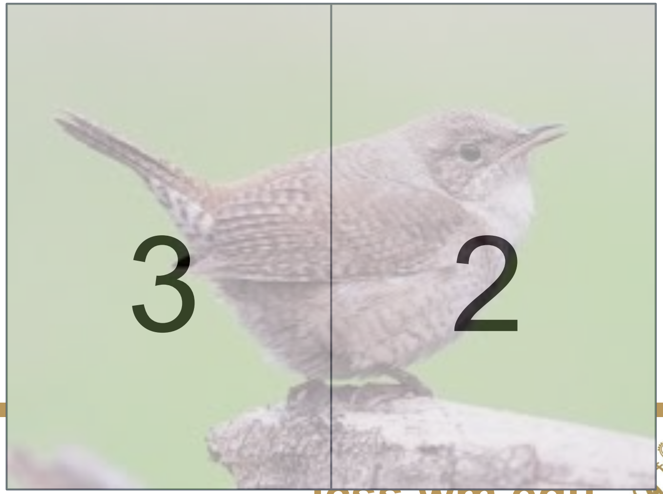


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

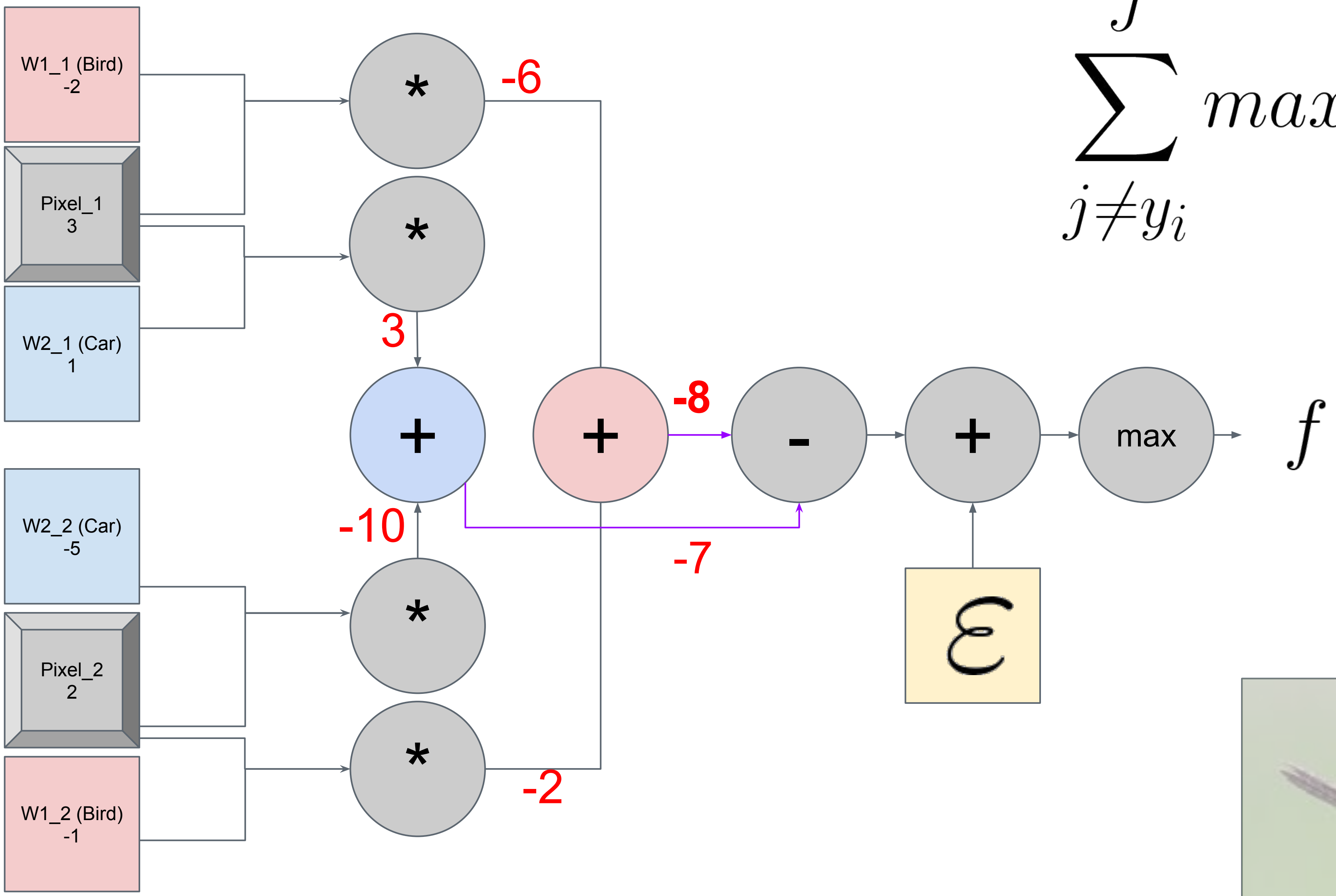


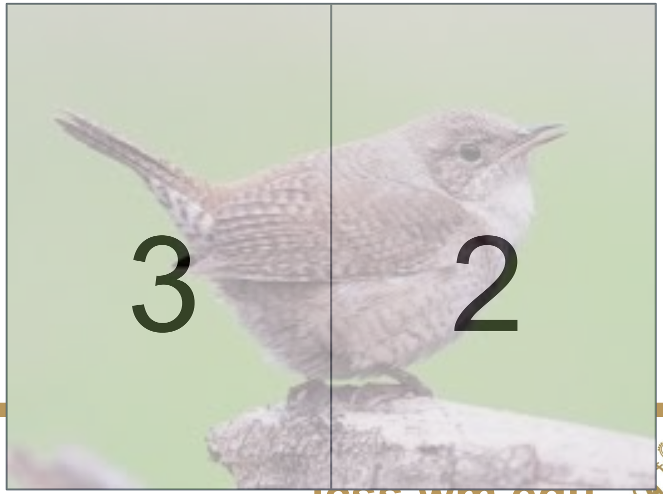
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



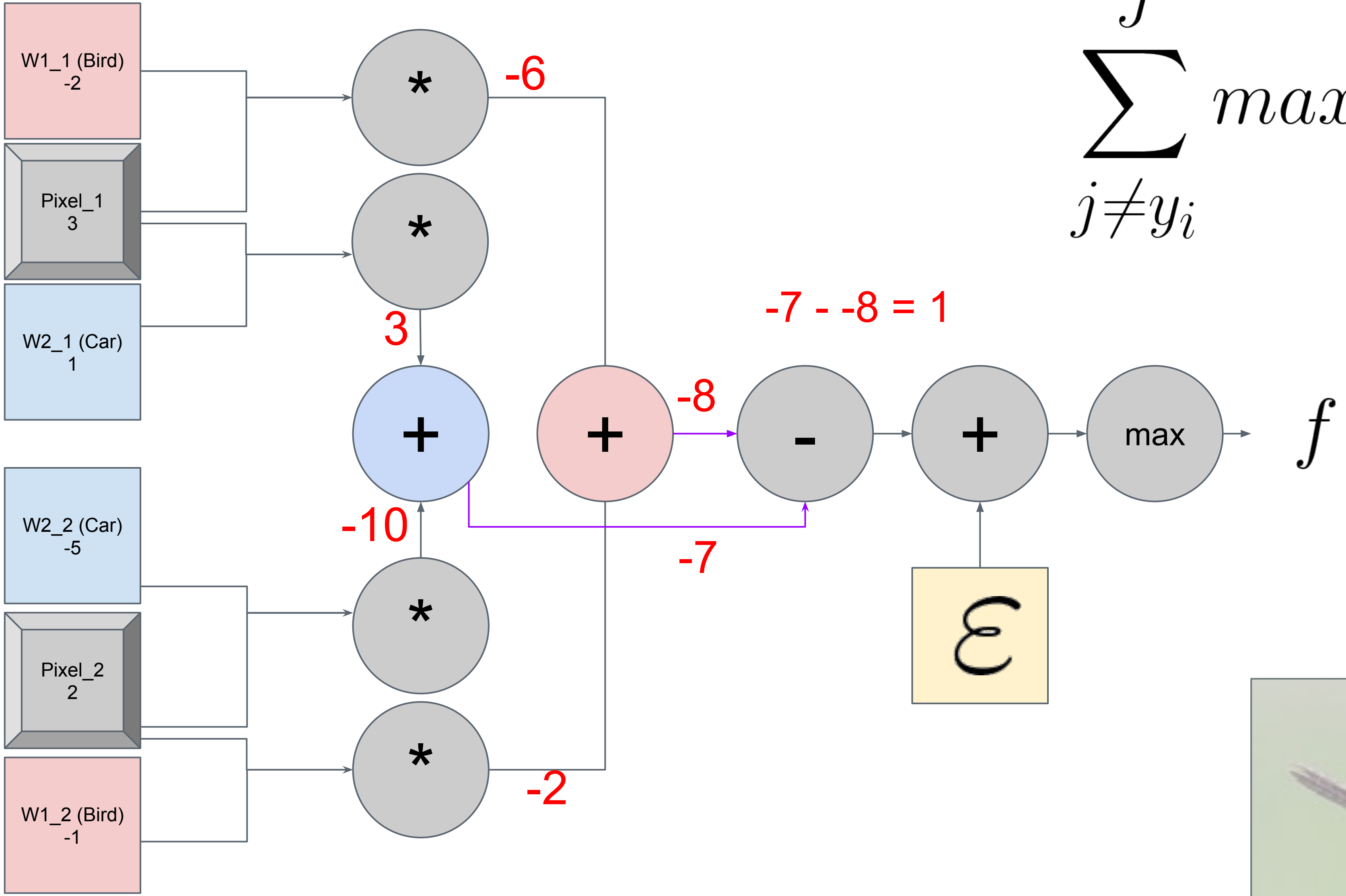


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

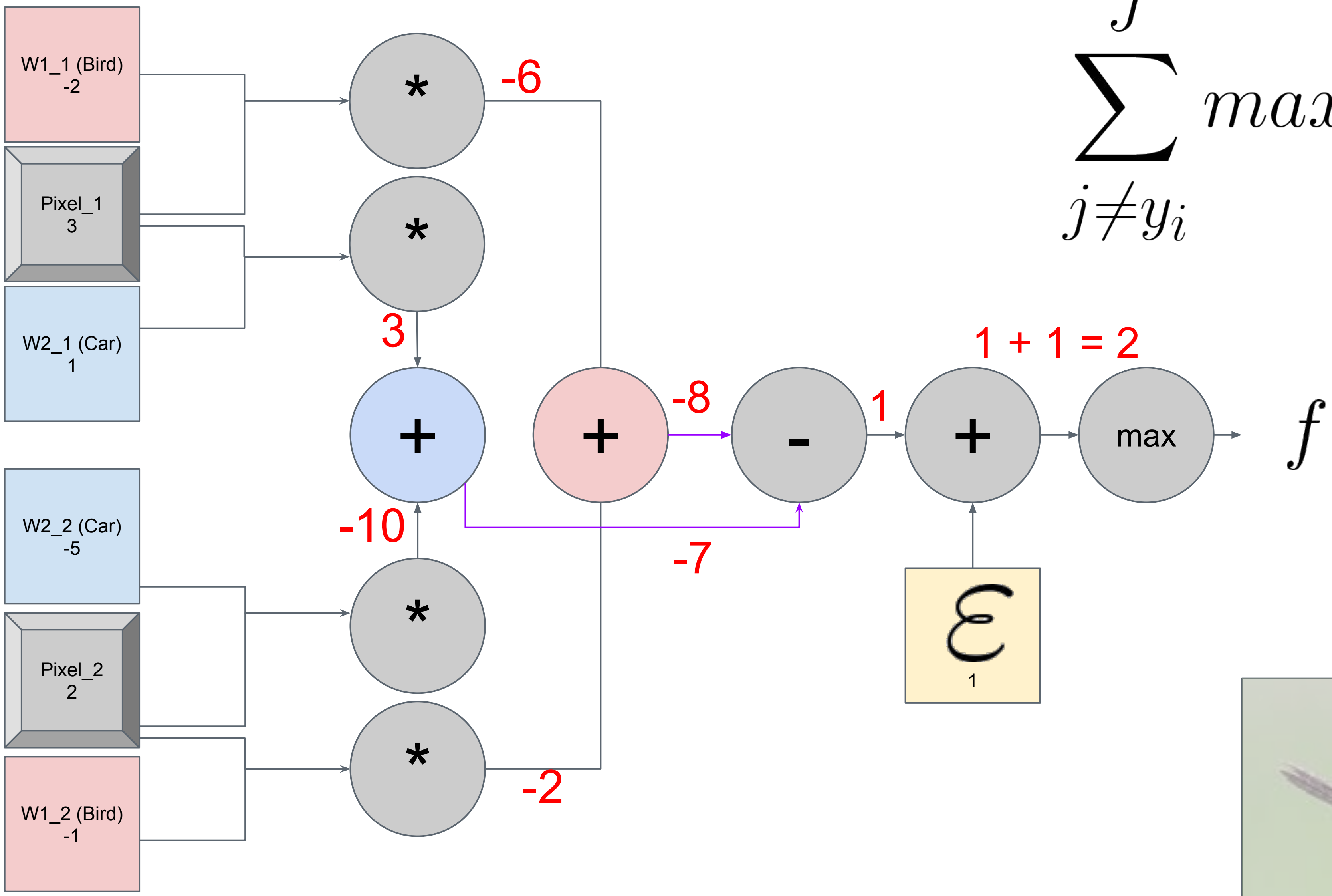


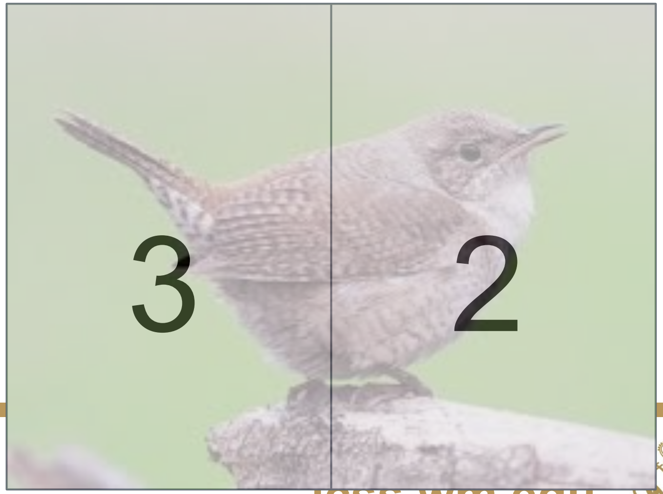


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

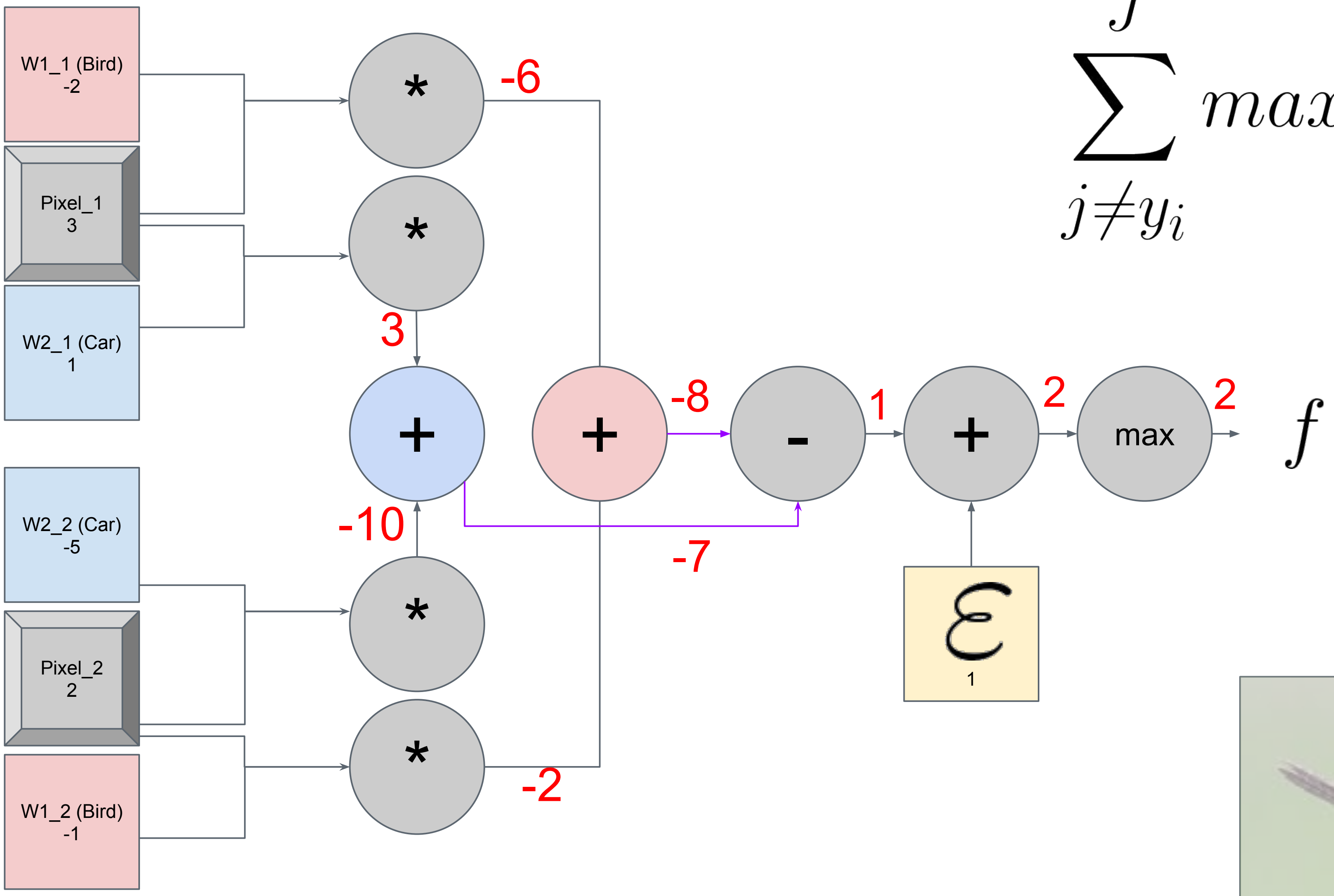


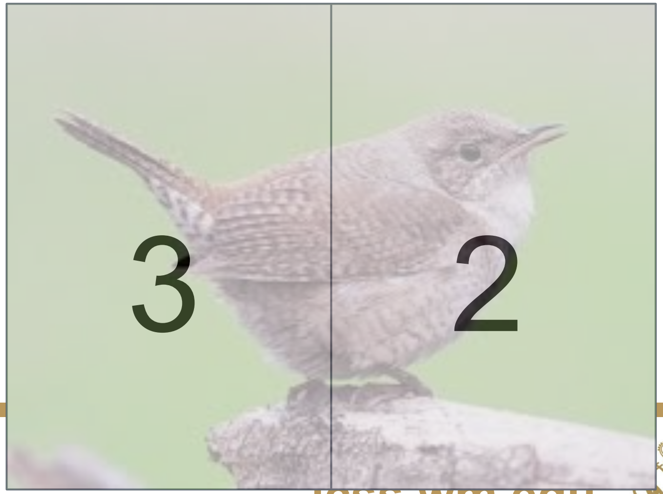
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



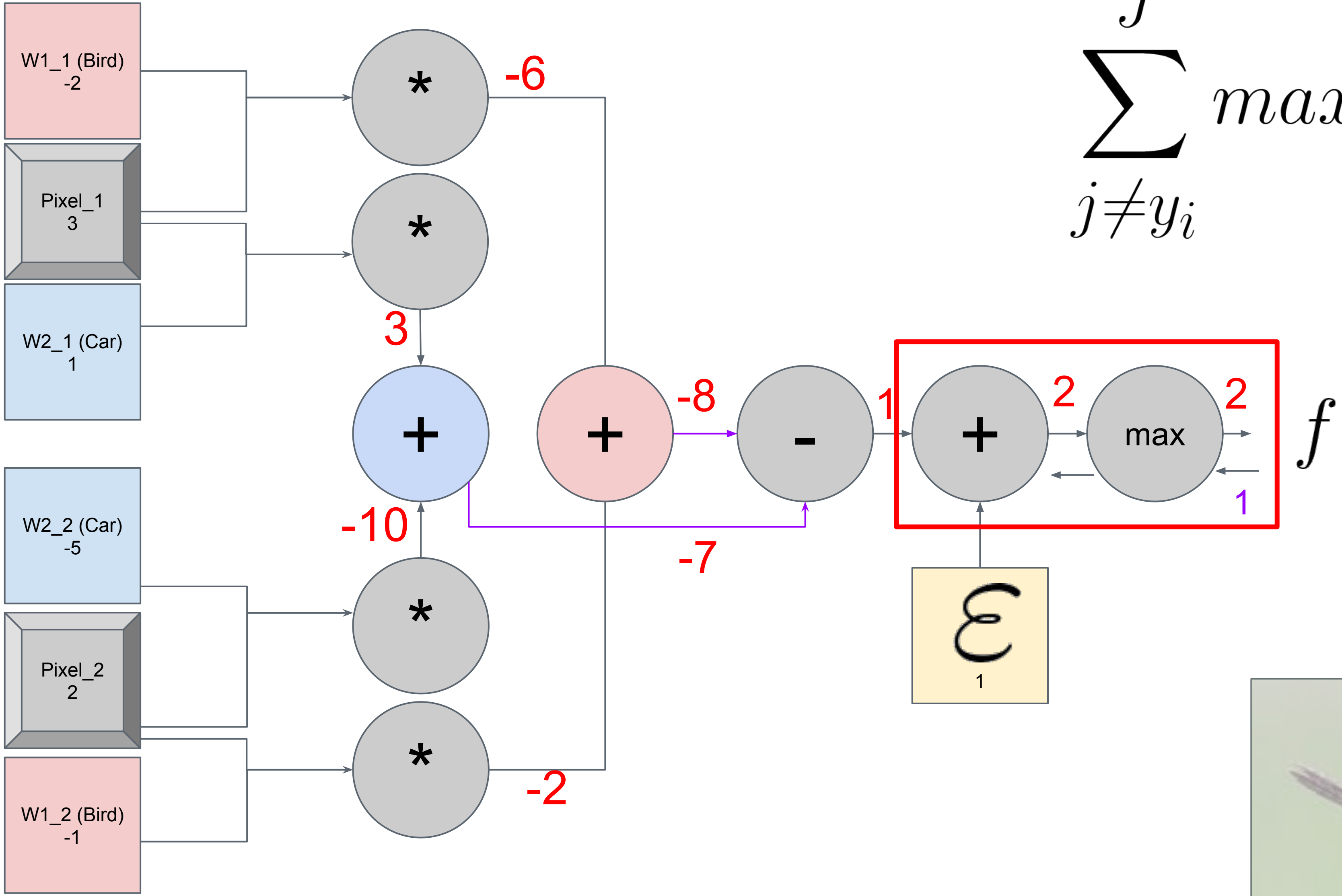


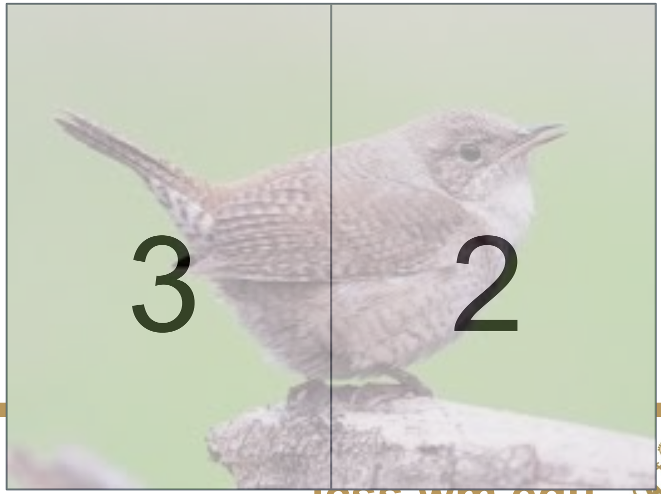
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



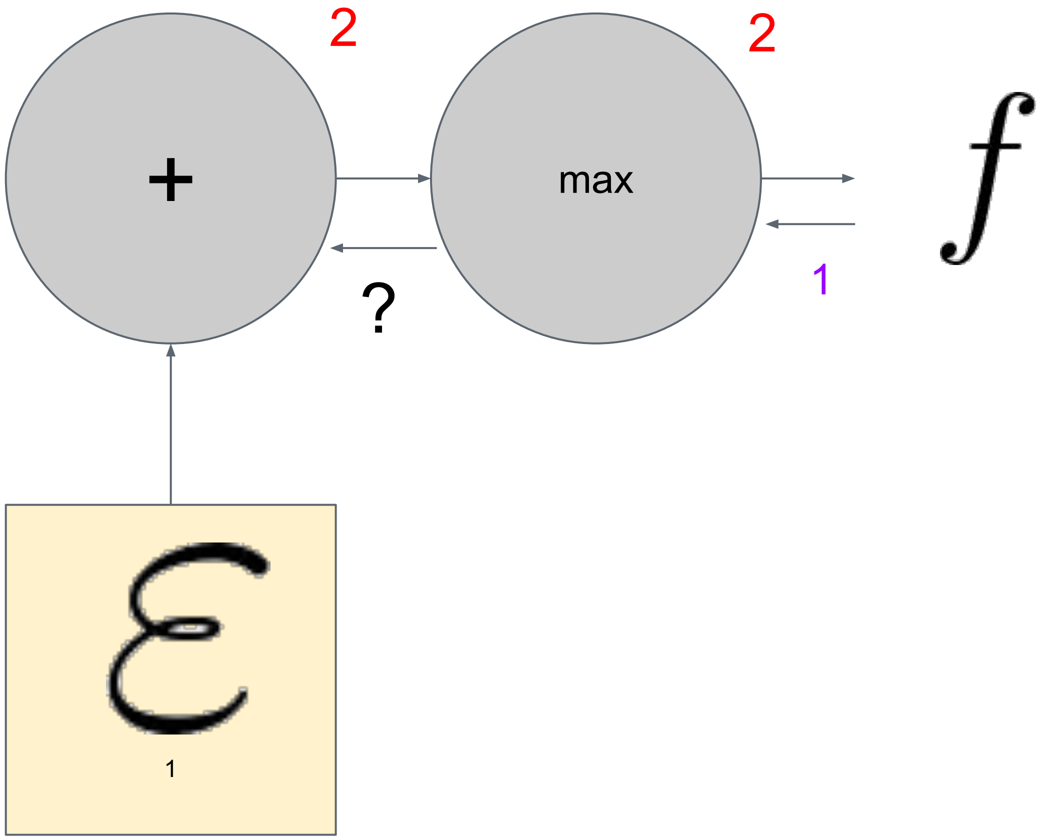


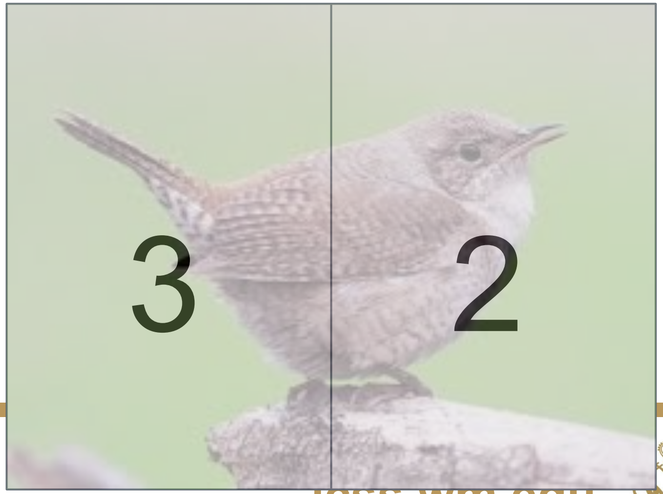
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



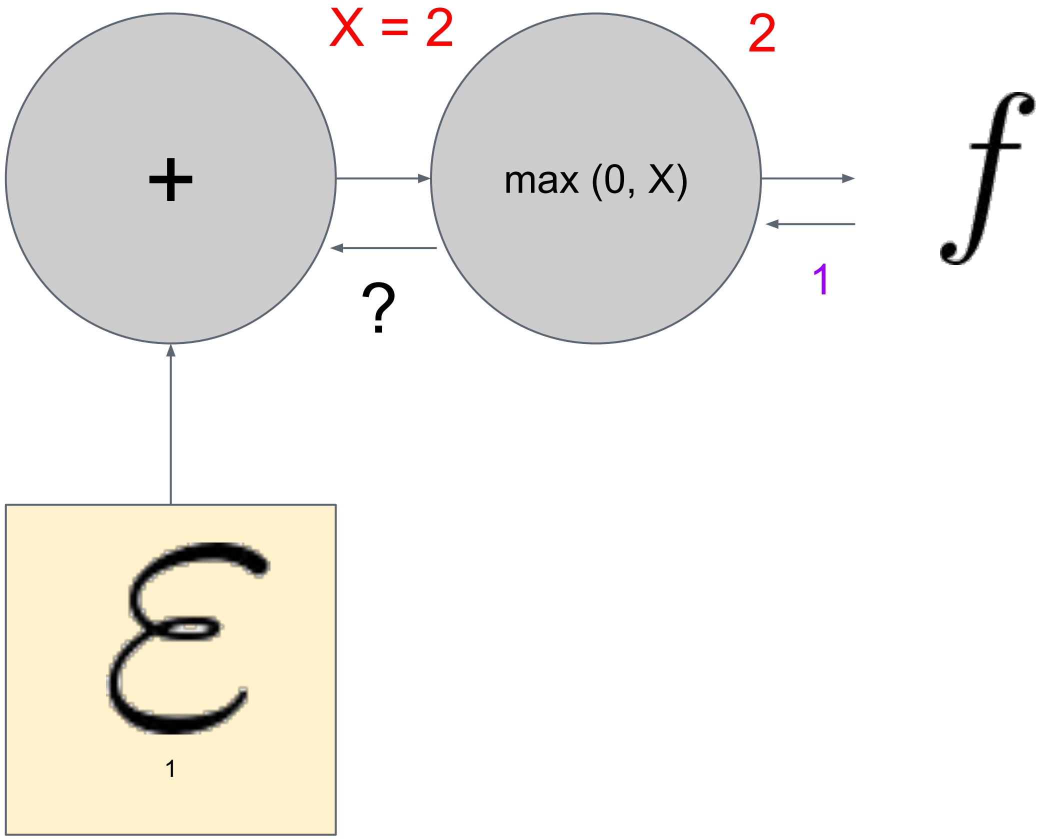


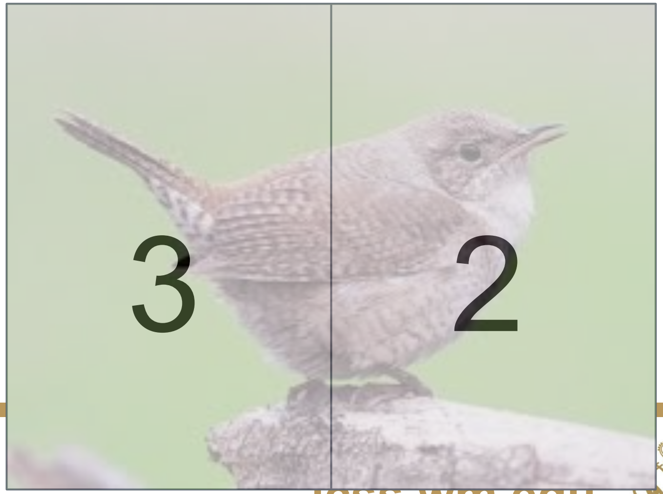
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



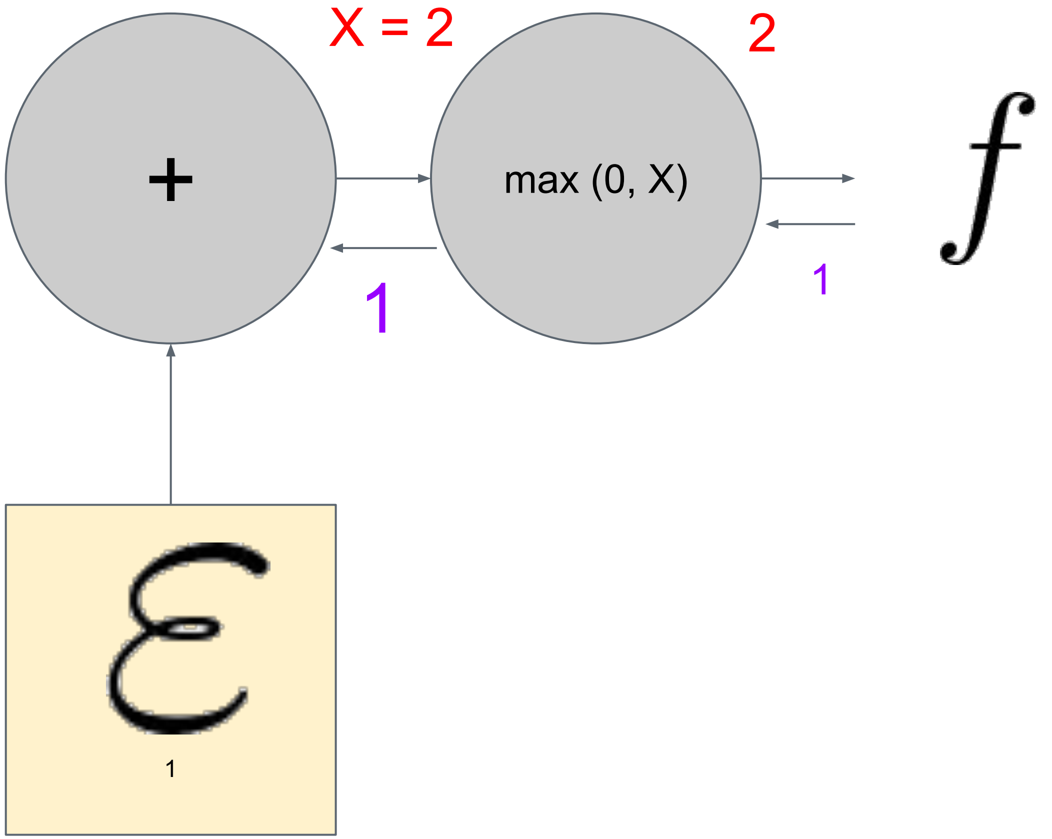


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

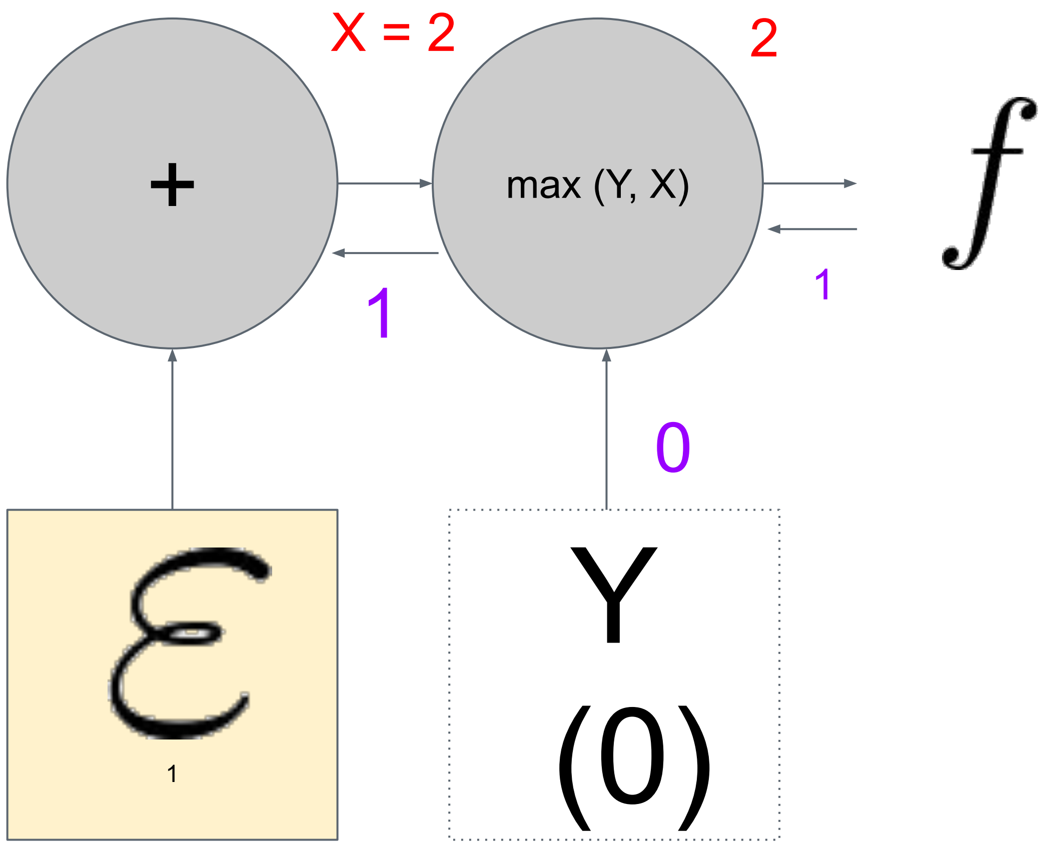


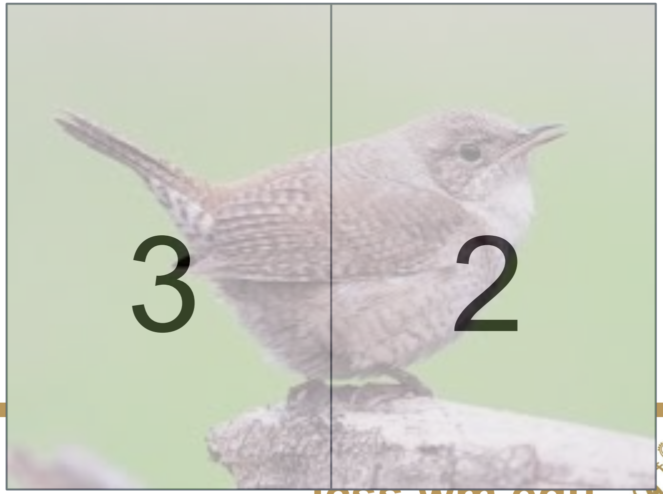


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

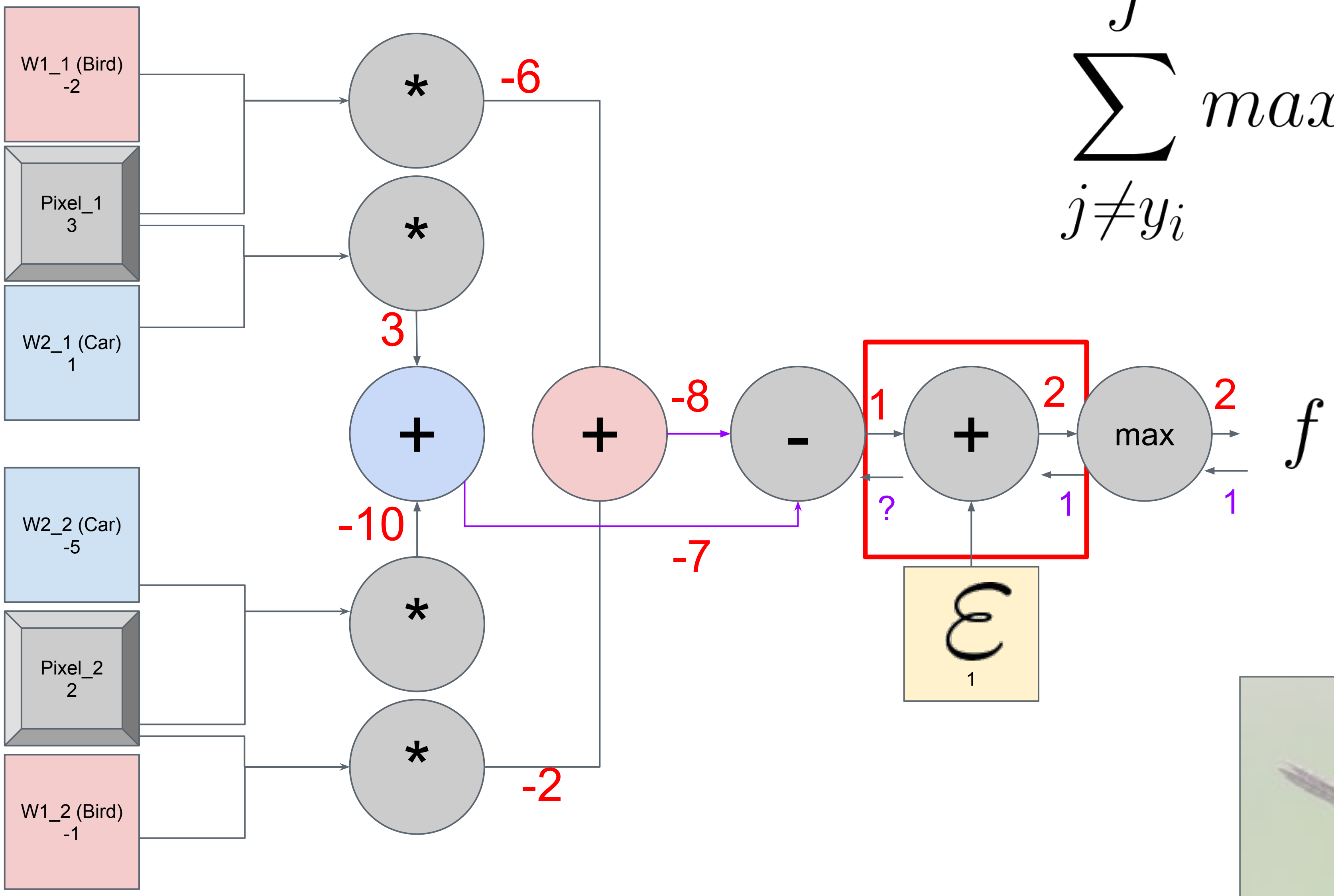


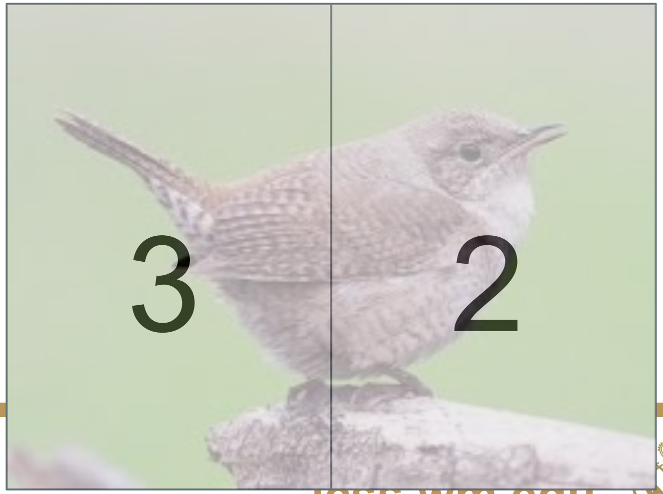
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



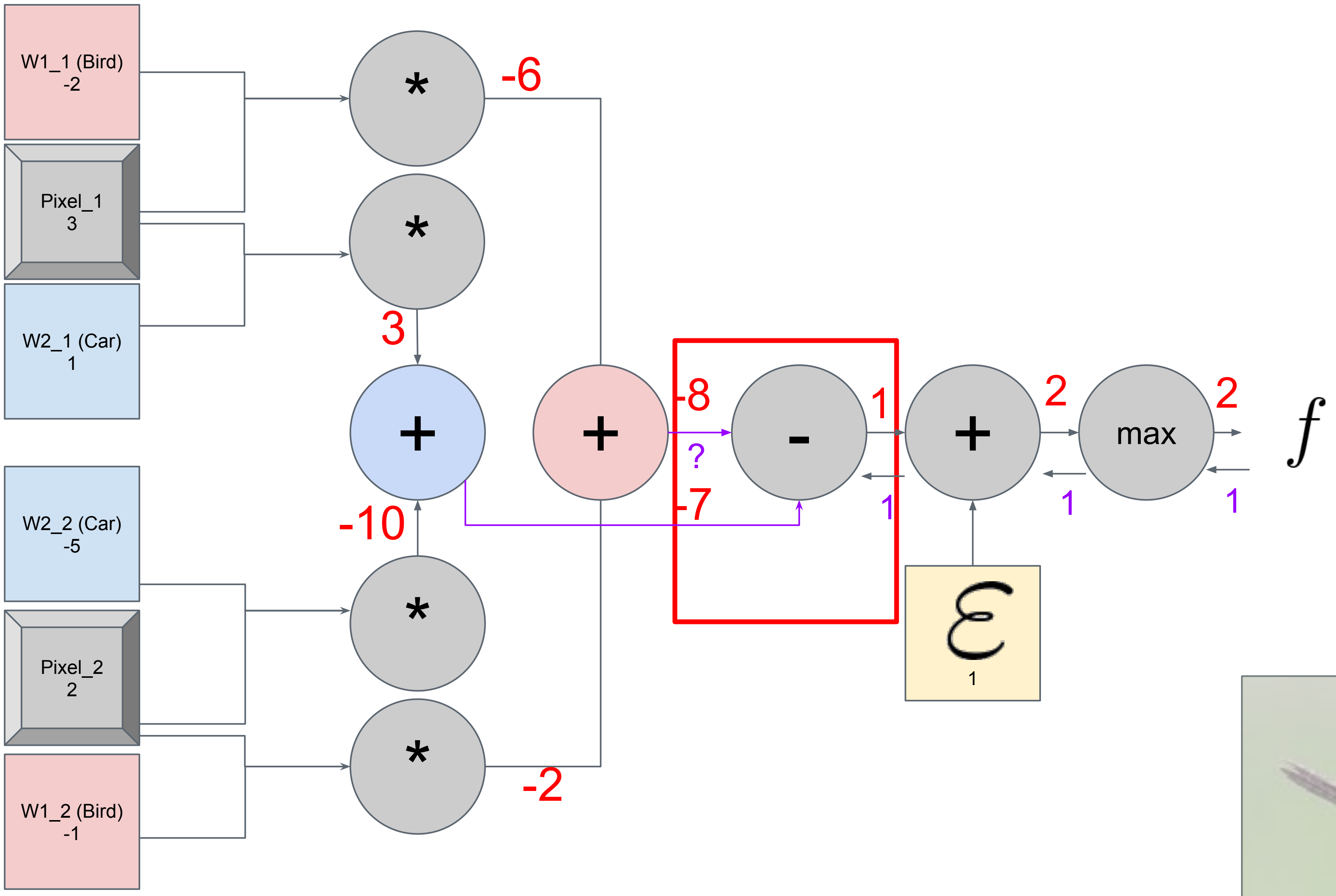


$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

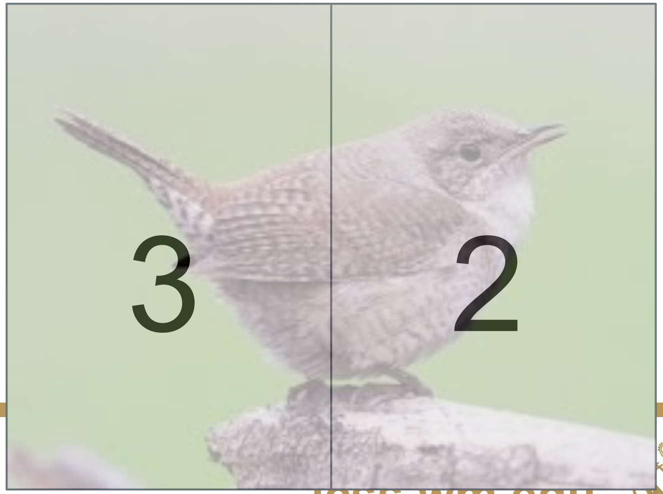




$$U = 1$$



W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
*
*
*
*
-6
3
-10
-2
+
+
-8
?
-7
-
1
1
+
2
1
max
2
1
f
ε
1

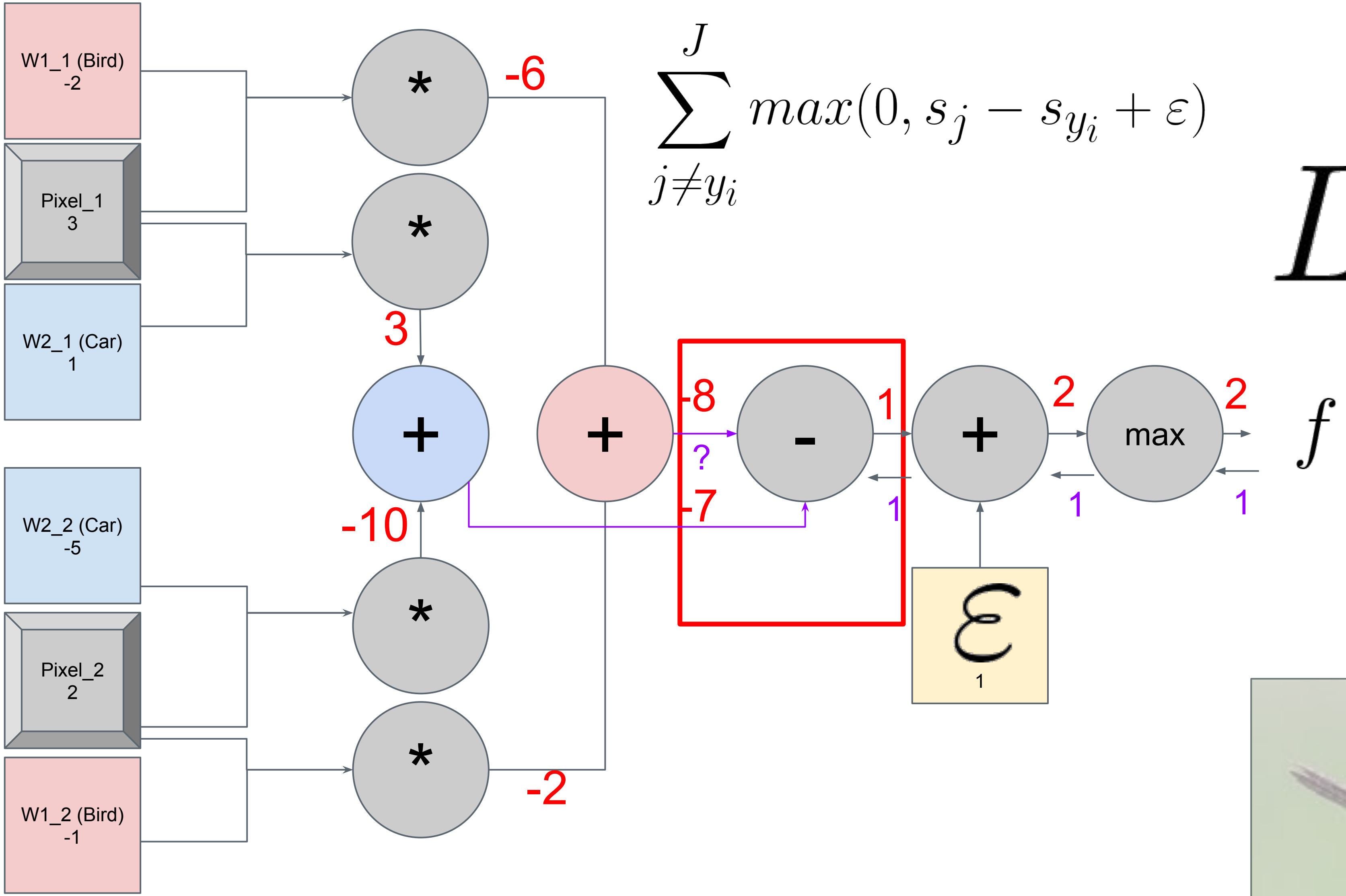


3
2

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$

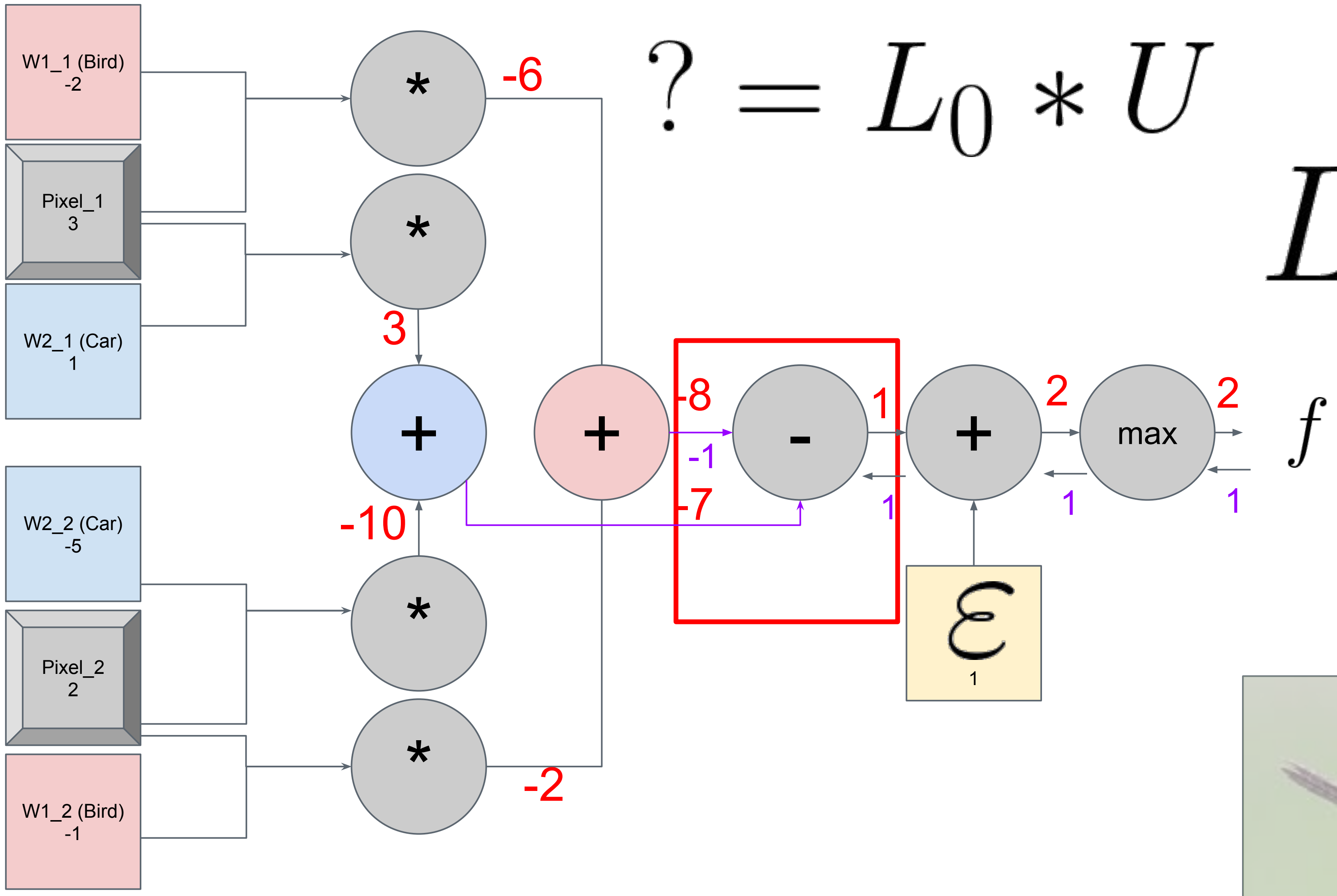
$$U = 1$$

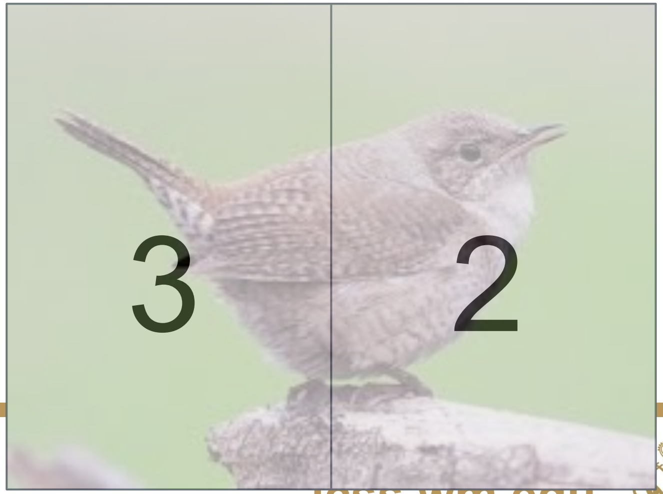
$$L_0 = -1$$



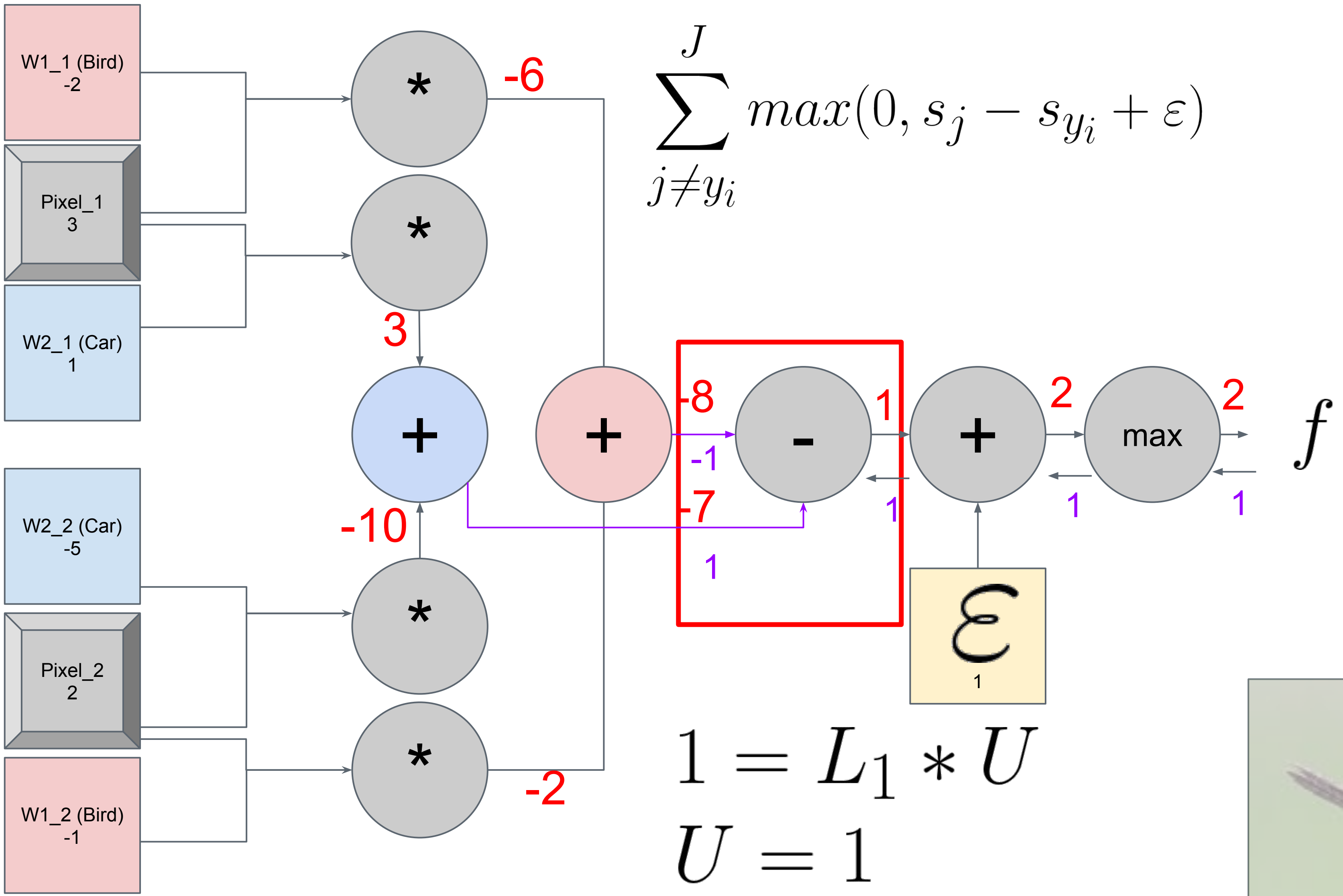
$$? = L_0 * U \qquad U = 1$$

$$L_0 = -1$$





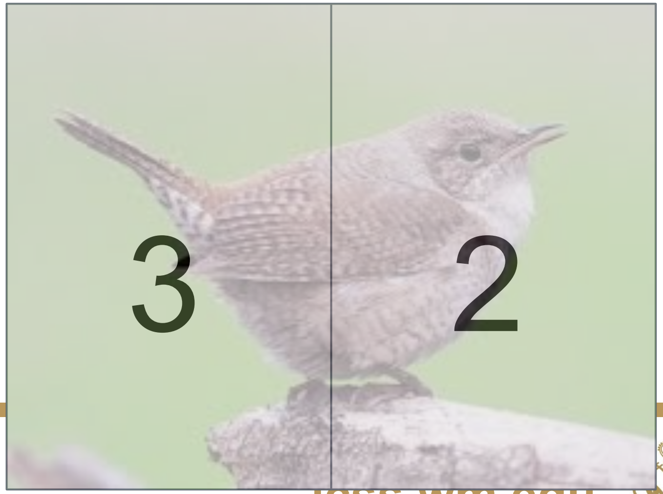
$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$



$$1 = L_1 * U$$

$$U = 1$$

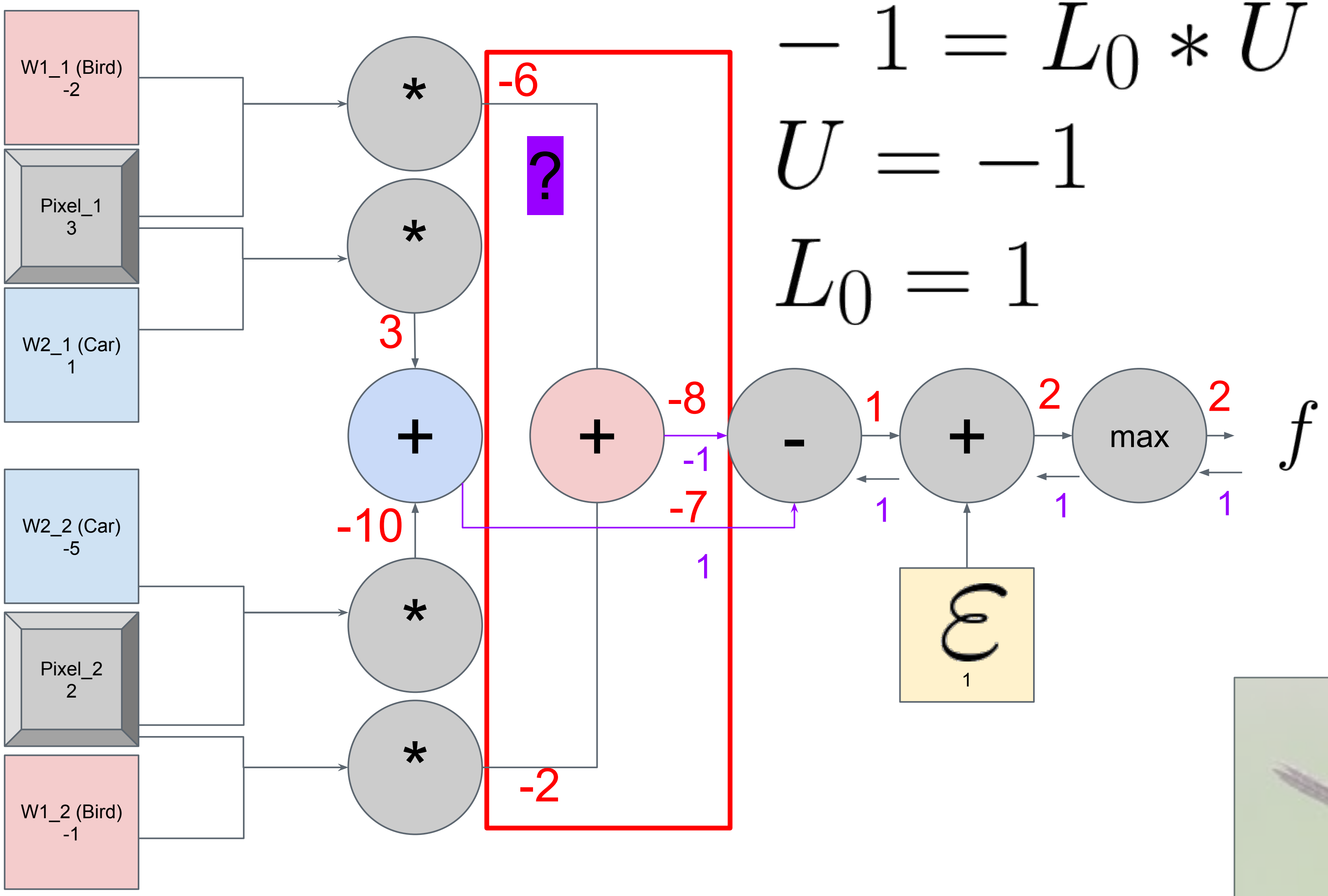
$$L_1 = 1$$



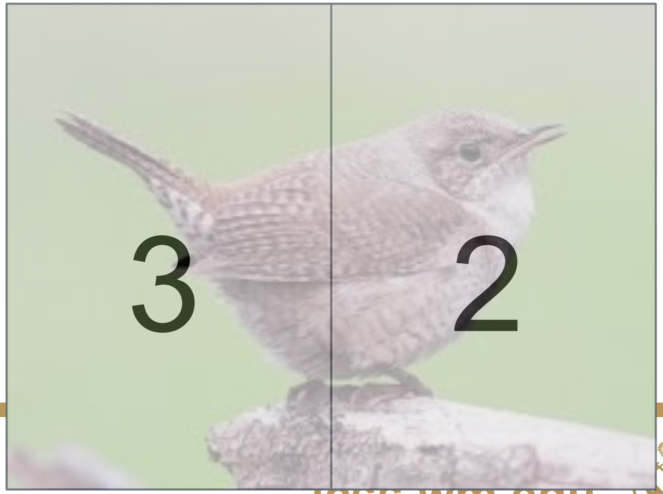
$$-1 = L_0 * U$$

$$U = -1$$

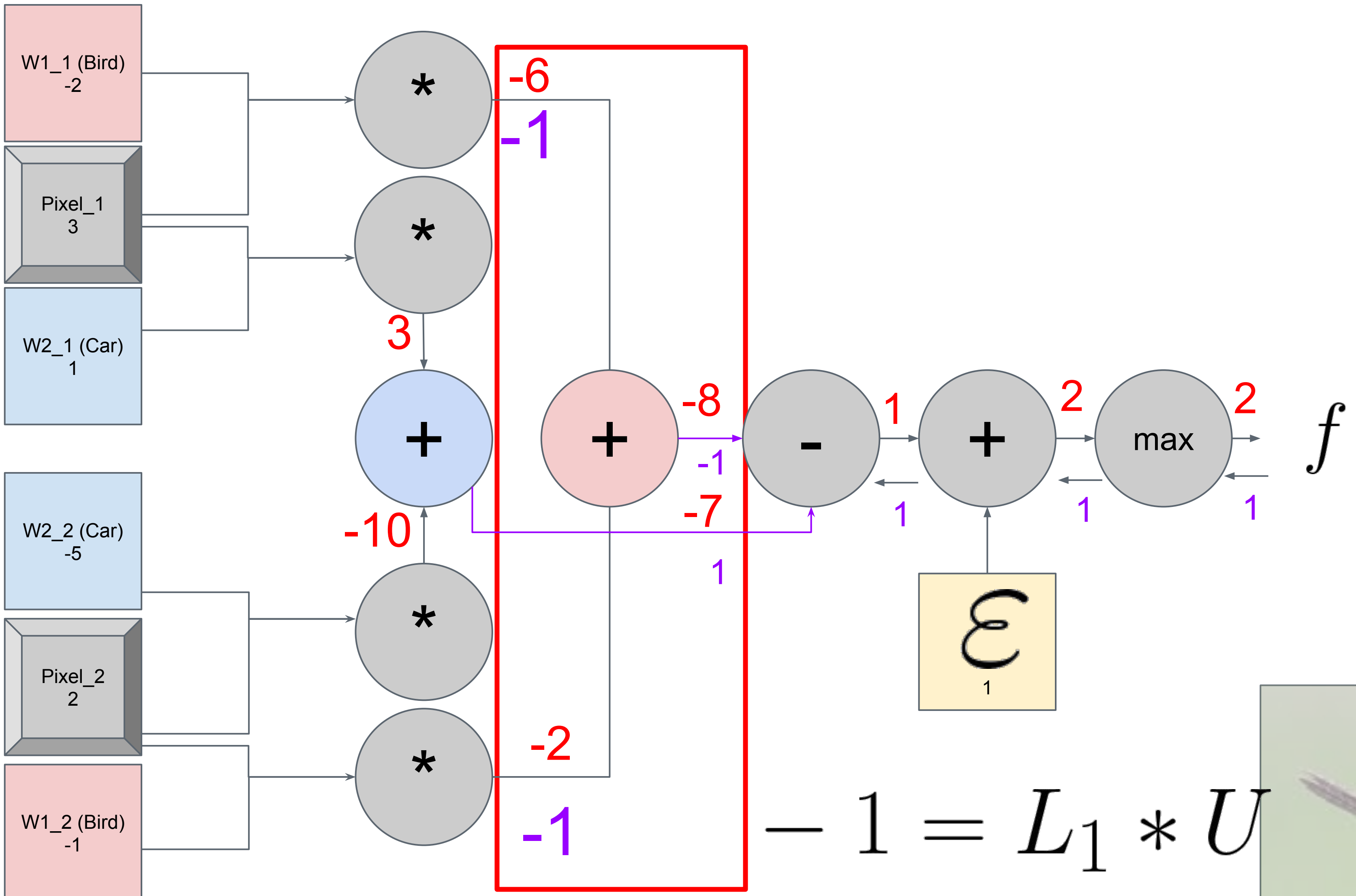
$$L_0 = 1$$

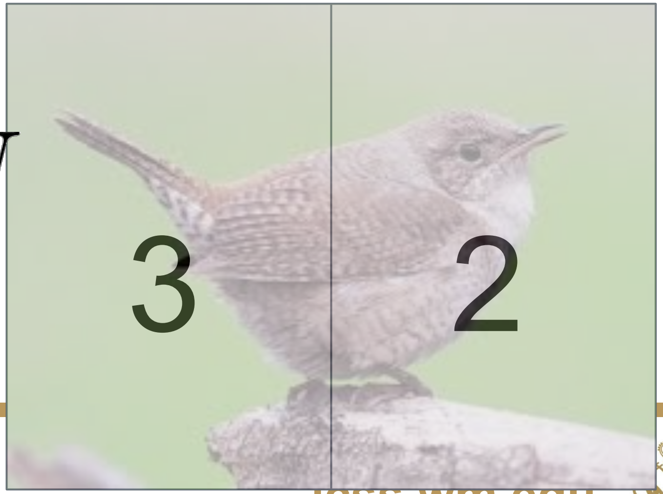


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
*
*
*
*
+
+
-
+
max
-6
?
3
-10
-2
-8
-1
-7
1
1
2
2
1
1
1
f
$\mathcal{E}$
1



3
2

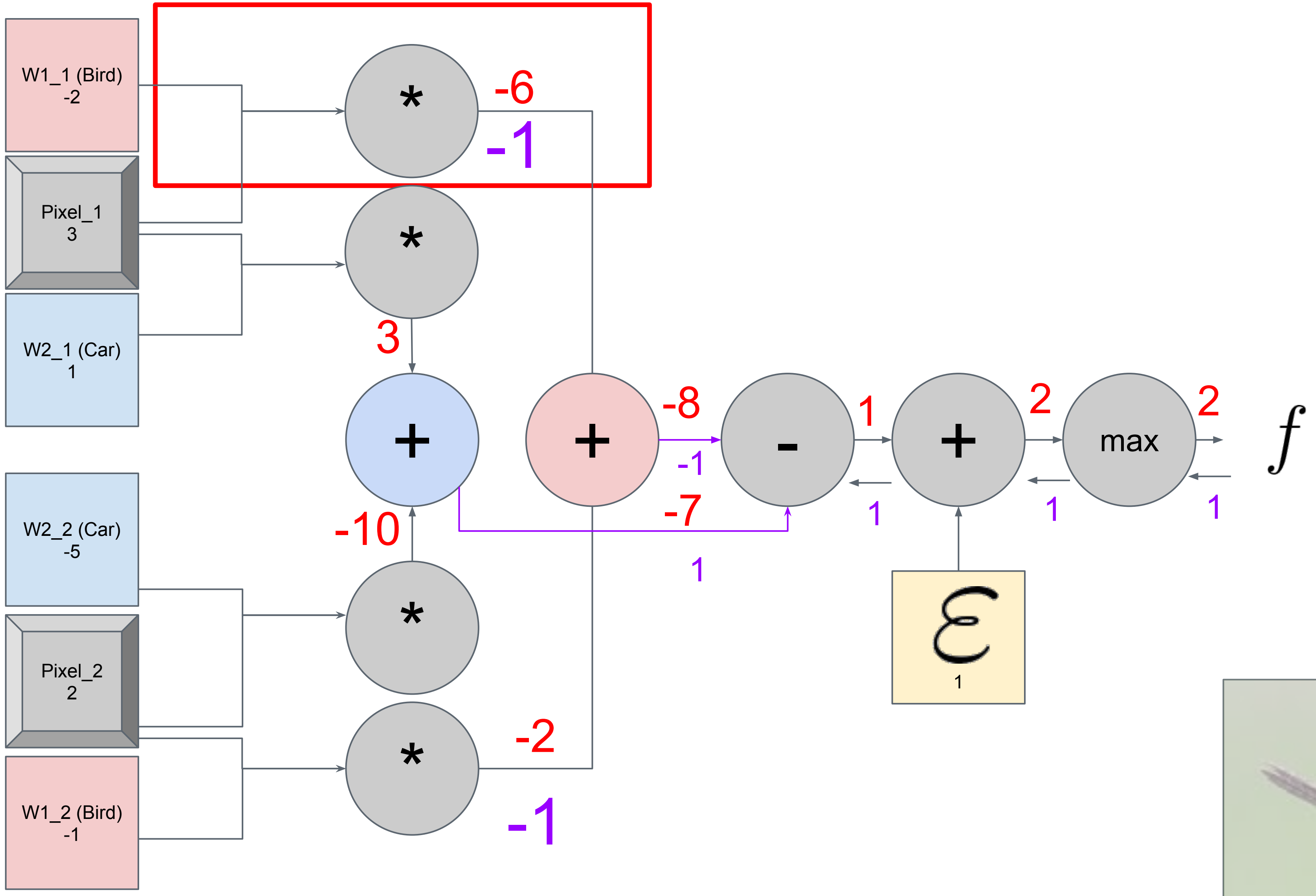

W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
*
*
+
*
*
+
-
+
max
-6
-1
3
-10
-8
-1
-7
1
1
2
1
2
1
-2
-1
1
1
f


3
2

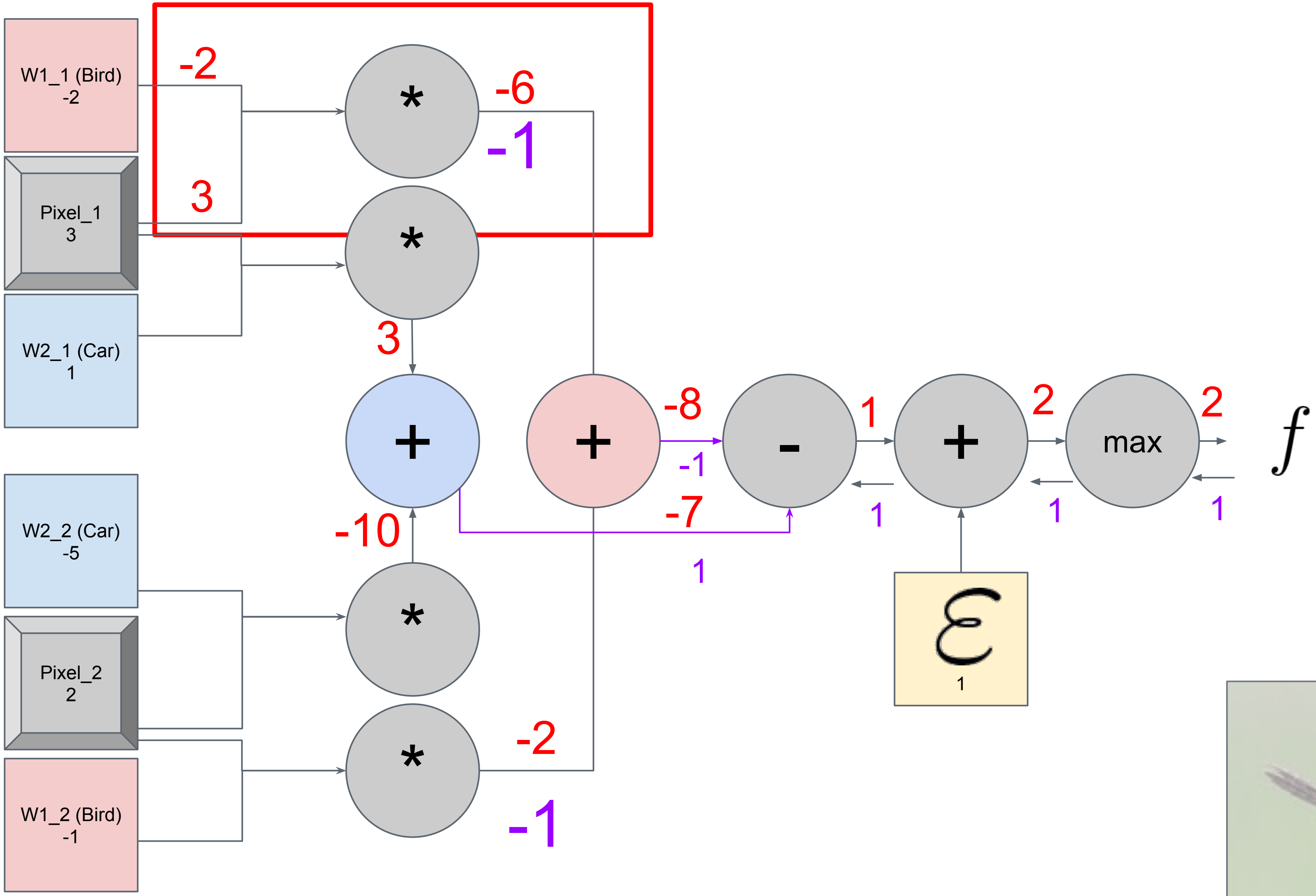
$$-1 = L_1 * U$$

$$U = -1$$

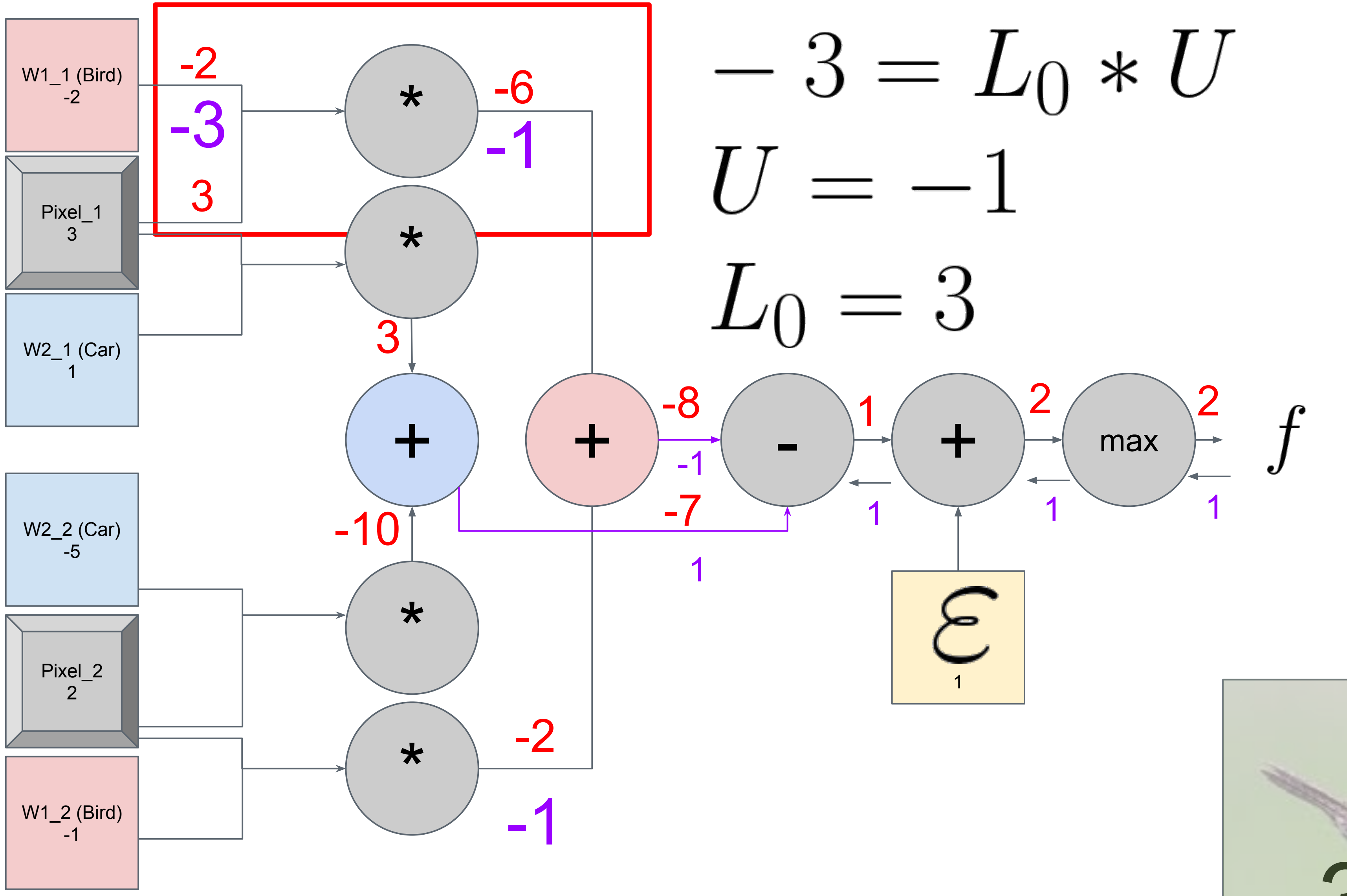
$$L_1 = 1$$


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
*
*
*
*
+
+
-
+
max
-6
-1
3
-10
-2
-1
-8
-1
-7
1
1
1
2
1
2
1
f
ε
1


3
2


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
-2
3
*
*
*
*
-6
-1
3
-10
-2
-1
+
+
-8
-1
-7
1
-
1
1
+
2
1
max
2
1
f
ε
1


3
2

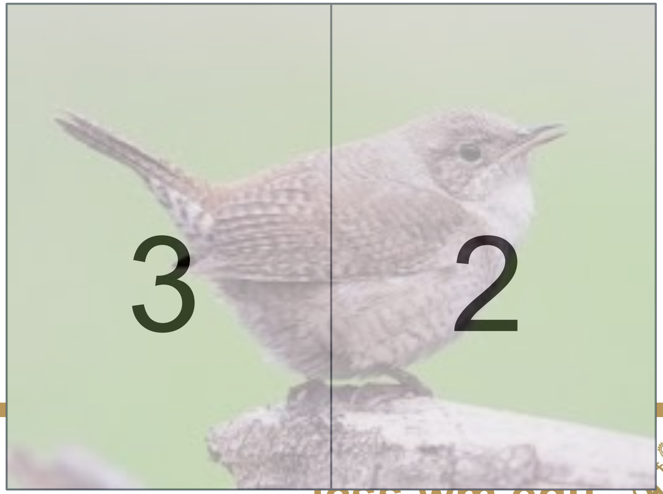


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
-2
-3
3
*
-6
-1
*
3
+
+
-8
-1
-
1
+
2
max
2
f
-10
-7
1
1
1
1
ε
1
*
*
-2
-1

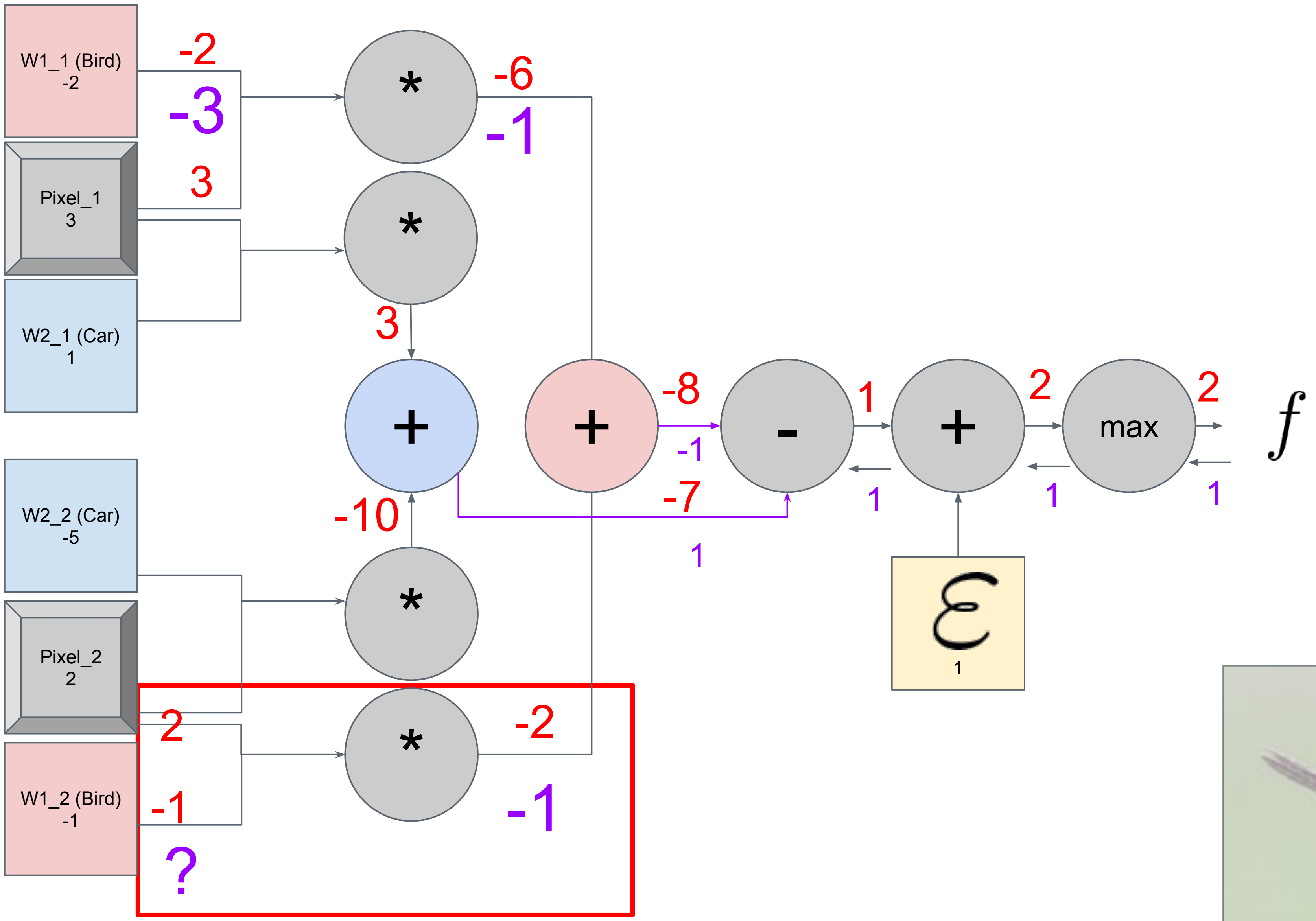
$$-3 = L_0 * U$$

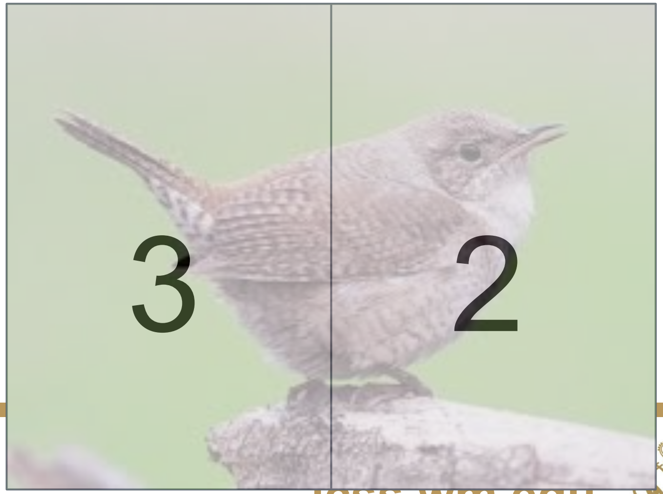
$$U = -1$$

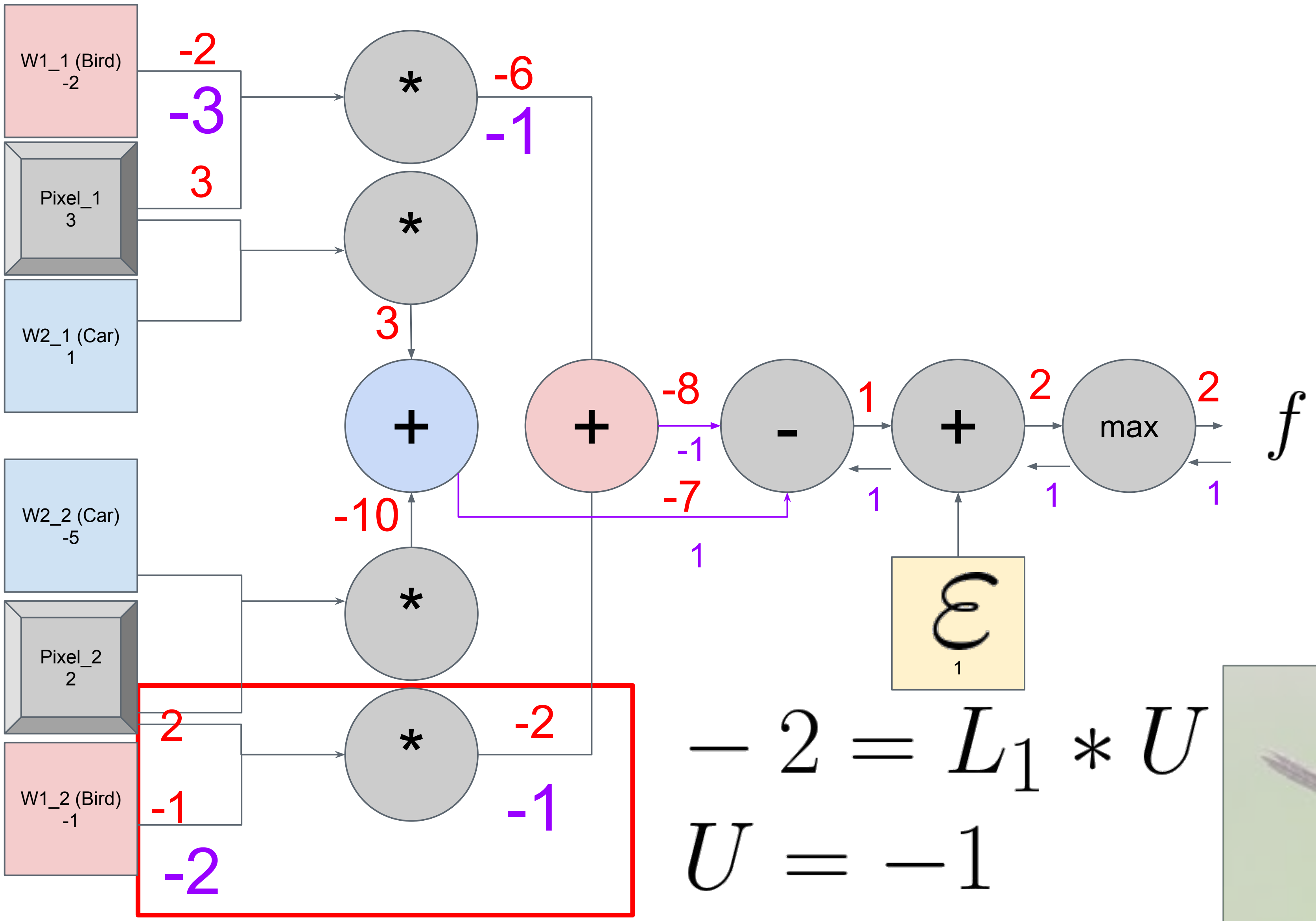
$$L_0 = 3$$



3
2


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
-2
-3
3
*
-6
-1
*
3
+
-10
+
-8
-1
-7
1
-
1
1
+
2
1
max
2
1
f
ε
1
2
-1
?
*
*
-2
-1

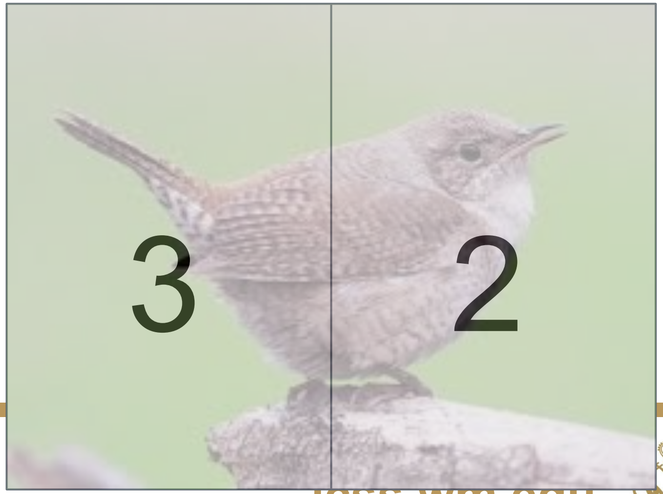

3
2

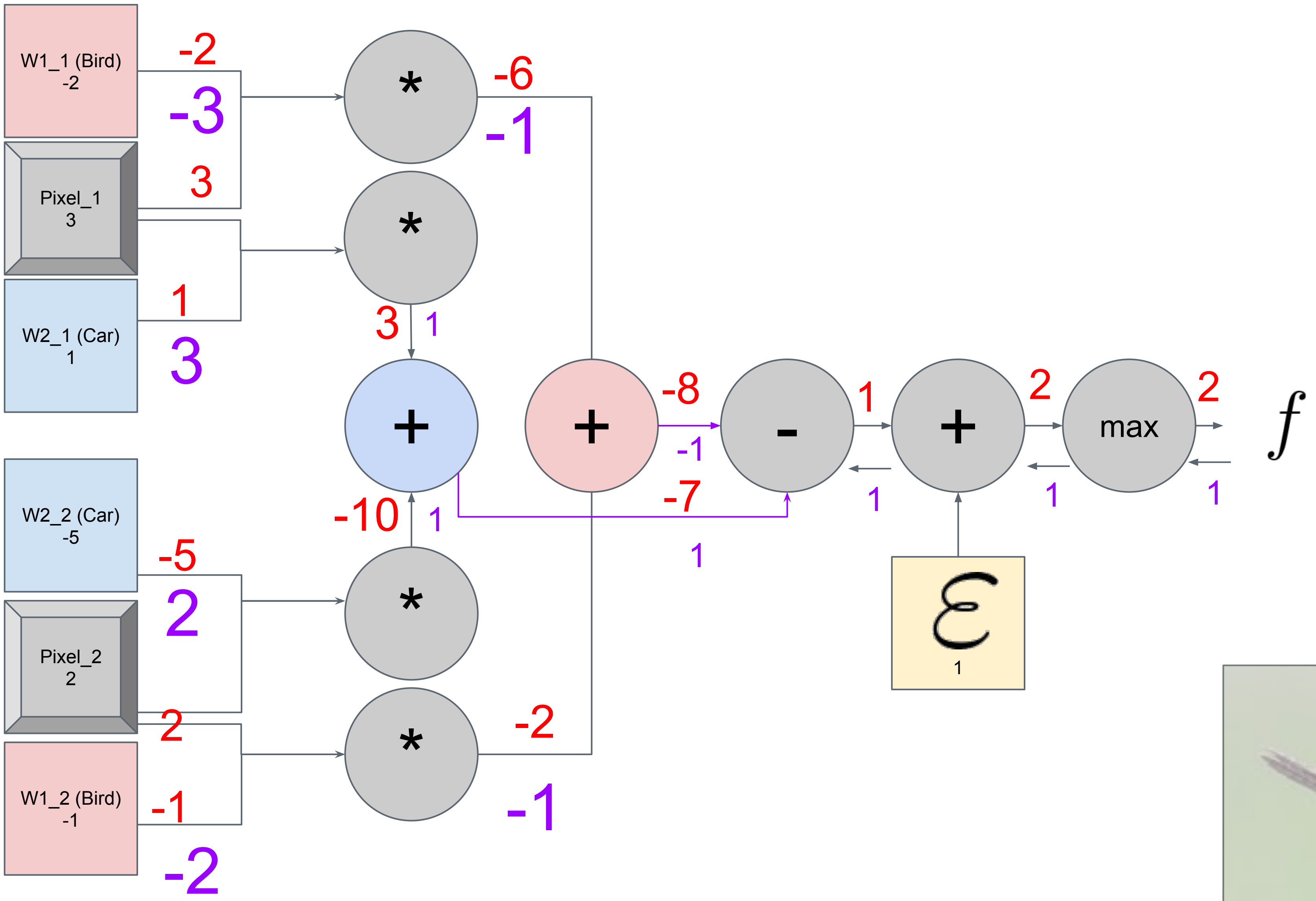

W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
-2
-3
3
*
-6
-1
*
3
+
-10
+
-8
-1
-7
1
-
1
1
+
2
1
max
2
1
f
ε
1
*
*
2
-1
-2
-2
-1

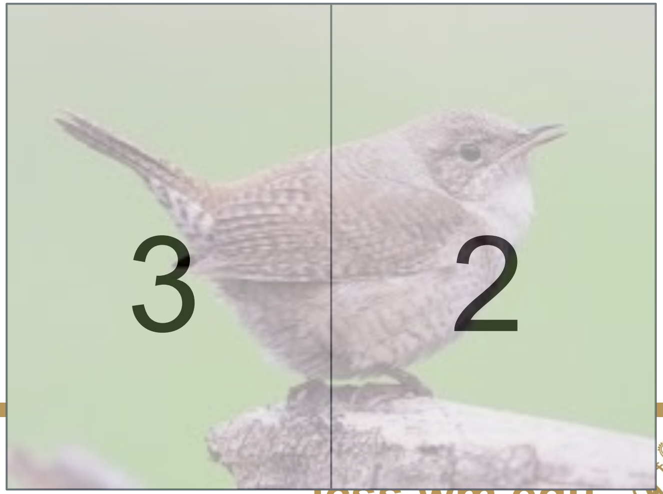
$$-2 = L_1 * U$$

$$U = -1$$

$$L_1 = 2$$


3
2


W1_1 (Bird)
-2
Pixel_1
3
W2_1 (Car)
1
W2_2 (Car)
-5
Pixel_2
2
W1_2 (Bird)
-1
-2
-3
3
1
3
-5
2
2
-1
-2
*
*
*
*
-6
-1
3
1
+
-10
1
-2
-1
+
-8
-1
-7
1
-
1
1
+
2
1
max
2
1
f
ε
1


3
2

Weights Gradient

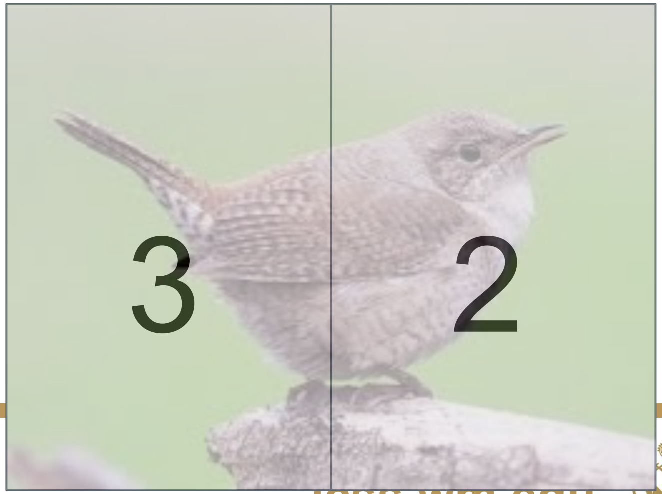
Weights	Gradient
W1_1 (Bird) -2	-3
W2_1 (Car) 1	3
W2_2 (Car) -5	2
W1_2 (Bird) -1	-2

For these weights, in our forward pass,

the **Car** score was **-7.**

The **bird** score was **-8.**

So in our classification, we guess Car. That isn't what we want (i.e., it's Bad).



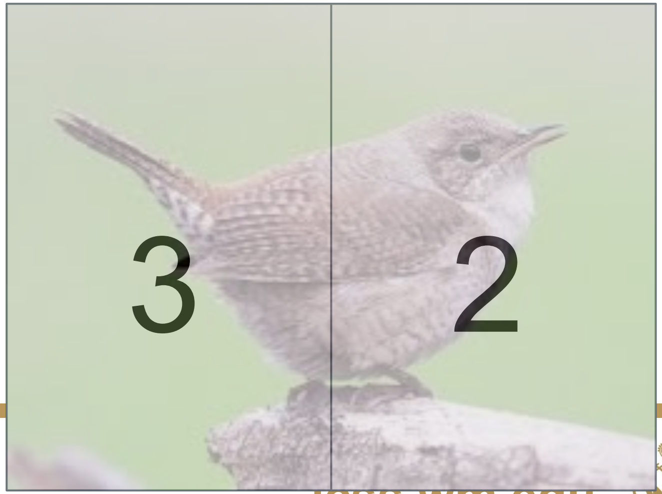
Weights	Gradient
W1_1 (Bird) -2	-3
W2_1 (Car) 1	3
W2_2 (Car) -5	2
W1_2 (Bird) -1	-2

For these weights, in our forward pass,

the **Car** score was **-7.**

The **bird** score was **-8.**

Our **loss** was **2.**



Weights	Gradient
W1_1 (Bird) -2	-3
W2_1 (Car) 1	3
W2_2 (Car) -5	2
W1_2 (Bird) -1	-2

For these weights, in our forward pass,

the **Car** score was **-7.**

The **bird** score was **-8.**

Our **loss** was **2.**

Weights	Gradient	Weights + (-1 * Gradient)
W1_1 (Bird) -2	-3	-2 + (-1 * -3) = -2 + 3 = 1
W2_1 (Car) 1	3	
W2_2 (Car) -5	2	
W1_2 (Bird) -1	-2	

For these weights, in our forward pass,

the **Car** score was **-7.**

The **bird** score was **-8.**

Our **loss** was **2.**

Weights	Gradient	Weights + (-1 * Gradient)
W1_1 (Bird) -2	-3	1
W2_1 (Car) 1	3	-2
W2_2 (Car) -5	2	-7
W1_2 (Bird) -1	-2	1

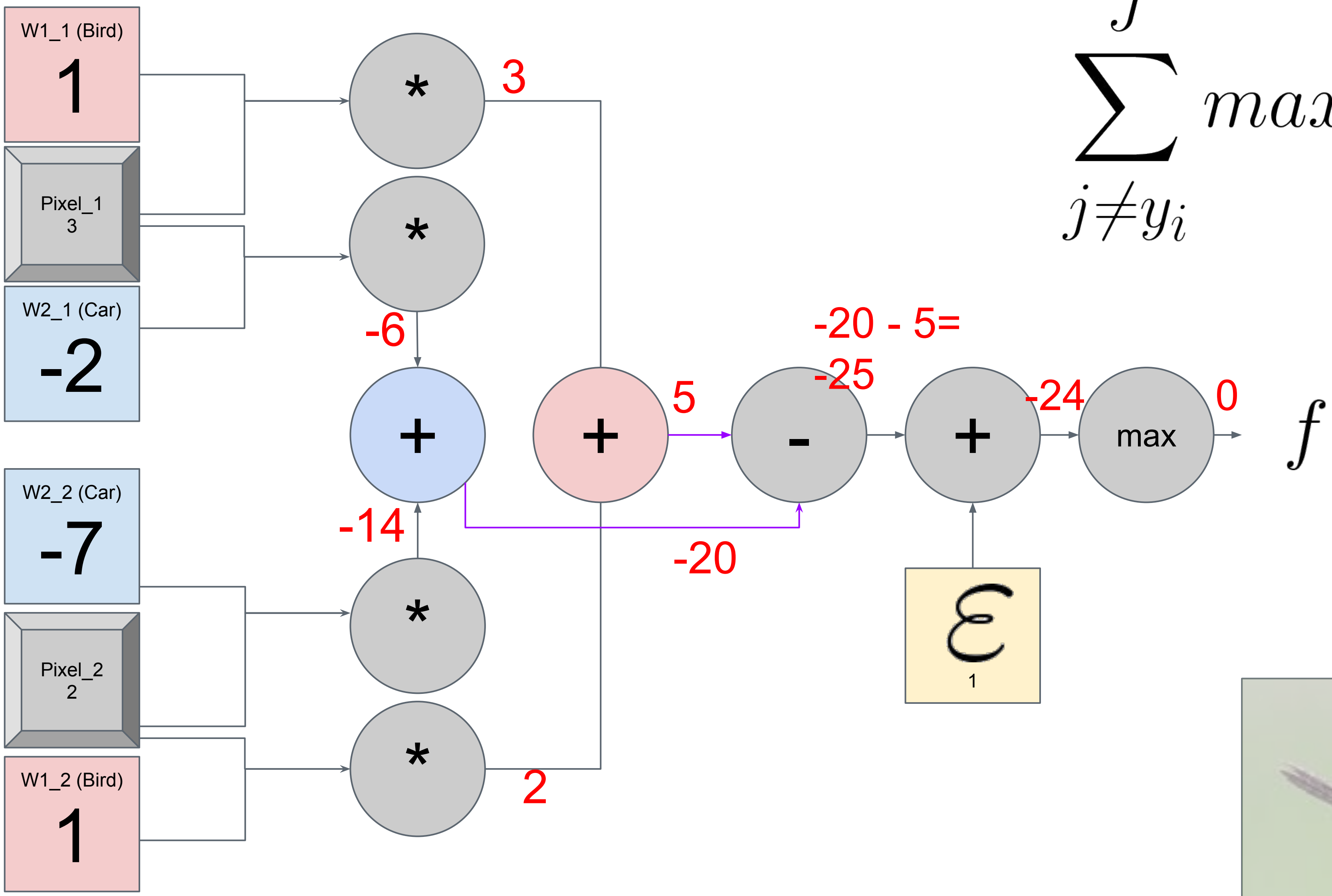
For these weights, in our forward pass,

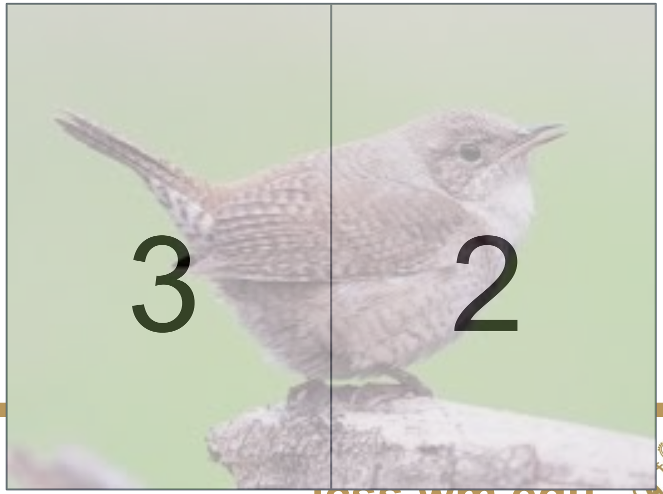
the **Car** score was **-7.**

The **bird** score was **-8.**

Our **loss** was **2.**

$$\sum_{j \neq y_i}^{J} max(0, s_j - s_{y_i} + \varepsilon)$$





Summary

- Computational Graphs
- Gradients & Partial Derivatives
- Backpropagation with a small example
- Next time: Matrix and vectorized backpropogation