Simple Neural Network

You are given the following neural network model parametrized by weight vector **w**. Model takes as a input vector **x** and outputs y:

$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$

Where:

$$\mathbf{x} = [2, 1],\ \mathbf{w} = [\pi/2, \pi],\ b = 0,\ \tilde{y} = 2$$

1) Draw a computational graph of forward pass of this small neural network
2) Compute feedforward pass with initial weights **w** and input data feature **x**
3) Calculate gradients of output y with respect to **w**, i. e $\frac{\partial y}{\partial \mathbf{w}}$
4) Use L_2 loss (Mean square error) to compute loss value between forward prediction y and label $\tilde{y}$. Add loss into computational graph.
5) Use chain rule to compute the gradient $\frac{\partial L}{\partial \mathbf{w}}$ and update weights with learning rate parameter α = 0.5

1) Draw computational graph

$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$

1) Draw computational graph

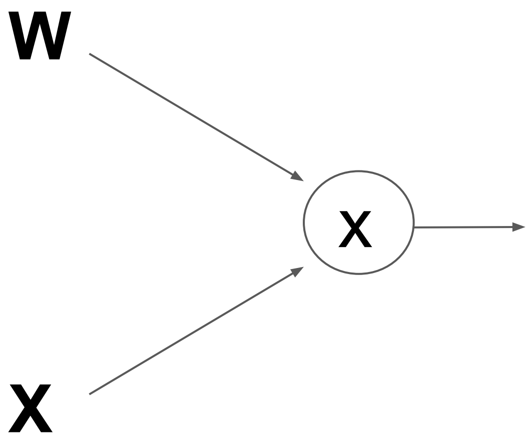
$$y = sin(\mathbf{w}^T\mathbf{x}) - b$$

w

x

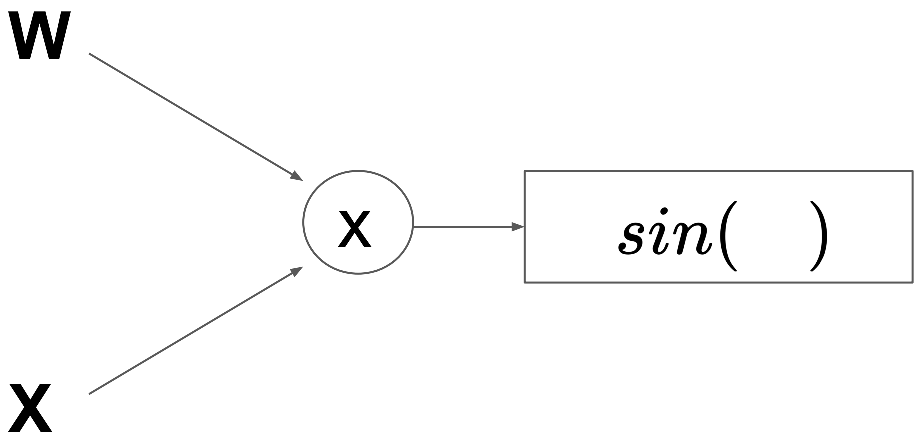
1) Draw computational graph

$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$



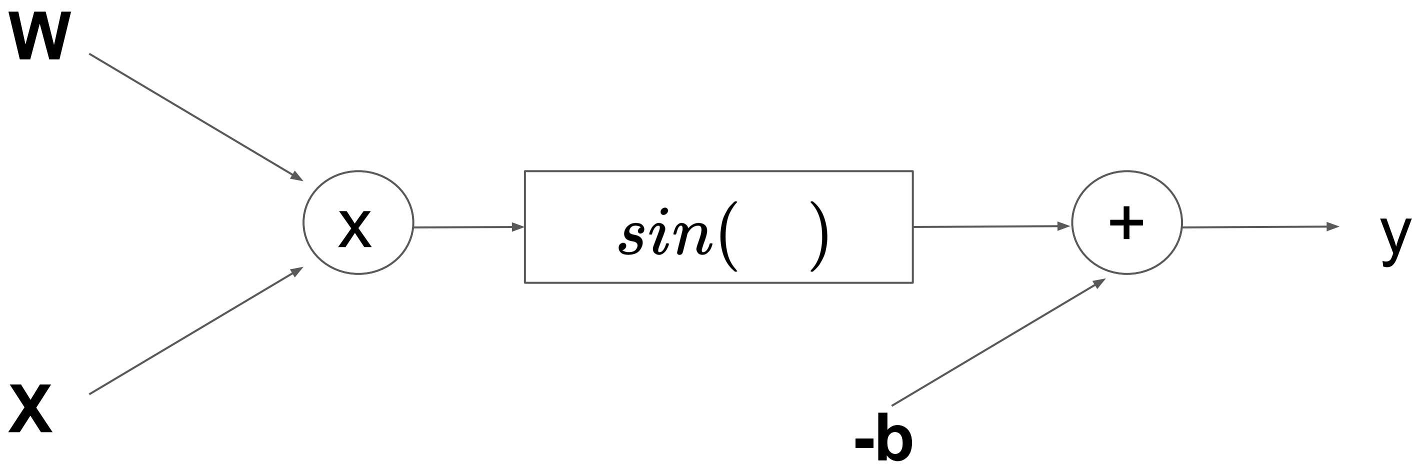
1) Draw computational graph

$$y = sin(\mathbf{w}^T\mathbf{x}) - b$$



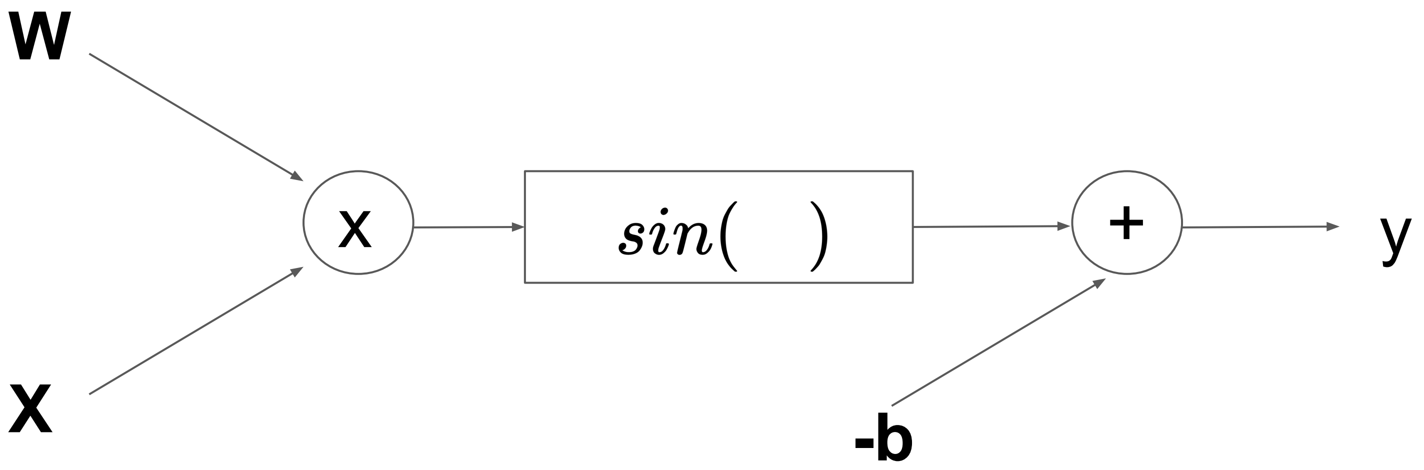
1) Draw computational graph

$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$



1) Draw computational graph

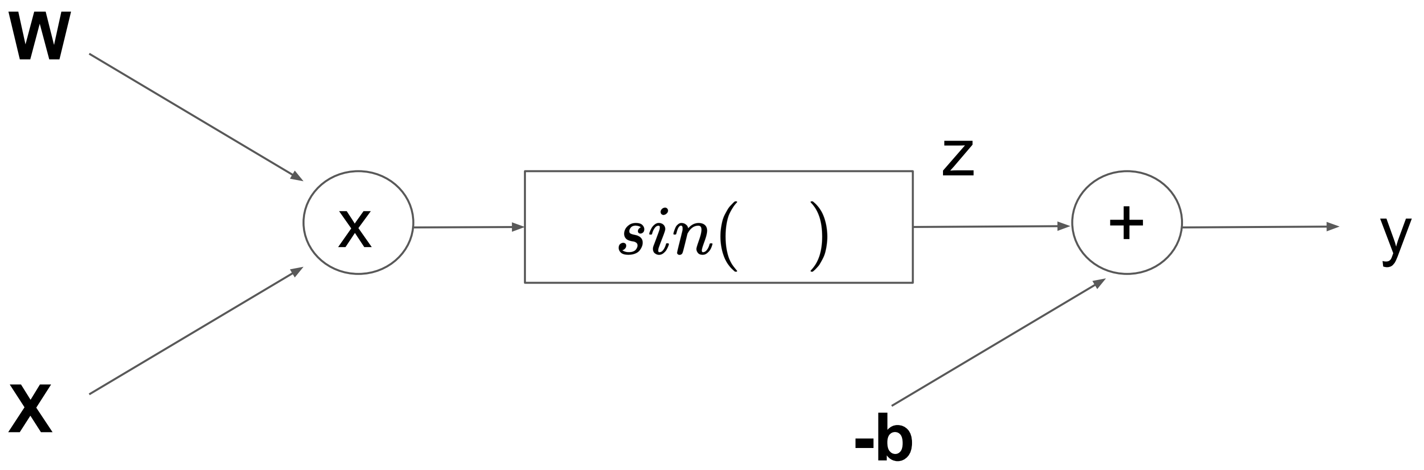
$$y = sin(\mathbf{w}^T\mathbf{x}) - b$$



2) Feedforward pass

1) Draw computational graph

$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$

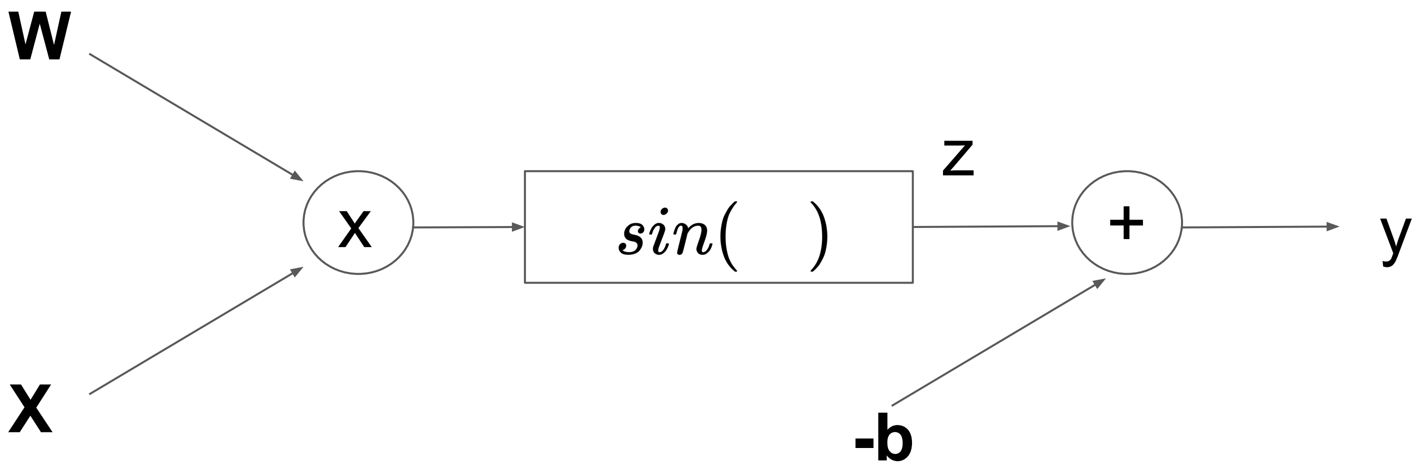


2) Feedforward pass

$$y = sin(\mathbf{w}^T \mathbf{x}) - b =$$

1) Draw computational graph

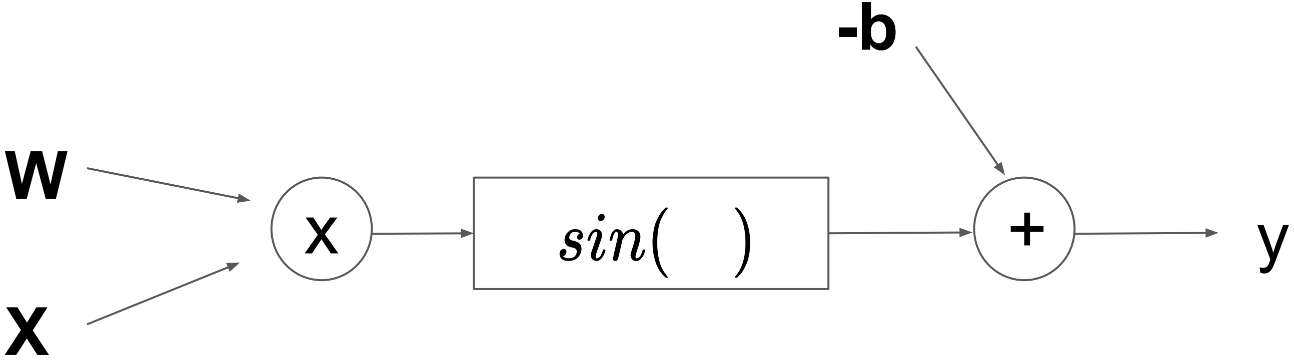
$$y = sin(\mathbf{w}^T \mathbf{x}) - b$$



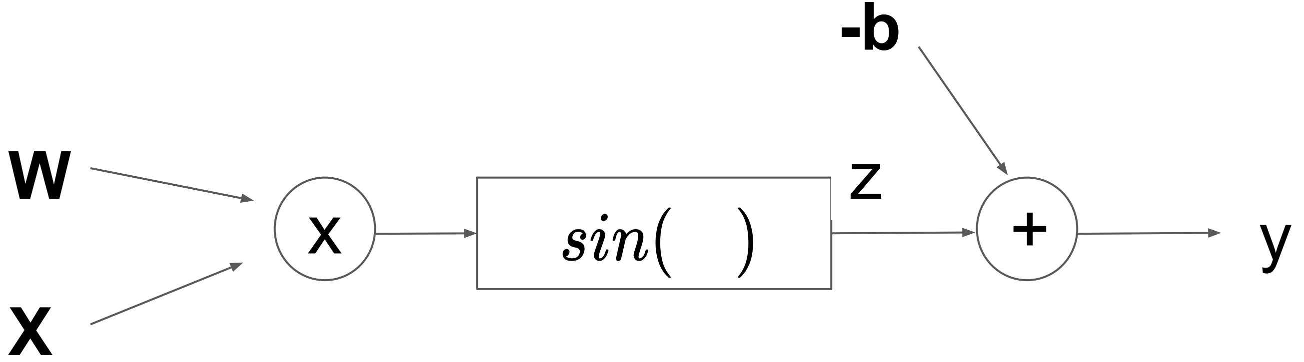
2) Feedforward pass

$$y = sin(\mathbf{w}^T \mathbf{x}) - b = sin((\pi/2 \quad \pi)\begin{pmatrix} 2 \\ 1 \end{pmatrix}) - 0 = 0$$

3) Gradients

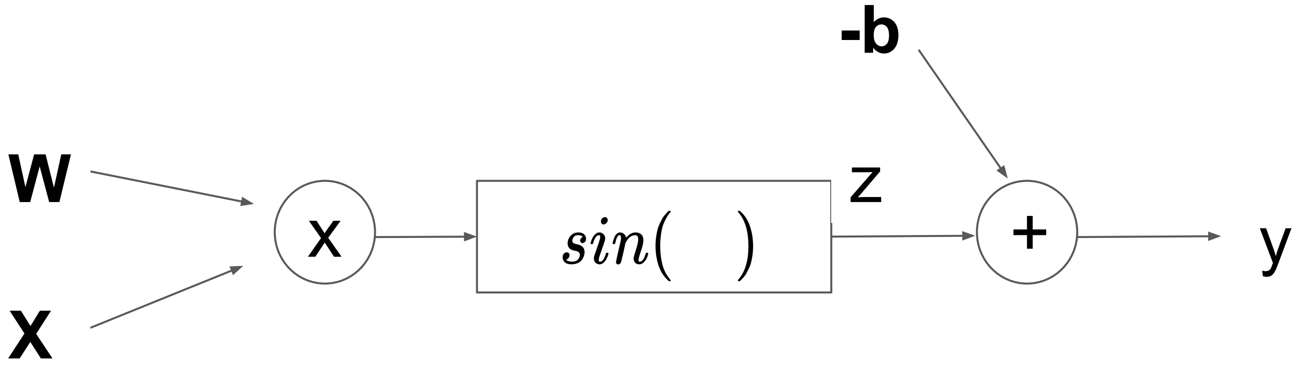


3) Gradients



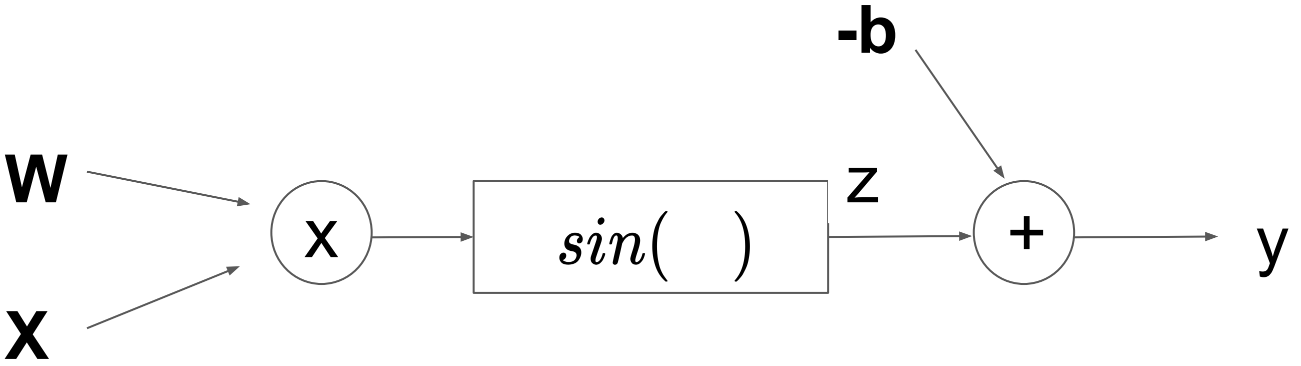
$$\frac{\partial y}{\partial \mathbf{w}}$$

3) Gradients



$$\frac{\partial y}{\partial \mathbf{w}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial (w^T x)} \frac{\partial (w^T x)}{\partial w}$$

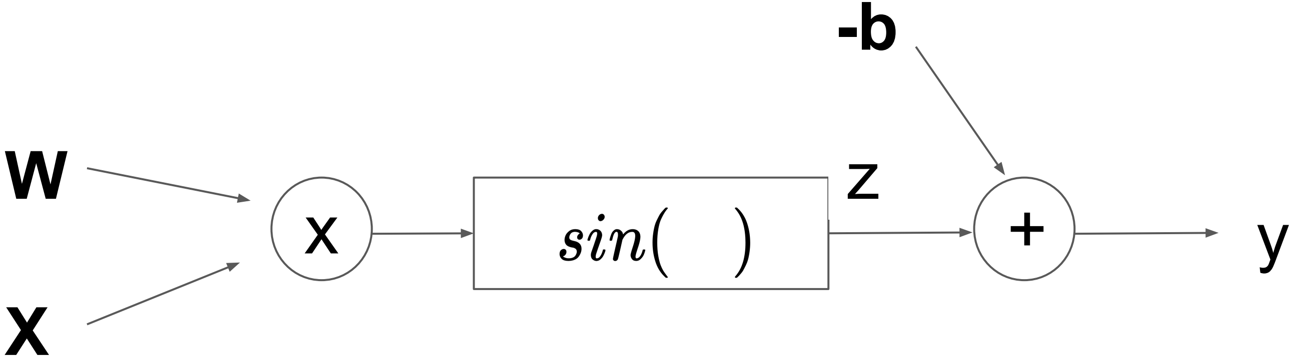
3) Gradients



$$\frac{\partial y}{\partial z} = \frac{\partial (z - b)}{\partial z} = 1$$

$$\frac{\partial y}{\partial \mathbf{w}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial (w^T x)} \frac{\partial (w^T x)}{\partial w}$$

3) Gradients

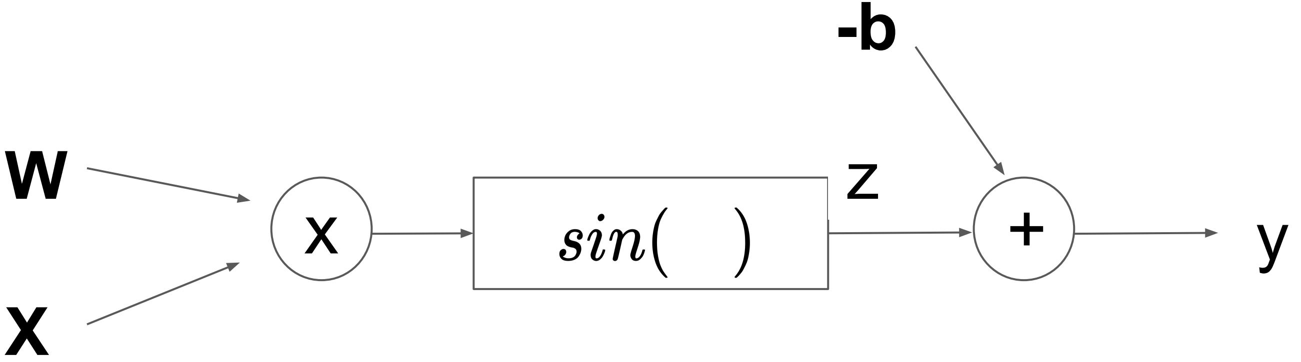


$$\frac{\partial y}{\partial z} = \frac{\partial (z - b)}{\partial z} = 1$$

$$\frac{\partial z}{\partial (w^T x)} = \frac{\partial sin(w^T x)}{\partial w^T x} = cos(w^T x)$$

$$\frac{\partial y}{\partial \mathbf{w}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial (w^T x)} \frac{\partial (w^T x)}{\partial w}$$

3) Gradients



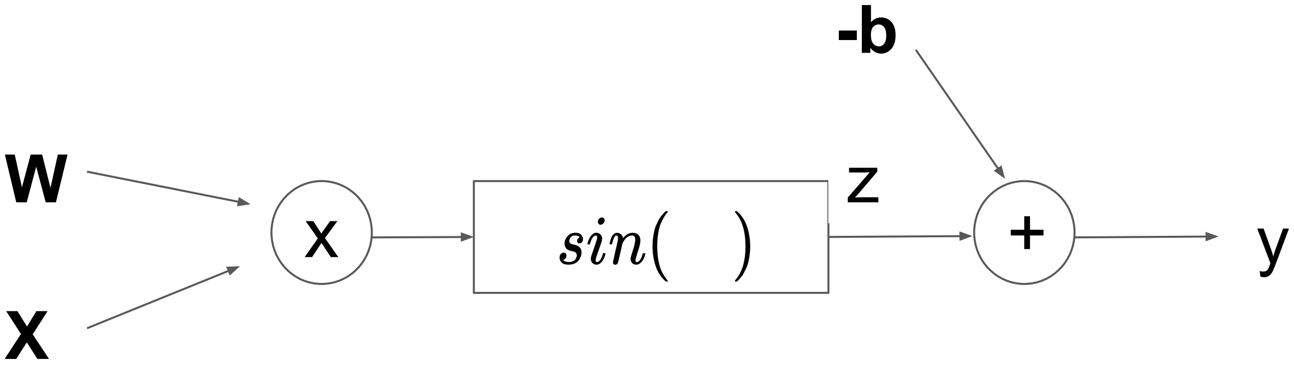
$$\frac{\partial y}{\partial z} = \frac{\partial (z - b)}{\partial z} = 1$$

$$\frac{\partial z}{\partial (w^T x)} = \frac{\partial sin(w^T x)}{\partial w^T x} = cos(w^T x)$$

$$\frac{\partial (w^T x)}{\partial w} = x$$

$$\frac{\partial y}{\partial \mathbf{w}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial (w^T x)} \frac{\partial (w^T x)}{\partial w}$$

3) Gradients



$$\frac{\partial y}{\partial z} = \frac{\partial (z - b)}{\partial z} = 1$$

$$\frac{\partial z}{\partial (w^T x)} = \frac{\partial sin(w^T x)}{\partial w^T x} = cos(w^T x)$$

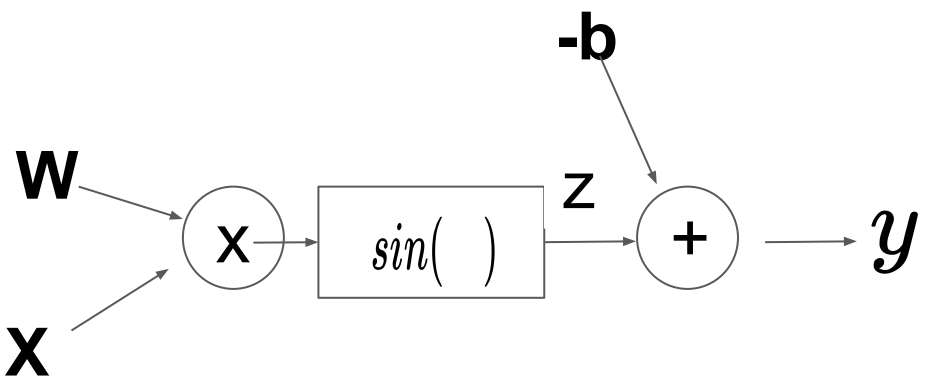
$$\frac{\partial (w^T x)}{\partial w} = x$$

- Use Chain rule

$$\frac{\partial y}{\partial \mathbf{w}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial (w^T x)} \frac{\partial (w^T x)}{\partial w} = 1 * cos(w^T x) * x = cos(2\pi) * x = x$$

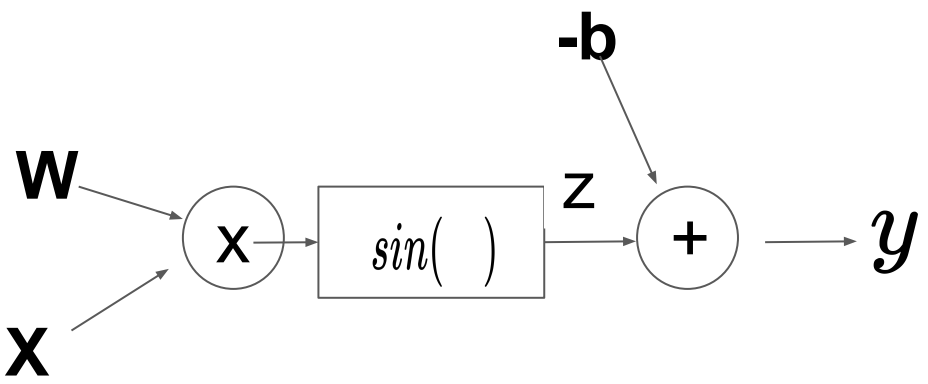
4) Compute L2 loss from prediction and label. Add loss into computational graph

$L_2\ loss =$



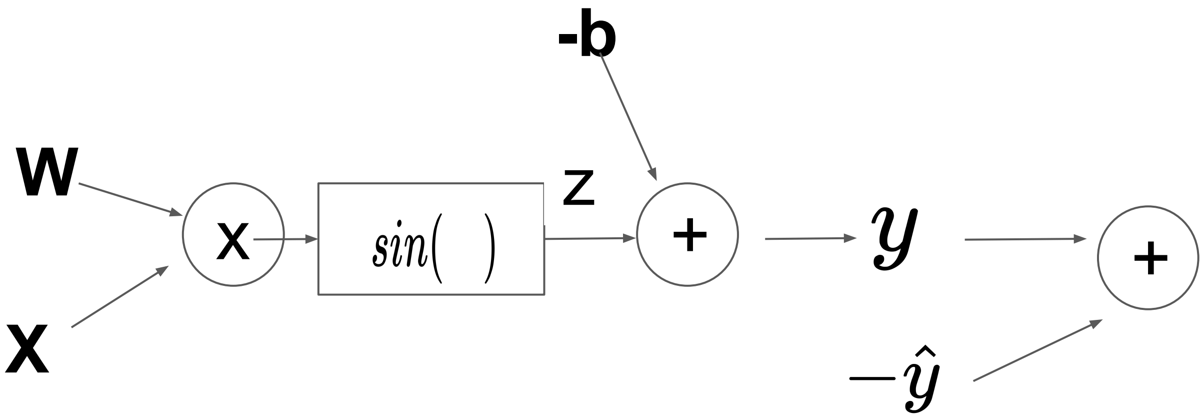
4) Compute L2 loss from prediction and label. Add loss into computational graph

$$L_2\ loss = \ ||y - \hat{y}||^2 \ = (0-2)^2 = 4$$



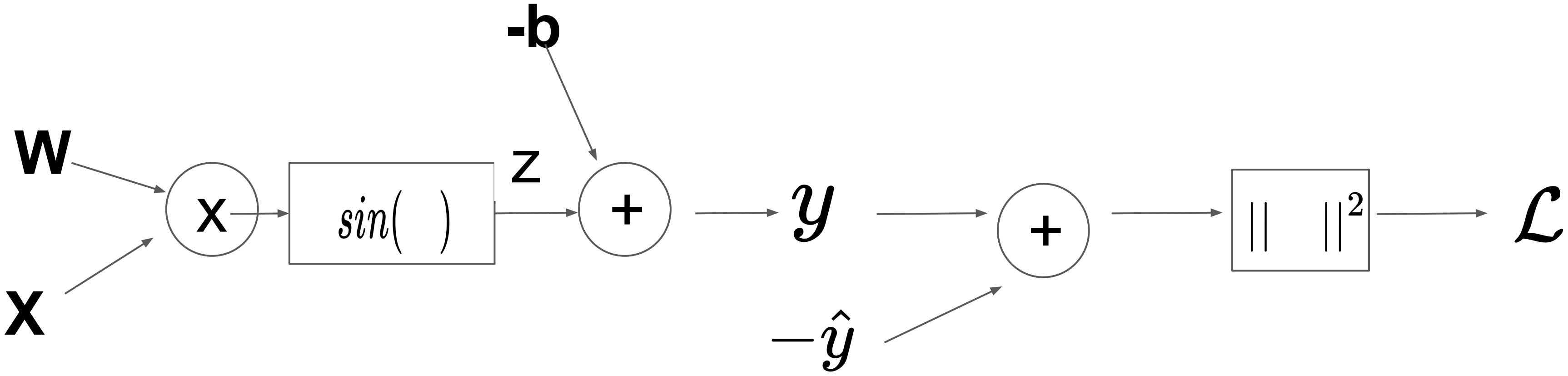
4) Compute L2 loss from prediction and label. Add loss into computational graph

$$L_2\ loss = \ ||y - \hat{y}||^2 \ = (0-2)^2 \ = 4$$



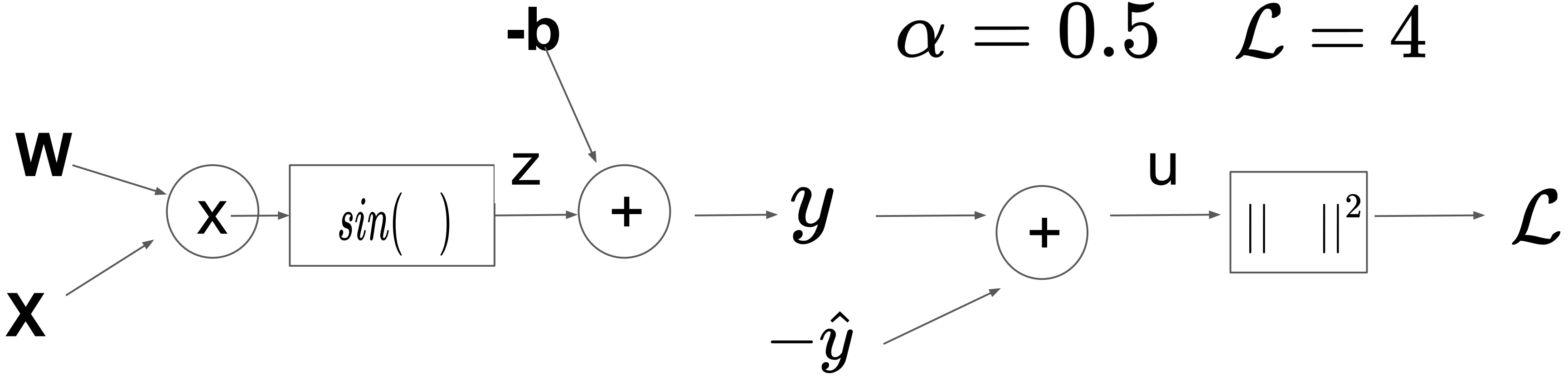
4) Compute L2 loss from prediction and label. Add loss into computational graph

$$L_2\ loss = \ ||y - \hat{y}||^2 \ = (0-2)^2 \ = 4$$



5) Compute loss gradients. Update weights.

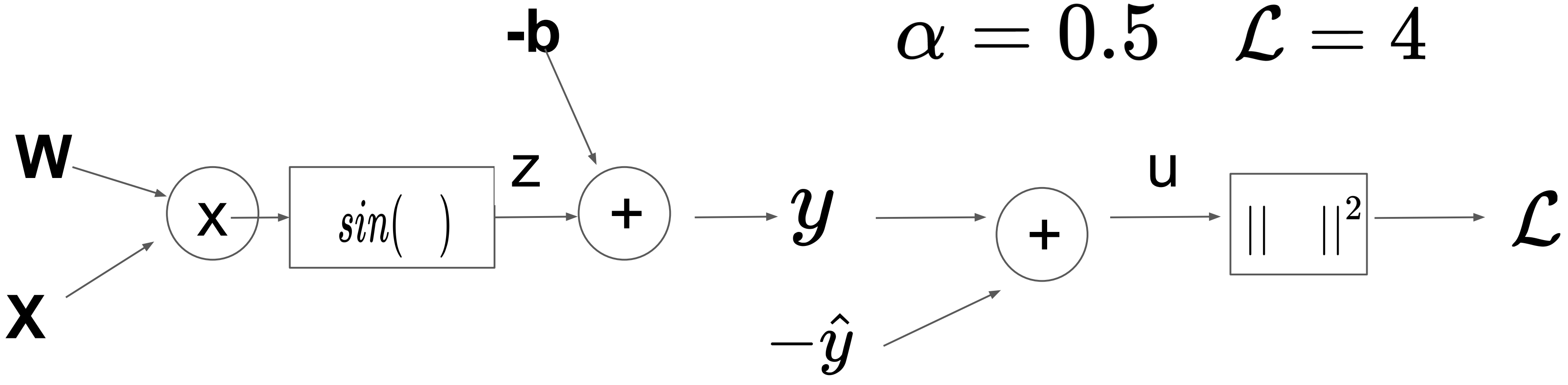
$$\alpha = 0.5 \quad \mathcal{L} = 4$$



$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} =$$

5) Compute loss gradients. Update weights.

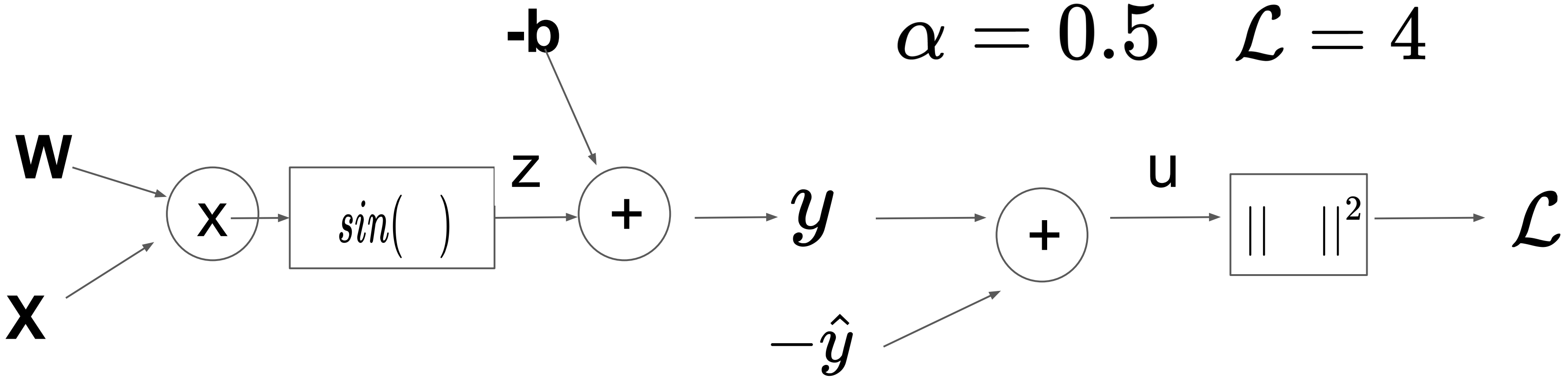
$$\alpha = 0.5 \quad \mathcal{L} = 4$$



$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial u} \frac{\partial u}{\partial y} \frac{\partial y}{\partial z} \frac{\partial z}{\partial w^T x} \frac{\partial w^T x}{\partial w} =$$

5) Compute loss gradients. Update weights.

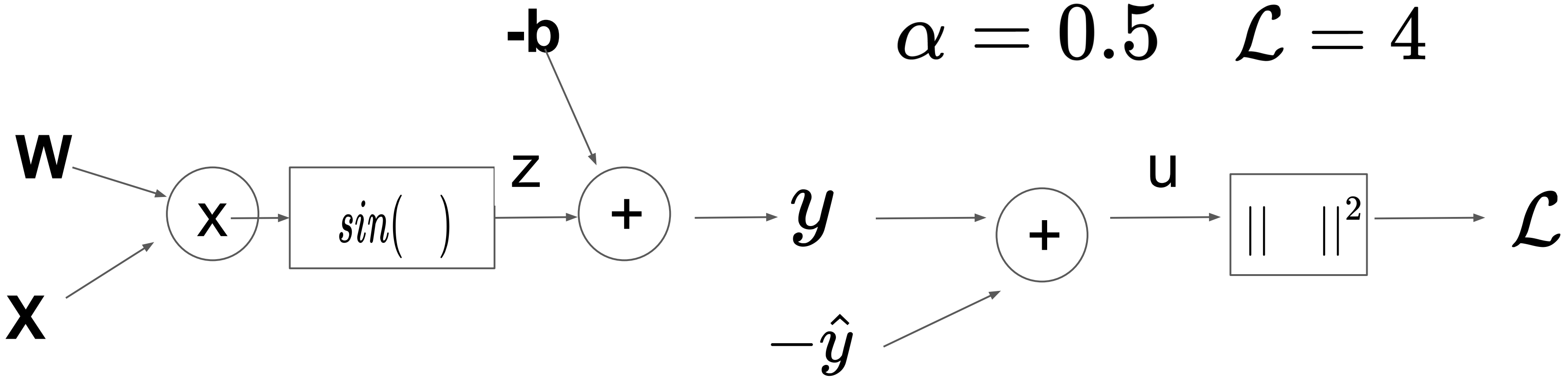
$$\alpha = 0.5 \quad \mathcal{L} = 4$$



$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial w^T x}\frac{\partial w^T x}{\partial w} =$$

5) Compute loss gradients. Update weights.

$$\alpha = 0.5 \quad \mathcal{L} = 4$$

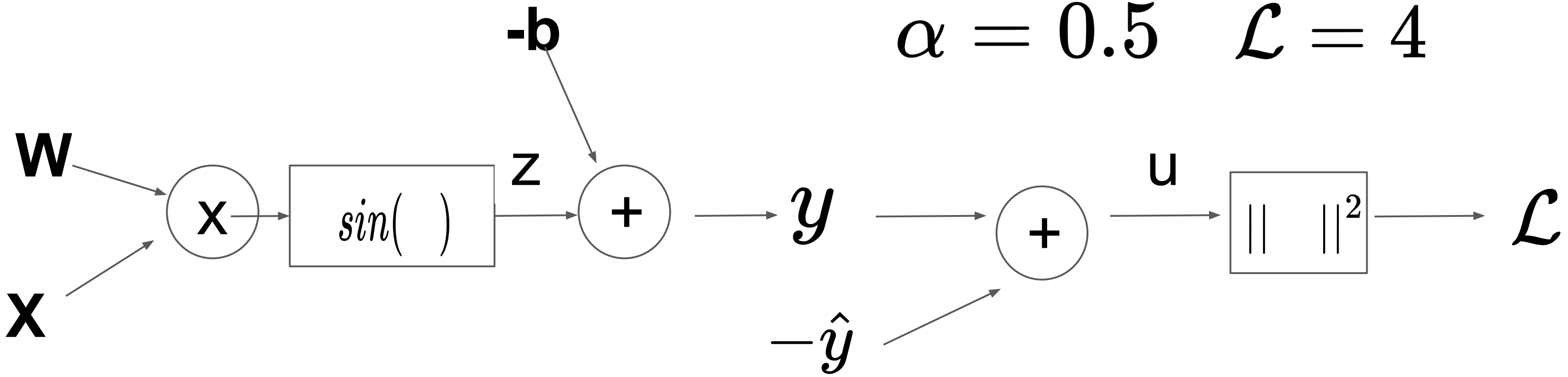


$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial w^T x}\frac{\partial w^T x}{\partial w} =$$

$$\frac{\partial u}{\partial y} = \frac{\partial (y - \hat{y})}{\partial y} = 1$$

5) Compute loss gradients. Update weights.

$$\alpha = 0.5 \quad \mathcal{L} = 4$$



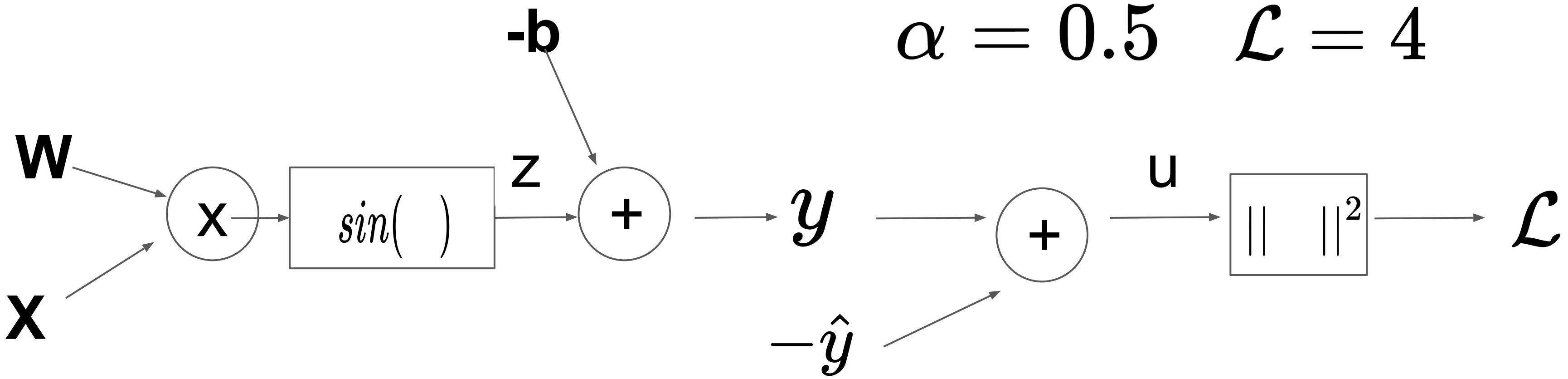
$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial w^T x}\frac{\partial w^T x}{\partial w} = 2(y - \hat{y}) * 1 * x = -4x$$

$$\frac{\partial u}{\partial y} = \frac{\partial (y - \hat{y})}{\partial y} = 1$$

$$\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial ||y - \hat{y}||^2}{\partial ||y - \hat{y}||} = 2(y - \hat{y})$$

5) Compute loss gradients. Update weights.

$$\alpha = 0.5 \quad \mathcal{L} = 4$$



$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial w^T x}\frac{\partial w^T x}{\partial w} = 2(y-\hat{y}) * 1 * x = -4x$$

$$\frac{\partial u}{\partial y} = \frac{\partial (y-\hat{y})}{\partial y} = 1$$

$$\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial ||y-\hat{y}||^2}{\partial ||y-\hat{y}||} = 2(y-\hat{y})$$

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \begin{pmatrix} \pi/2 + 4 \\ \pi + 2 \end{pmatrix}$$

Convolutional Layer

You are given input feature map **x** and kernel **w**:

$$\mathbf{x} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & -1 \\ 0 & 0 & 2 \end{pmatrix}, \mathbf{w} = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$$

Stride denotes length of convolutional stride, padding denotes symetric zero-padding.
Compute outputs of following layers:

$$1)\ conv(\ \mathbf{x}, \mathbf{w},\ stride = 1,\ padding = 0)\ =$$

$$2)\ conv(\ \mathbf{x}, \mathbf{w},\ stride = 3,\ padding = 1)\ =$$

$$3)\ max(\ \mathbf{x},\ 2\ x\ 2)\ =$$

Convolution

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 1,\ \ pad = 0)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

Convolution

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 1,\ \ pad = 0)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$$z_{11} = sum(\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 3$$

Convolution

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 1,\ \ pad = 0)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$$z_{11} = sum(\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 3 \qquad z_{12} = sum(\begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = -4$$

Convolution

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 1,\ \ pad = 0)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$$z_{11} = sum(\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 3 \qquad z_{12} = sum(\begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = -4$$

$$z_{21} = sum(\begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 1 \qquad z_{22} = sum(\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 6$$

Convolution

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 1,\ \ pad = 0)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{z} =$

3	-4
1	6

$$z_{11} = sum(\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 3 \qquad z_{12} = sum(\begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = -4$$

$$z_{21} = sum(\begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 1 \qquad z_{22} = sum(\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}) = 6$$

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 3,\ \ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 3,\ \ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 3,\ \ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 3,\ \ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ \ stride = 3,\ \ padding = 1)$$

$$\mathbf{x} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{x}_{padding} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 2 & 1 & -1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \mathbf{z} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ stride = 3,\ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

$\mathbf{z} =$

2	0
0	1

$$\mathbf{p} = max(\ \mathbf{x},\ 2\ x\ 2)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$$\mathbf{z} = conv(\ \mathbf{x}, \mathbf{w},\ stride = 3,\ padding = 1)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

$\mathbf{w} =$

1	-1
0	2

$\mathbf{x}_{padding} =$

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

→ $\mathbf{z} =$

2	0
0	1

$$\mathbf{p} = max(\ \mathbf{x},\ 2\ x\ 2)$$

$\mathbf{x} =$

1	0	2
2	1	-1
0	0	2

→ $\mathbf{p} =$

2	2
2	2