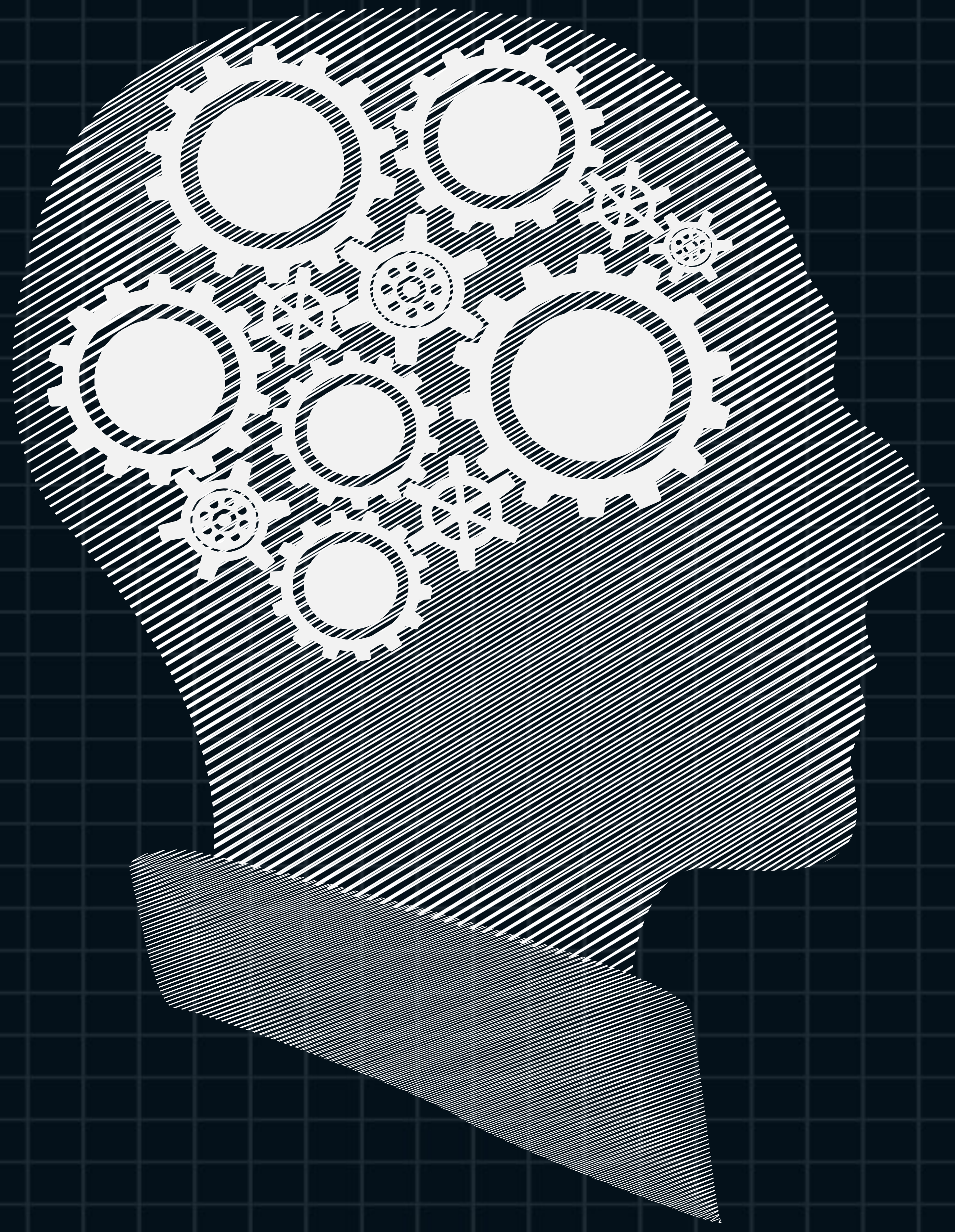


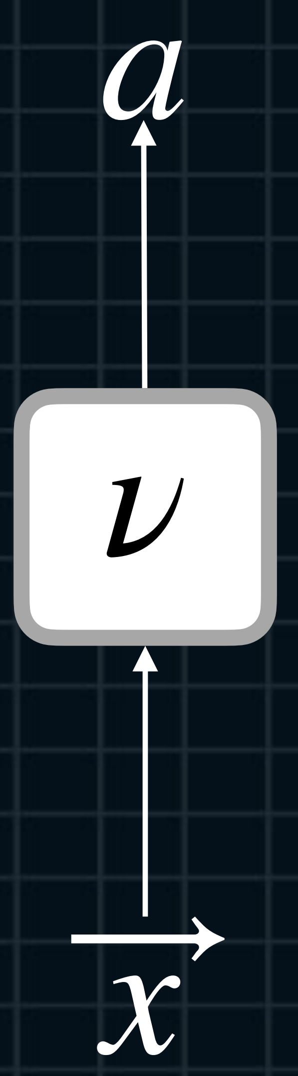
# Backpropagation and Jacobian Matrices

## Lecture.1
## Why Backpropagation and Jacobians?

# - Trainable Models and Params

## Artificial Neurons

$$\hat{y} = xw + b \qquad \hat{y} = \vec{x}^T \cdot \vec{w} + b \qquad \hat{y} = g\left(\vec{x}^T \cdot \vec{w} + b\right)$$



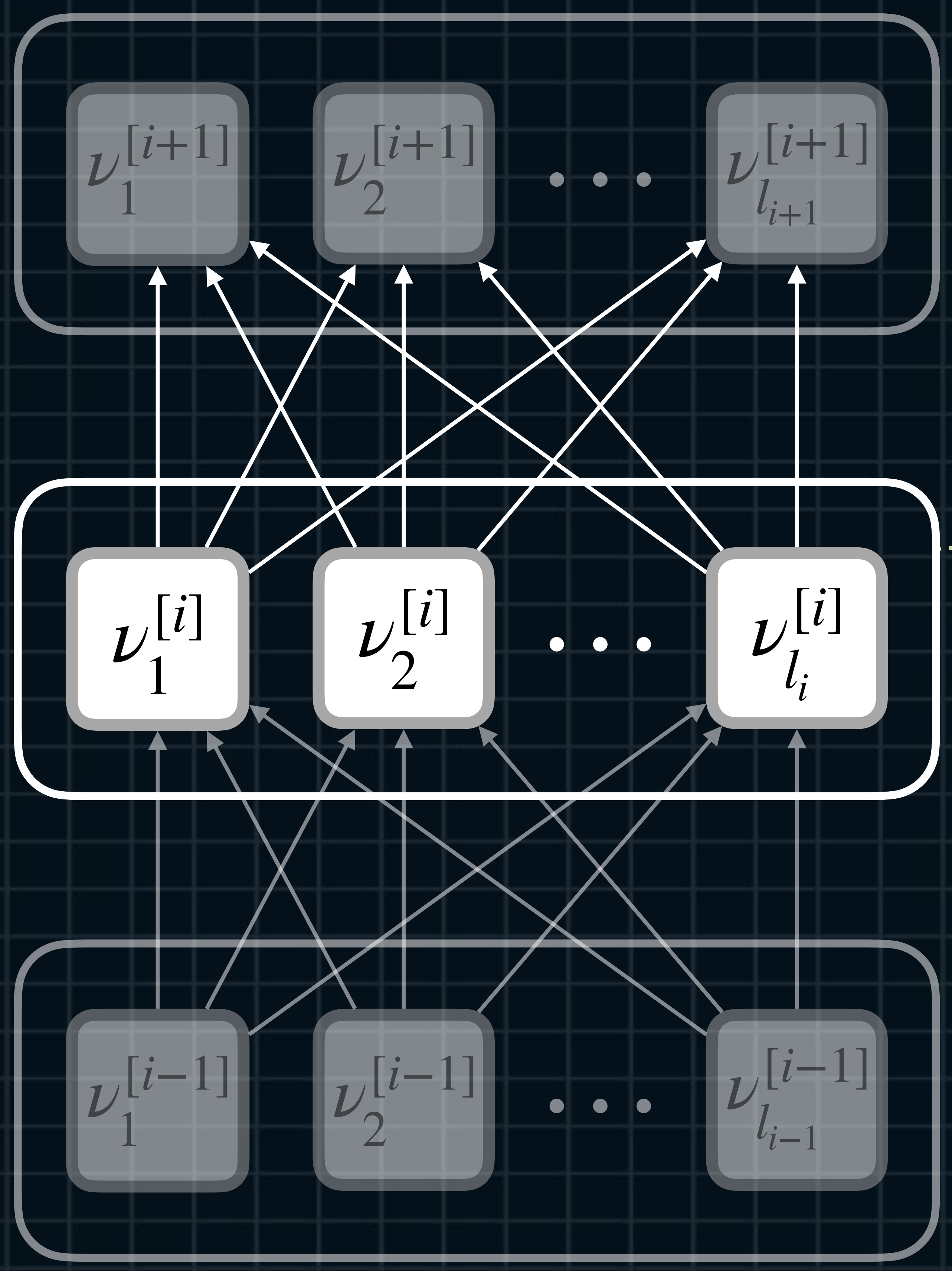

$$\nu\left(\vec{x}; \vec{w}, b\right) = g\left((\vec{x})^T \vec{w} + b\right)$$

$\vec{w} \in \mathbb{R}^{l_l \times 1}$

$b \in \mathbb{R}$

# - Trainable Models and Params

## Dense Layers



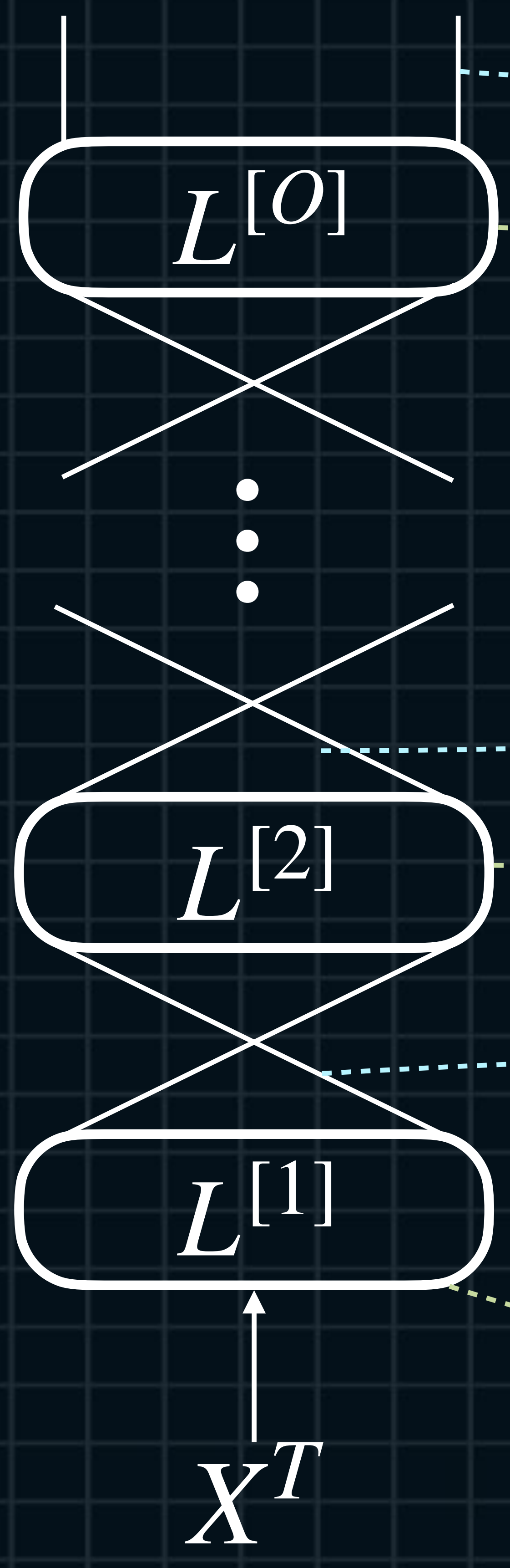

$$W^{[i]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{i-1} \times l_i}$$

$$(\vec{b}^{[i]})^T = \begin{pmatrix} b_1^{[i]} & b_2^{[i]} & \dots & b_{l_i}^{[i]} \end{pmatrix} \in \mathbb{R}^{1 \times l_i}$$

Lecture.1
Why Backpropagation
and Jacobians?

# - Trainable Models and Params

## Dense Layers




$$W^{[O]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[O]} & \vec{w}_2^{[O]} & \cdots & \vec{w}_{l_O}^{[O]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{O-1} \times l_O}$$

$$(\vec{b}^{[O]})^T = \begin{pmatrix} b_1^{[O]} & b_2^{[O]} & \dots & b_{l_O}^{[O]} \end{pmatrix} \in \mathbb{R}^{1 \times l_O}$$

$$\underset{\mathbb{R}^{N \times l_O}}{(A^{[O]})^T} = \underset{\mathbb{R}^{N \times l_{O-1}}}{(A^{[O-1]})^T} \underset{\mathbb{R}^{l_{O-1} \times l_O}}{W^{[O]}} + \underset{\mathbb{R}^{1 \times l_O}}{(\vec{b}^{[O]})^T}$$

$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_1 \times l_2}$$

$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \dots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1 \times l_2}$$

$$\underset{\mathbb{R}^{N \times l_2}}{(A^{[2]})^T} = \underset{\mathbb{R}^{N \times l_1}}{(A^{[1]})^T} \underset{\mathbb{R}^{l_1 \times l_2}}{W^{[2]}} + \underset{\mathbb{R}^{1 \times l_2}}{(\vec{b}^{[2]})^T}$$

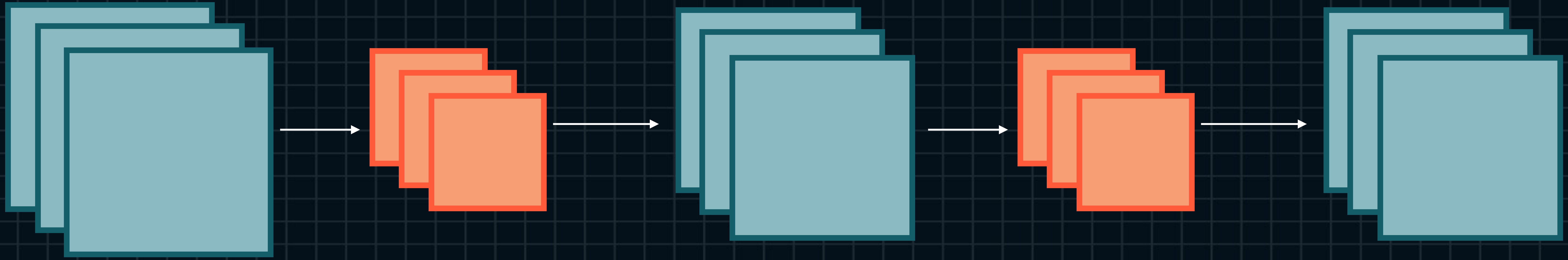
$$\underset{\mathbb{R}^{N \times l_1}}{(A^{[1]})^T} = \underset{\mathbb{R}^{N \times l_I}}{X^T} \underset{\mathbb{R}^{l_I \times l_1}}{W^{[1]}} + \underset{\mathbb{R}^{1 \times l_1}}{(\vec{b}^{[1]})^T}$$

$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

$$(\vec{b}^{[1]})^T = \begin{pmatrix} b_1^{[1]} & b_2^{[1]} & \dots & b_{l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{1 \times l_1}$$

# - Trainable Models and Params

## Conv Layers



$X \in \mathbb{R}^{n_H^{[I]} \times n_W^{[I]} \times l_I}$

$\mathscr{K}^{[1]} \in \mathbb{R}^{f^{[1]} \times f^{[1]} \times l_I \times l_1}$

$\overrightarrow{b}^{[1]} \in \mathbb{R}^{l_1}$

$A^{[1]} \in \mathbb{R}^{n_H^{[1]} \times n_W^{[1]} \times l_1}$

$\mathscr{K}^{[2]} \in \mathbb{R}^{f^{[2]} \times f^{[2]} \times l_1 \times l_2}$

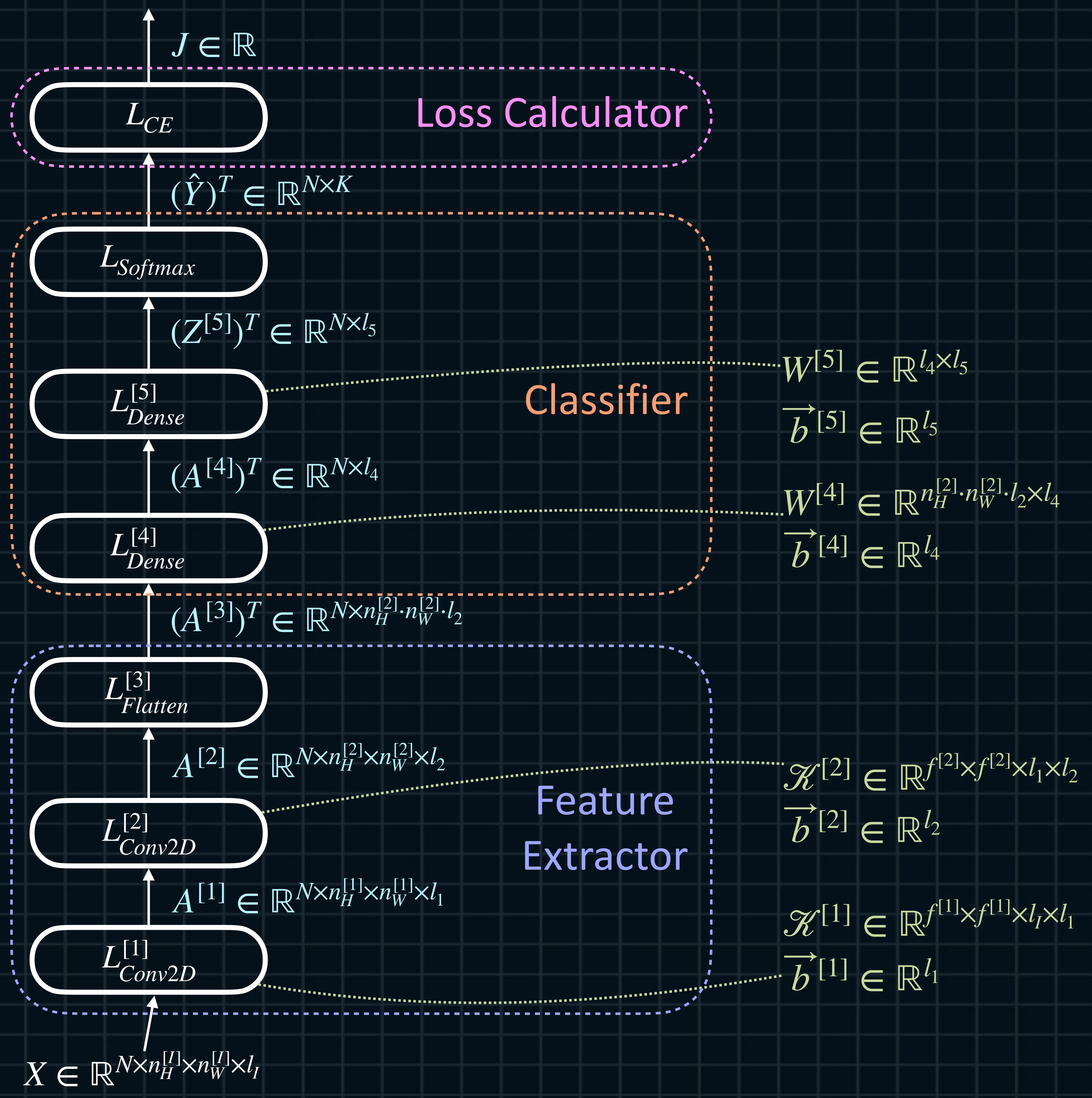
$\overrightarrow{b}^{[2]} \in \mathbb{R}^{l_2}$

$A^{[2]} \in \mathbb{R}^{n_H^{[2]} \times n_W^{[2]} \times l_2}$

Lecture.1
Why Backpropagation
and Jacobians?

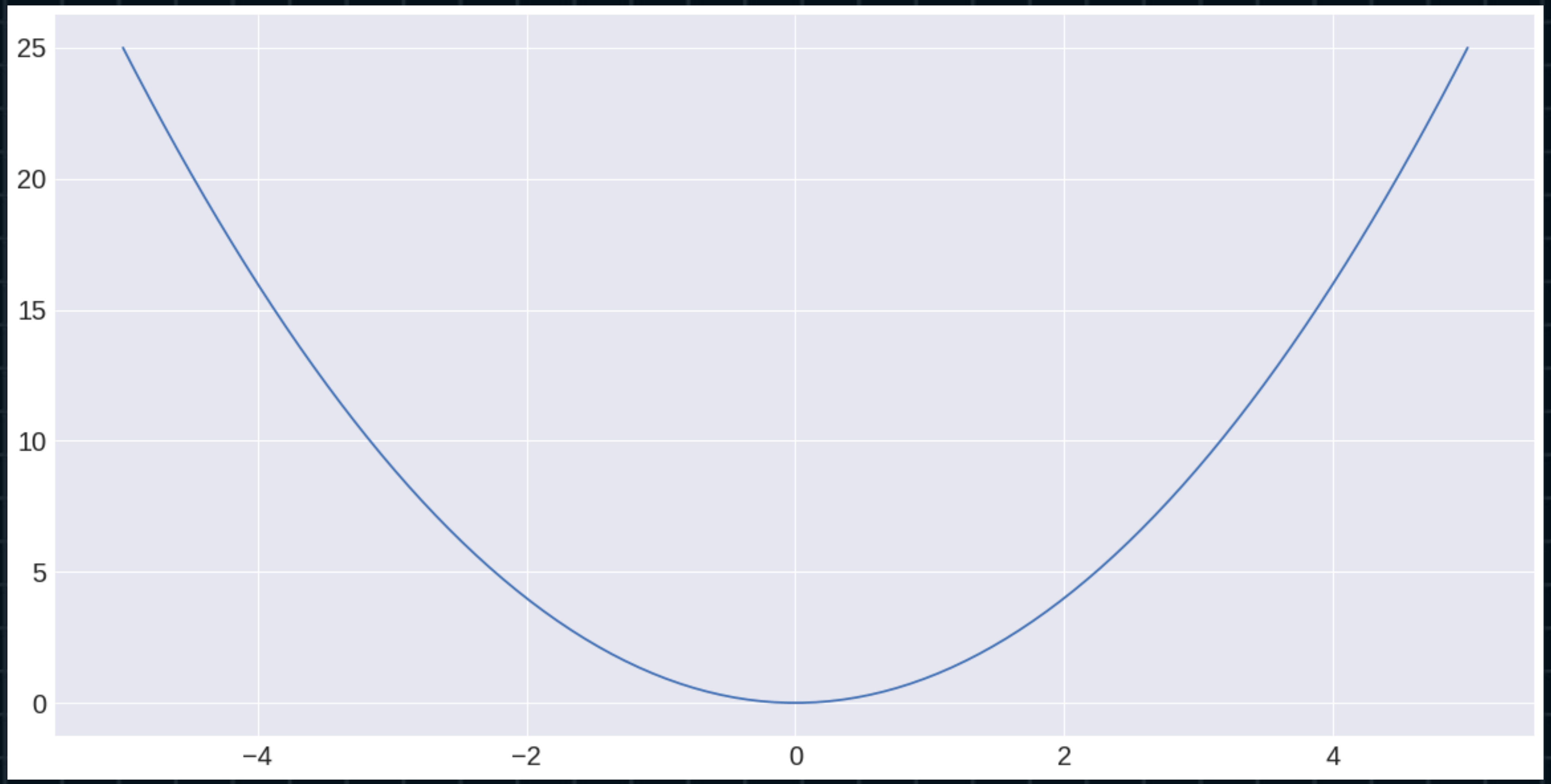
# - Trainable Models and Params

CNNs



# - Gradient-based Learning

## Differential Coefficient in DL




$$y' = 2x$$

$$y'|_{x=2} = 2 \cdot 2 = +4$$

$$y'|_{x=-2} = 2 \cdot (-2) = -4$$

$$y' = 2x$$

$$y'|_{x=1} = 2 \cdot 1 = +2$$

$$y'|_{x=2} = 2 \cdot 2 = +4$$

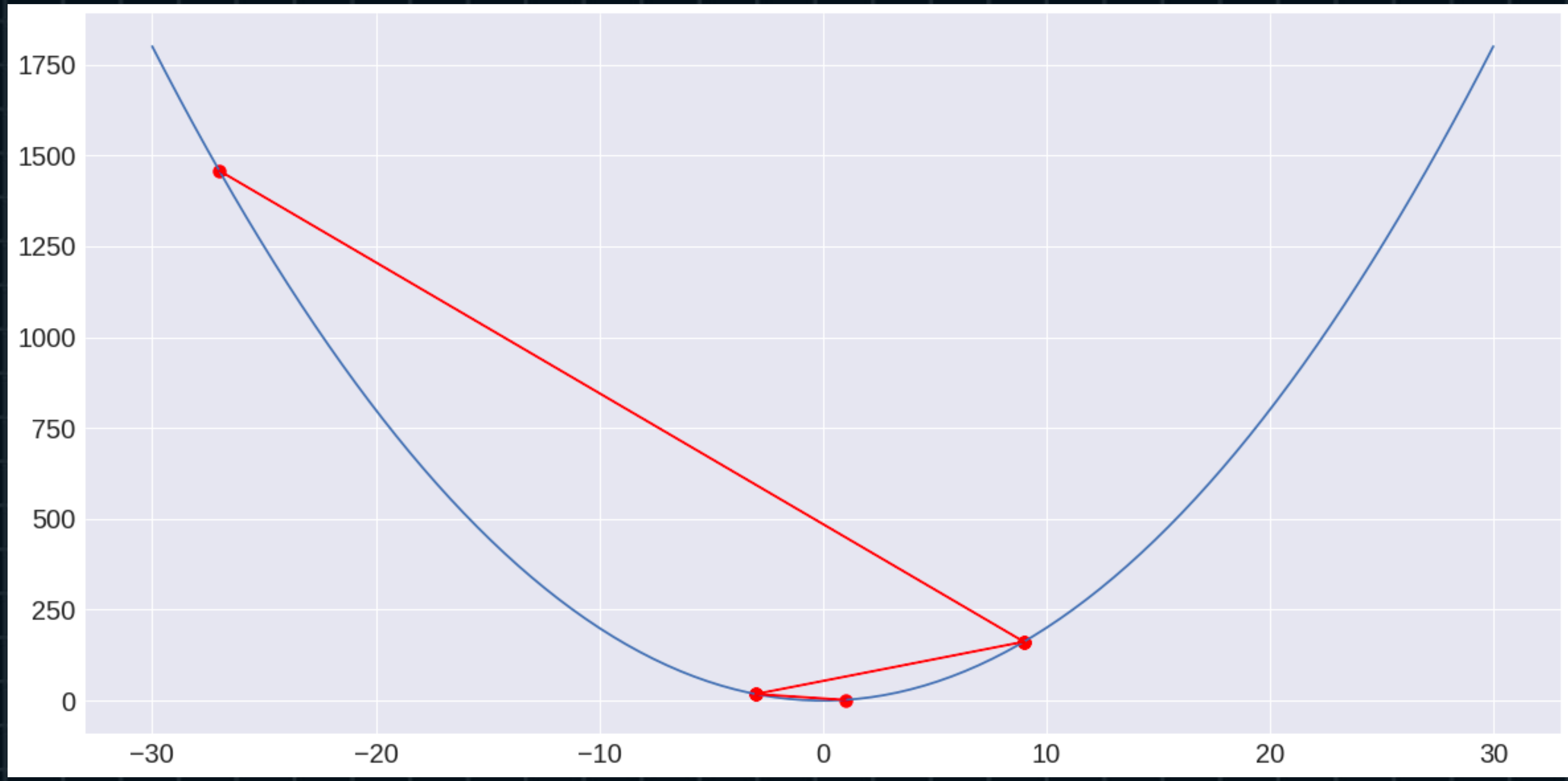
## - Gradient-based Learning

Update Notation

$$x := x + a$$

# - Gradient-based Learning

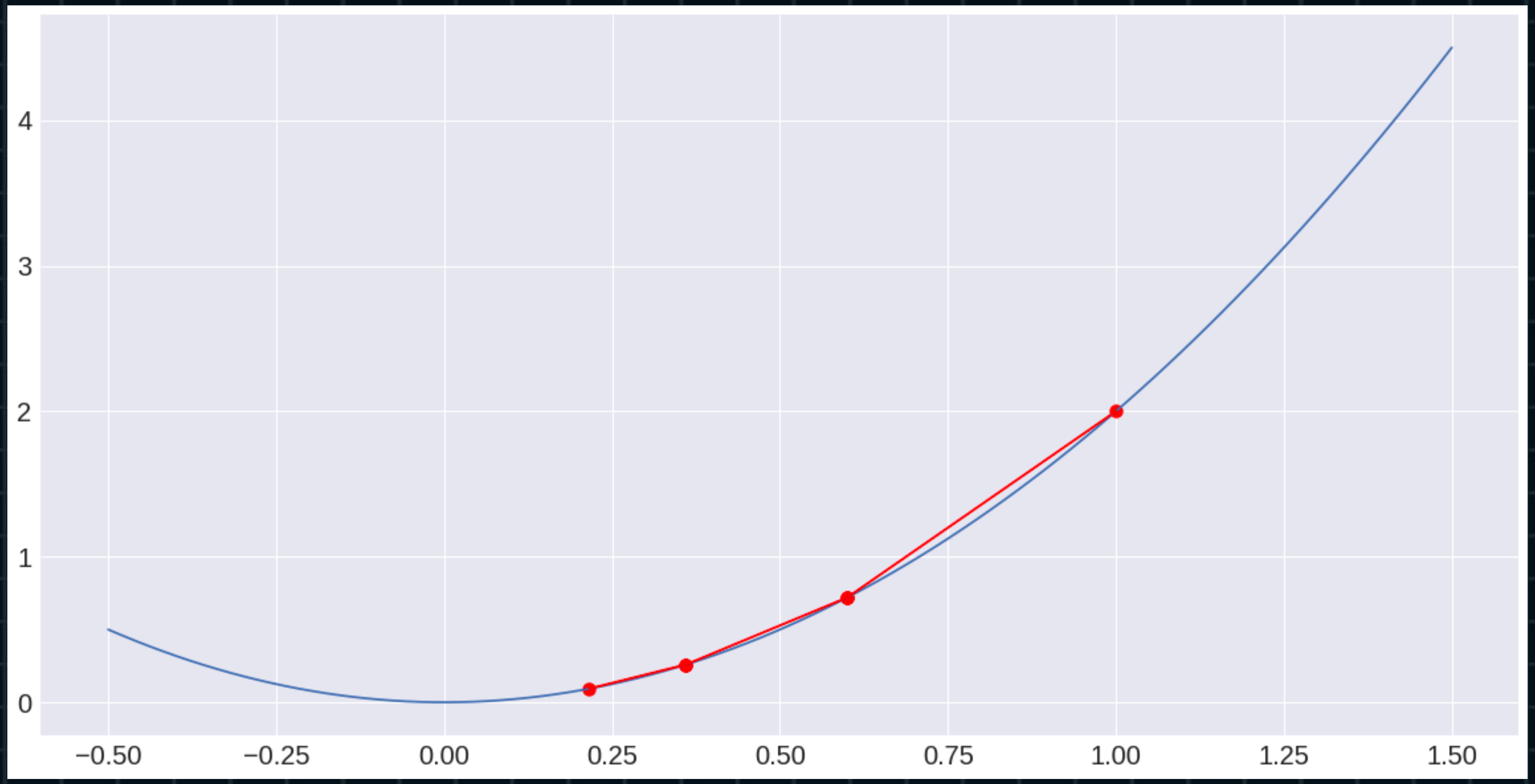
## Effectiveness of Gradients




$$
\begin{aligned}
x &:= x - f'(x) \\
&= x - 4x \\
&= -3x
\end{aligned}
$$

$$
\begin{aligned}
x &:= -3 \cdot 1 = -3 \\
x &:= -3 \cdot (-3) = 9 \\
x &:= -3 \cdot 9 = -27
\end{aligned}
$$

# Learning Rate and Gradient-based Learning




$$x := x - \alpha f'(x)$$

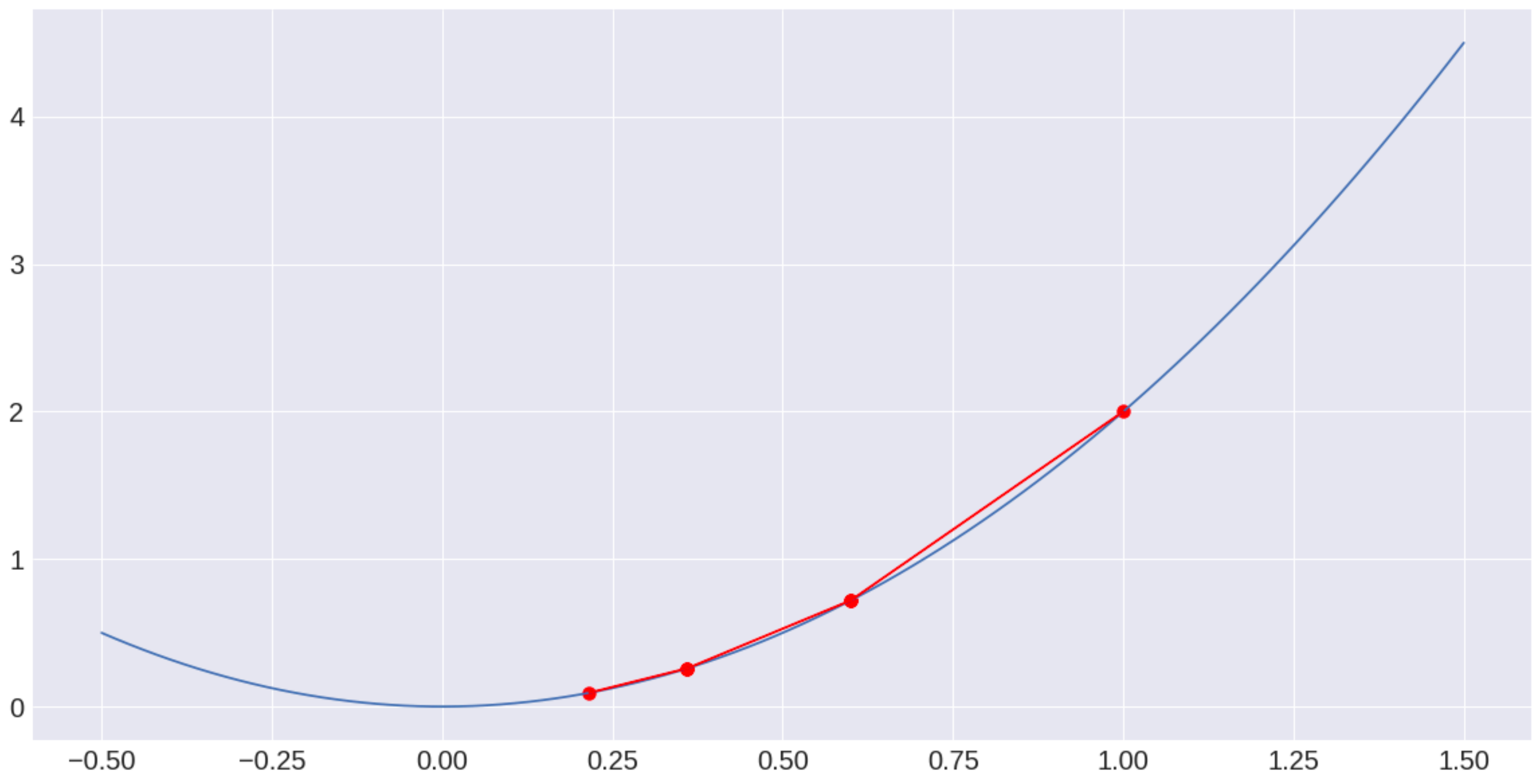
$$x := 0.6x$$

$$\begin{aligned} x &:= x - 0.1 \cdot f'(x) \\ &= x - 0.4x \\ &= 0.6x \end{aligned}$$

$$\begin{aligned} x &:= 0.6 \cdot 1 = 0.6 \\ x &:= 0.6 \cdot 0.6 = 0.36 \\ x &:= 0.6 \cdot 0.36 = 0.216 \end{aligned}$$

# - Gradient-based Learning

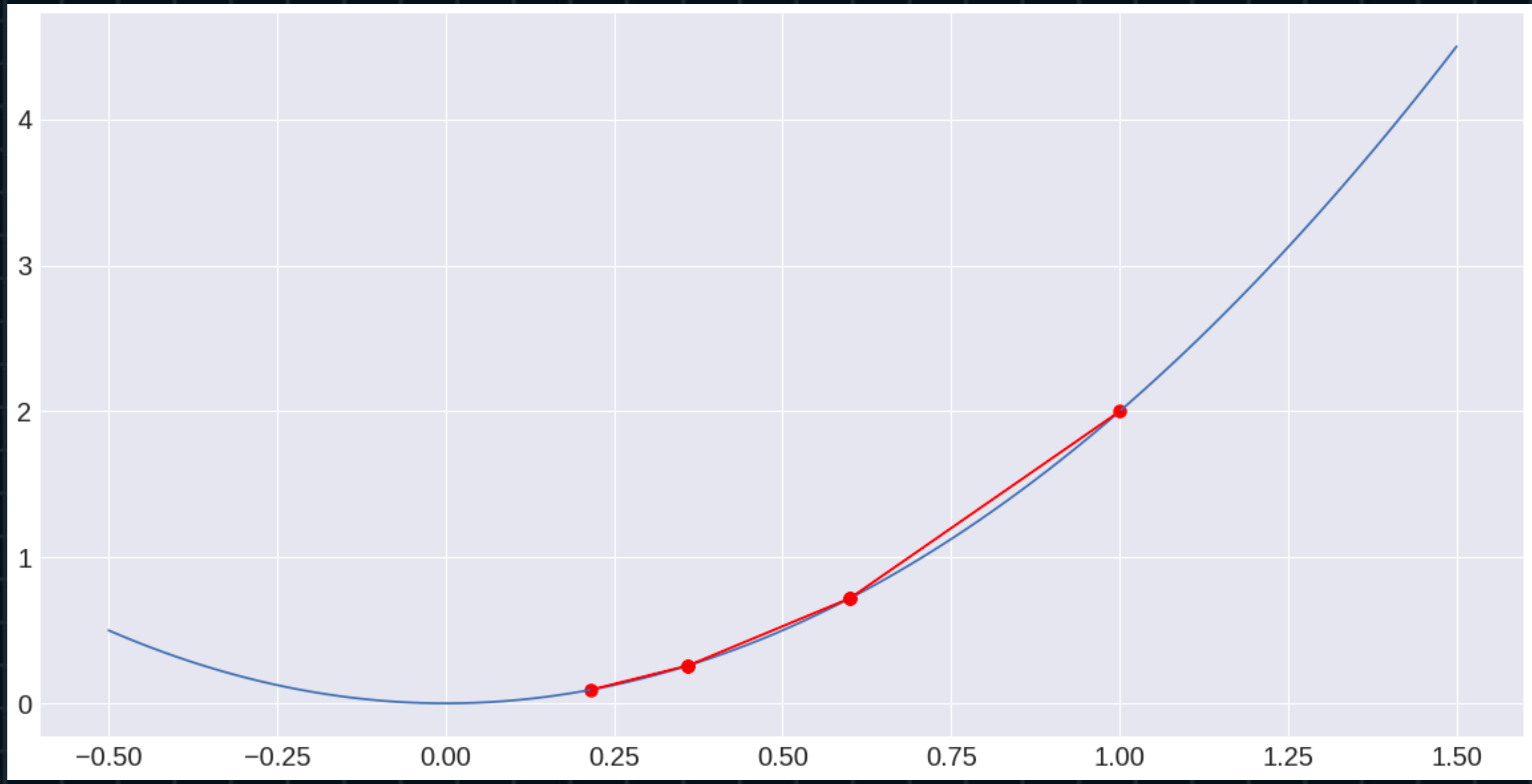
## Descending Without a Map




$$x := x - \alpha f'(x)$$

# - Gradient-based Learning

## Target of Gradient




$$J = \mathscr{L}(y, \hat{y})$$

$$x := x - \alpha \mathscr{L}'(x)$$

# - Backpropagation

## Chain Rule

$$y = f_4(f_3(f_2(f_1(x))))$$

$$y = f_4(u_3) \qquad \frac{\partial y}{\partial u_3}$$

$$u_3 = f_3(u_2) \qquad \frac{\partial u_3}{\partial u_2} \qquad \frac{\partial y}{\partial u_3}\frac{\partial u_3}{\partial u_2} = \frac{\partial y}{\partial u_2}$$

$$u_2 = f_2(u_1) \qquad \frac{\partial u_2}{\partial u_1} \qquad \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial u_1} = \frac{\partial y}{\partial u_1}$$

$$u_1 = f_1(x) \qquad \frac{\partial u_1}{\partial x} \qquad \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x} = \frac{\partial y}{\partial x}$$

# - Backpropagation

## Chain Rule

$$y = f_4(f_3(f_2(f_1(x))))$$

$$y = f_4(u_3)$$

$$u_3 = f_3(u_2)$$

$$u_2 = f_2(u_1)$$

$$u_1 = f_1(x)$$

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial y}{\partial u_1}$$

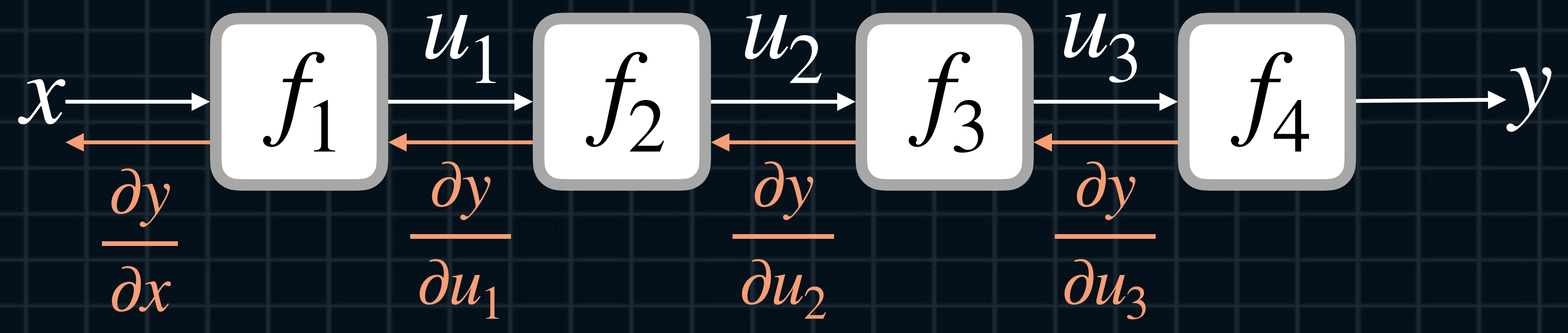
$$\frac{\partial y}{\partial u_2}$$

Forward

Backward

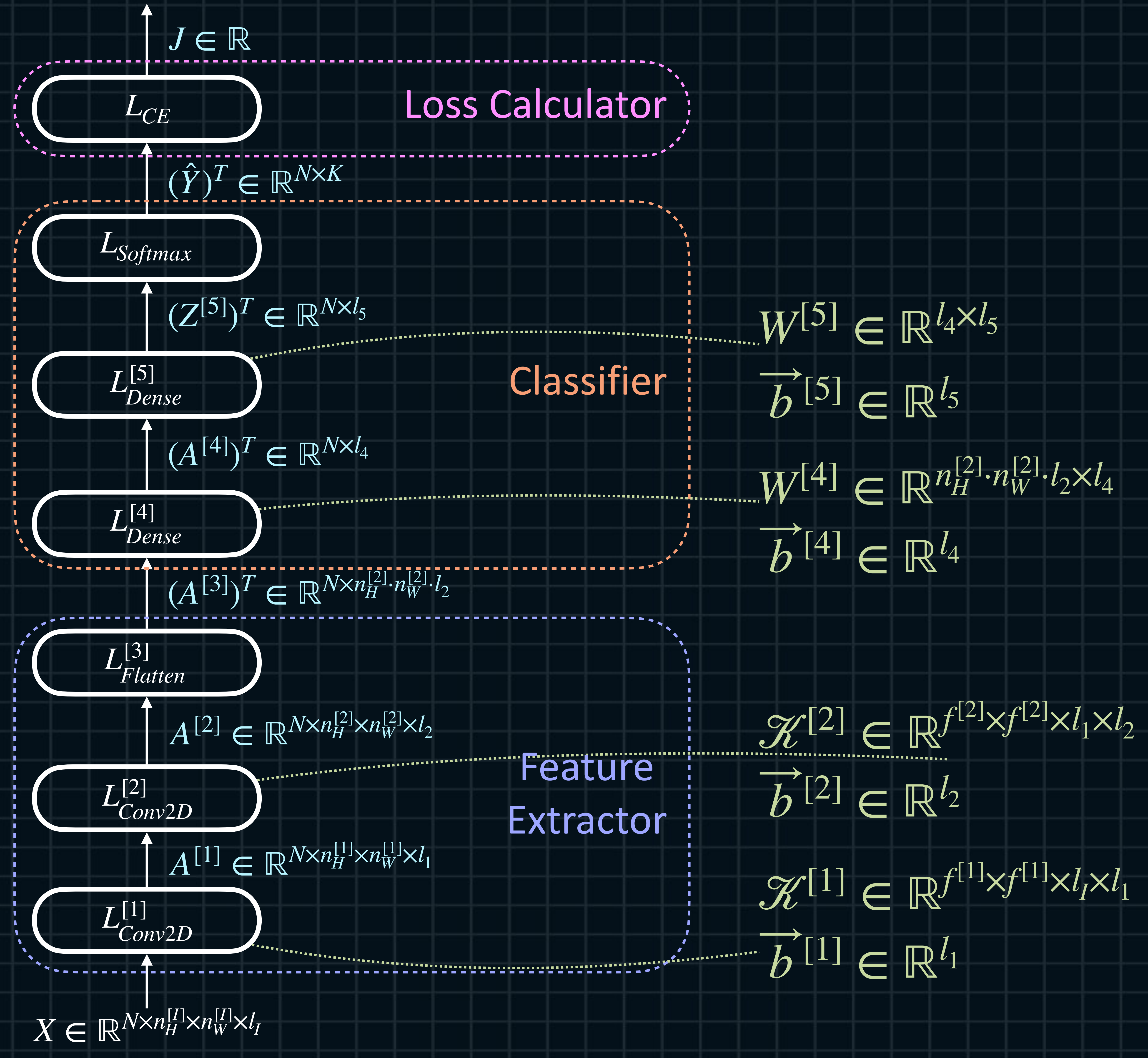
# - Backpropagation

## Chain Rule



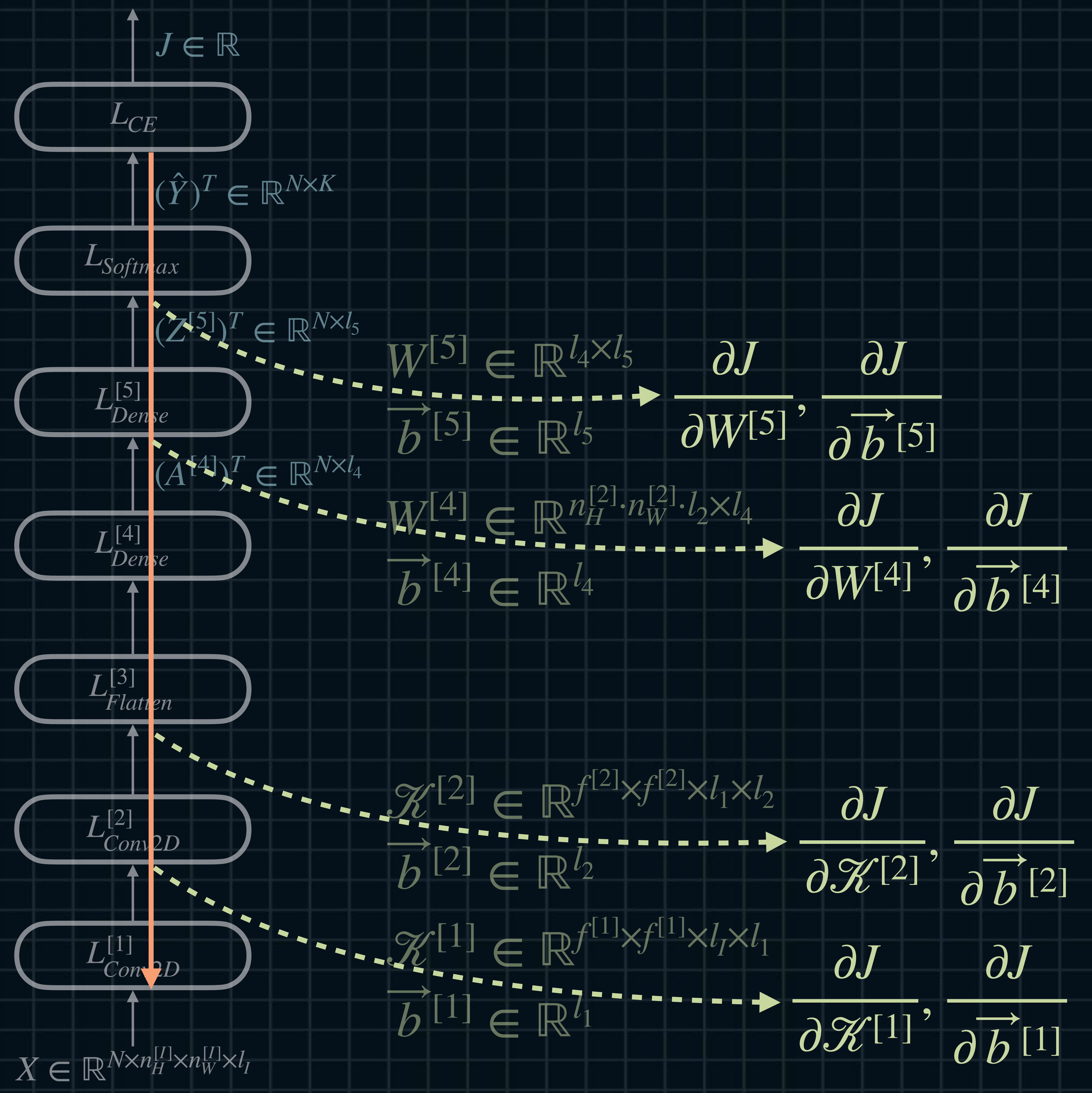
# - Backpropagation

## Chain Rule in Deep Learning



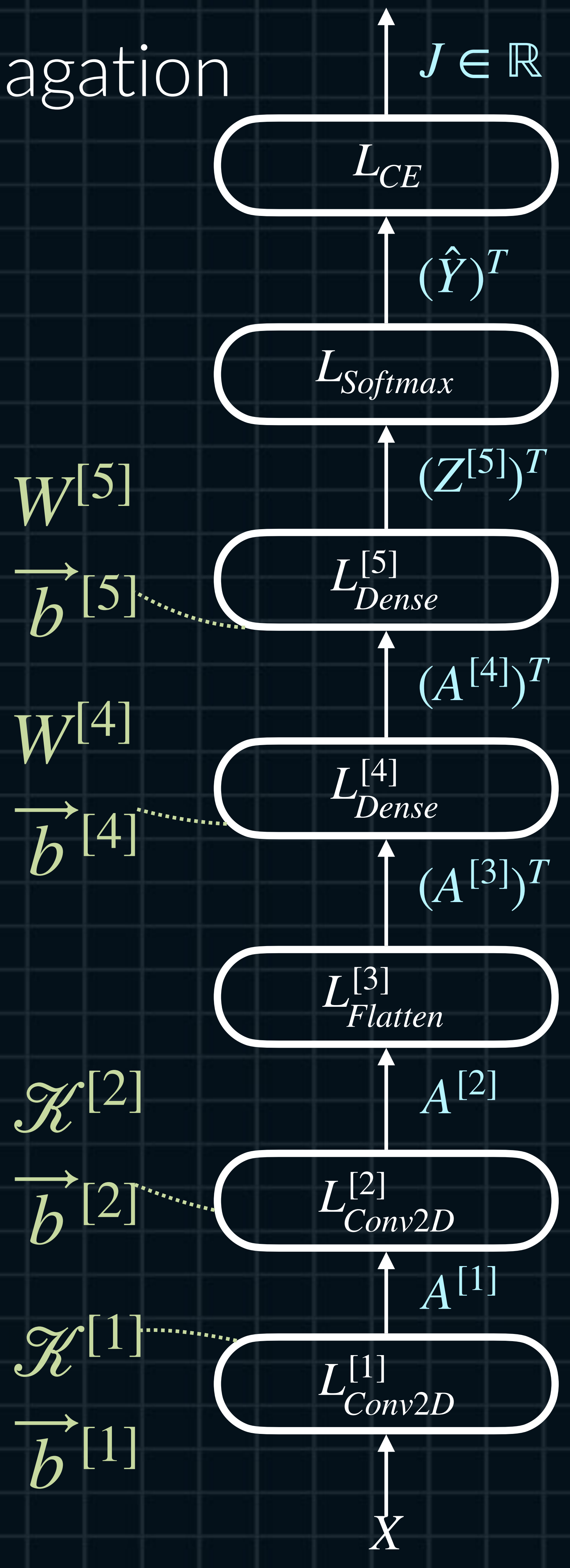
# - Backpropagation

## Backpropagation



# - Backpropagation

## Backpropagation




$$\frac{\partial J}{\partial \hat{Y}}$$

$$\frac{\partial \hat{Y}}{\partial Z^{[5]}}$$

$$\frac{\partial Z^{[5]}}{\partial A^{[4]}}, \frac{\partial Z^{[5]}}{\partial W^{[5]}}, \frac{\partial Z^{[5]}}{\partial \vec{b}^{[5]}}$$

$$\frac{\partial A^{[4]}}{\partial A^{[3]}}, \frac{\partial A^{[4]}}{\partial W^{[4]}}, \frac{\partial A^{[4]}}{\partial \vec{b}^{[4]}}$$

$$\frac{\partial A^{[3]}}{\partial A^{[2]}}$$

$$\frac{\partial A^{[2]}}{\partial A^{[1]}}, \frac{\partial A^{[2]}}{\partial \mathscr{K}^{[2]}}, \frac{\partial A^{[2]}}{\partial \vec{b}^{[2]}}$$

$$\frac{\partial A^{[1]}}{\partial \mathscr{K}^{[1]}}, \frac{\partial A^{[1]}}{\partial \vec{b}^{[1]}}$$

$$\frac{\partial J}{\partial \hat{Y}} \frac{\partial \hat{Y}}{\partial Z^{[5]}} = \frac{\partial J}{\partial Z^{[5]}}$$

$$\frac{\partial J}{\partial Z^{[5]}} \frac{\partial Z^{[5]}}{\partial A^{[4]}} = \frac{\partial J}{\partial A^{[4]}} \qquad \frac{\partial J}{\partial Z^{[5]}} \frac{\partial Z^{[5]}}{\partial W^{[5]}} = \frac{\partial J}{\partial W^{[5]}} \qquad \frac{\partial J}{\partial Z^{[5]}} \frac{\partial Z^{[5]}}{\partial \vec{b}^{[5]}} = \frac{\partial J}{\partial \vec{b}^{[5]}}$$

$$\frac{\partial J}{\partial A^{[4]}} \frac{\partial A^{[4]}}{\partial A^{[3]}} = \frac{\partial J}{\partial A^{[3]}} \qquad \frac{\partial J}{\partial A^{[4]}} \frac{\partial A^{[4]}}{\partial W^{[4]}} = \frac{\partial J}{\partial W^{[4]}} \qquad \frac{\partial J}{\partial A^{[4]}} \frac{\partial A^{[4]}}{\partial \vec{b}^{[4]}} = \frac{\partial J}{\partial \vec{b}^{[4]}}$$

$$\frac{\partial J}{\partial A^{[3]}} \frac{\partial A^{[3]}}{\partial A^{[2]}} = \frac{\partial J}{\partial A^{[2]}}$$

$$\frac{\partial J}{\partial A^{[2]}} \frac{\partial A^{[2]}}{\partial A^{[1]}} = \frac{\partial J}{\partial A^{[1]}} \qquad \frac{\partial J}{\partial A^{[2]}} \frac{\partial A^{[2]}}{\partial \mathscr{K}^{[2]}} = \frac{\partial J}{\partial \mathscr{K}^{[2]}} \qquad \frac{\partial J}{\partial A^{[2]}} \frac{\partial A^{[2]}}{\partial \vec{b}^{[2]}} = \frac{\partial J}{\partial \vec{b}^{[2]}}$$

$$\frac{\partial J}{\partial A^{[1]}} \frac{\partial A^{[1]}}{\partial \mathscr{K}^{[1]}} = \frac{\partial J}{\partial \mathscr{K}^{[1]}} \qquad \frac{\partial J}{\partial A^{[1]}} \frac{\partial A^{[1]}}{\partial \vec{b}^{[1]}} = \frac{\partial J}{\partial \vec{b}^{[1]}}$$

- Why Jacobians?

## Derivatives of Scalars

$$a \uparrow \boxed{\nu} \uparrow x$$

$$\nu(x; w, b) = g(xw + b)$$

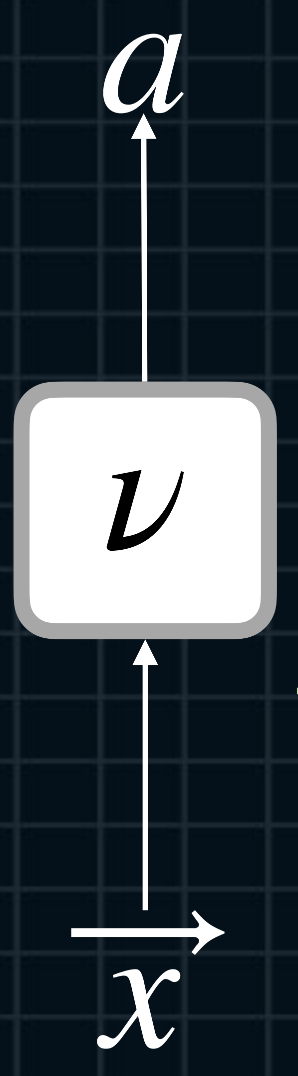
$$w \in \mathbb{R}$$

$$b \in \mathbb{R}$$

$$w := w - \alpha \frac{\partial J}{\partial w} \qquad b := b - \alpha \frac{\partial J}{\partial b}$$

# - Why Jacobians?

## Derivatives of Vectors




$$\nu\left(\overrightarrow{x}; \overrightarrow{w}, b\right) = g\left((\overrightarrow{x})^T \overrightarrow{w} + b\right)$$

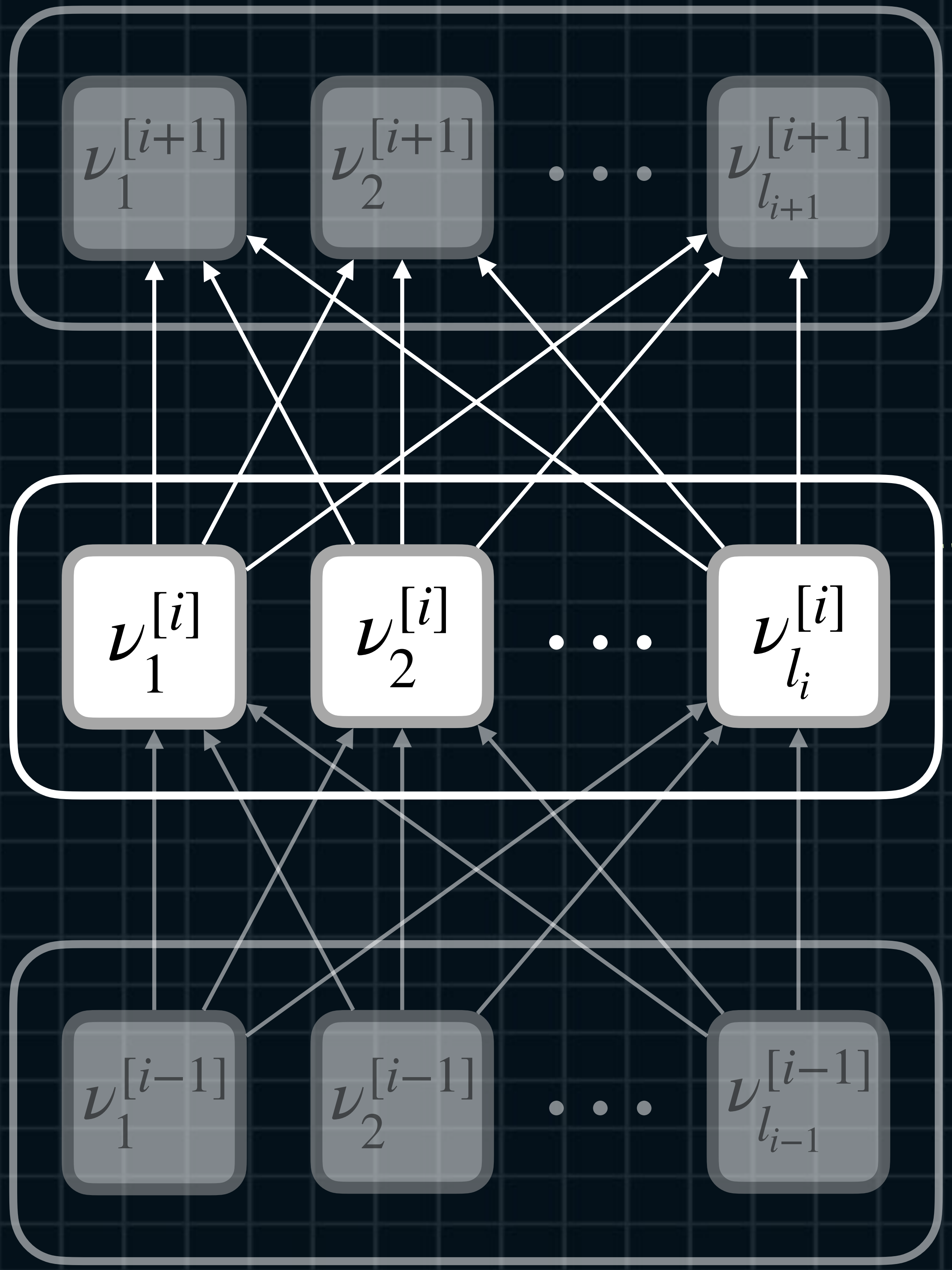
$$\overrightarrow{w} \in \mathbb{R}^{l_l \times 1}$$

$$b \in \mathbb{R}$$

$$\overrightarrow{w} := \overrightarrow{w} - \alpha \frac{\partial J}{\partial \overrightarrow{w}} \qquad b := b - \alpha \frac{\partial J}{\partial b}$$

# - Why Jacobians?

## Derivatives of Matrices




$$W^{[i]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{i-1} \times l_i}$$

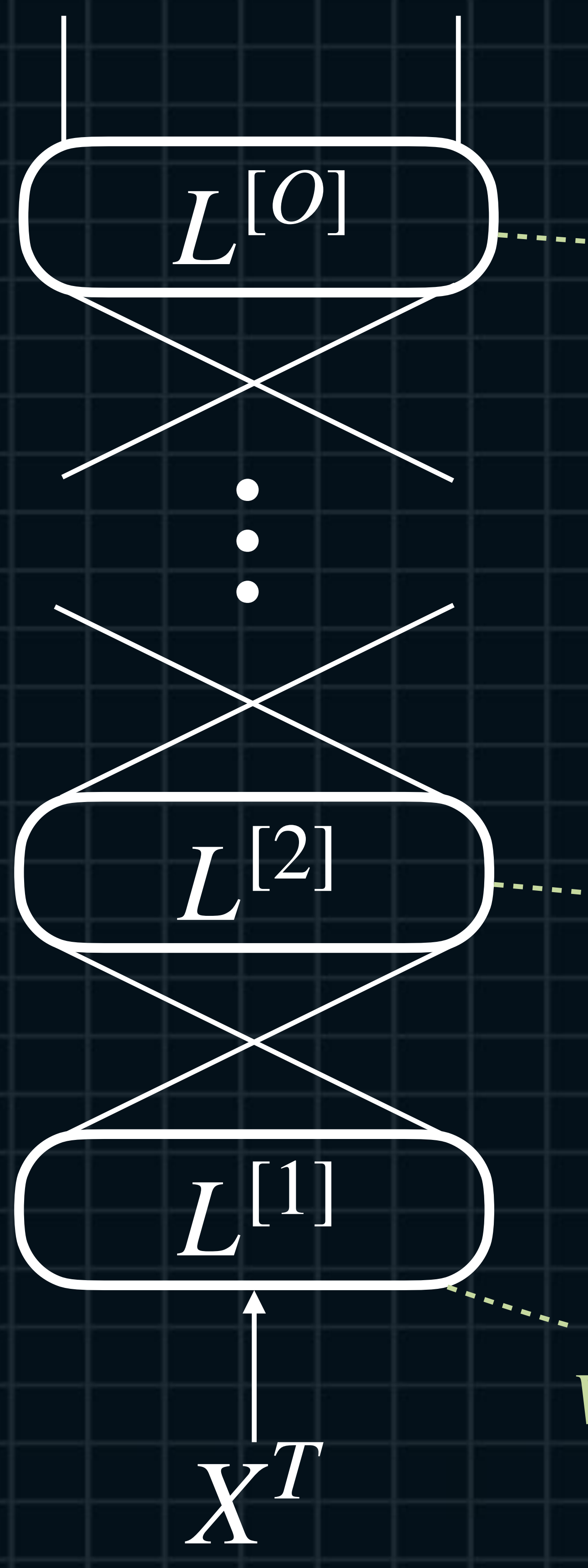
$$(\vec{b}^{[i]})^T = \begin{pmatrix} b_1^{[i]} & b_2^{[i]} & \dots & b_{l_i}^{[i]} \end{pmatrix} \in \mathbb{R}^{1 \times l_i}$$

$$W^{[i]} := W^{[i]} - \alpha \frac{\partial J}{\partial W^{[i]}}$$

$$\vec{b}^{[i]} := \vec{b}^{[i]} - \alpha \frac{\partial J}{\partial \vec{b}^{[i]}}$$

# - Why Jacobians?

## Derivatives of Matrices




$$W^{[O]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[O]} & \vec{w}_2^{[O]} & \cdots & \vec{w}_{l_O}^{[O]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{O-1}\times l_O}$$

$$(\vec{b}^{[O]})^T = \begin{pmatrix} b_1^{[O]} & b_2^{[O]} & \dots & b_{l_O}^{[O]} \end{pmatrix} \in \mathbb{R}^{1\times l_O}$$

$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_1\times l_2}$$

$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \dots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1\times l_2}$$

$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I\times l_1}$$

$$(\vec{b}^{[1]})^T = \begin{pmatrix} b_1^{[1]} & b_2^{[1]} & \dots & b_{l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{1\times l_1}$$

$$W^{[O]} := W^{[O]} - \alpha \frac{\partial J}{\partial W^{[O]}} \qquad \vec{b}^{[O]} := \vec{b}^{[O]} - \alpha \frac{\partial J}{\partial \vec{b}^{[O]}}$$

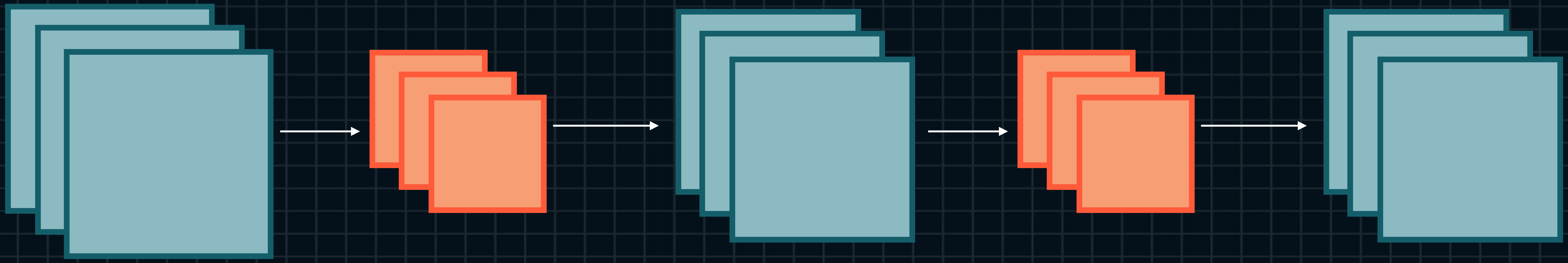
$$W^{[2]} := W^{[2]} - \alpha \frac{\partial J}{\partial W^{[2]}} \qquad \vec{b}^{[2]} := \vec{b}^{[2]} - \alpha \frac{\partial J}{\partial \vec{b}^{[2]}}$$

$$W^{[1]} := W^{[1]} - \alpha \frac{\partial J}{\partial W^{[1]}} \qquad \vec{b}^{[1]} := \vec{b}^{[1]} - \alpha \frac{\partial J}{\partial \vec{b}^{[1]}}$$

Lecture.1
Why Backpropagation
and Jacobians?

# - Why Jacobians?

## Derivatives of Tensors



$X \in \mathbb{R}^{n_H^{[l]} \times n_W^{[l]} \times l_l}$

$\mathscr{K}^{[1]} \in \mathbb{R}^{f^{[1]} \times f^{[1]} \times l_l \times l_1}$

$\vec{b}^{[1]} \in \mathbb{R}^{l_1}$

$A^{[1]} \in \mathbb{R}^{n_H^{[1]} \times n_W^{[1]} \times l_1}$

$\mathscr{K}^{[2]} \in \mathbb{R}^{f^{[2]} \times f^{[2]} \times l_1 \times l_2}$

$\vec{b}^{[2]} \in \mathbb{R}^{l_2}$

$A^{[2]} \in \mathbb{R}^{n_H^{[2]} \times n_W^{[2]} \times l_2}$

$$\mathscr{K}^{[1]} := \mathscr{K}^{[1]} - \alpha \frac{\partial J}{\partial \mathscr{K}^{[1]}}$$

$$\vec{b}^{[1]} := \vec{b}^{[1]} - \alpha \frac{\partial J}{\partial \vec{b}^{[1]}}$$

$$\mathscr{K}^{[2]} := \mathscr{K}^{[2]} - \alpha \frac{\partial J}{\partial \mathscr{K}^{[2]}}$$

$$\vec{b}^{[2]} := \vec{b}^{[2]} - \alpha \frac{\partial J}{\partial \vec{b}^{[2]}}$$

# Lecture.1 Why Backpropagation and Jacobians?

## - Why Jacobians?

### Derivatives of Tensors

$W^{[5]} \in \mathbb{R}^{l_4 \times l_5}$

$\vec{b}^{[5]} \in \mathbb{R}^{l_5}$

$W^{[4]} \in \mathbb{R}^{n_H^{[2]} \cdot n_W^{[2]} \cdot l_2 \times l_4}$

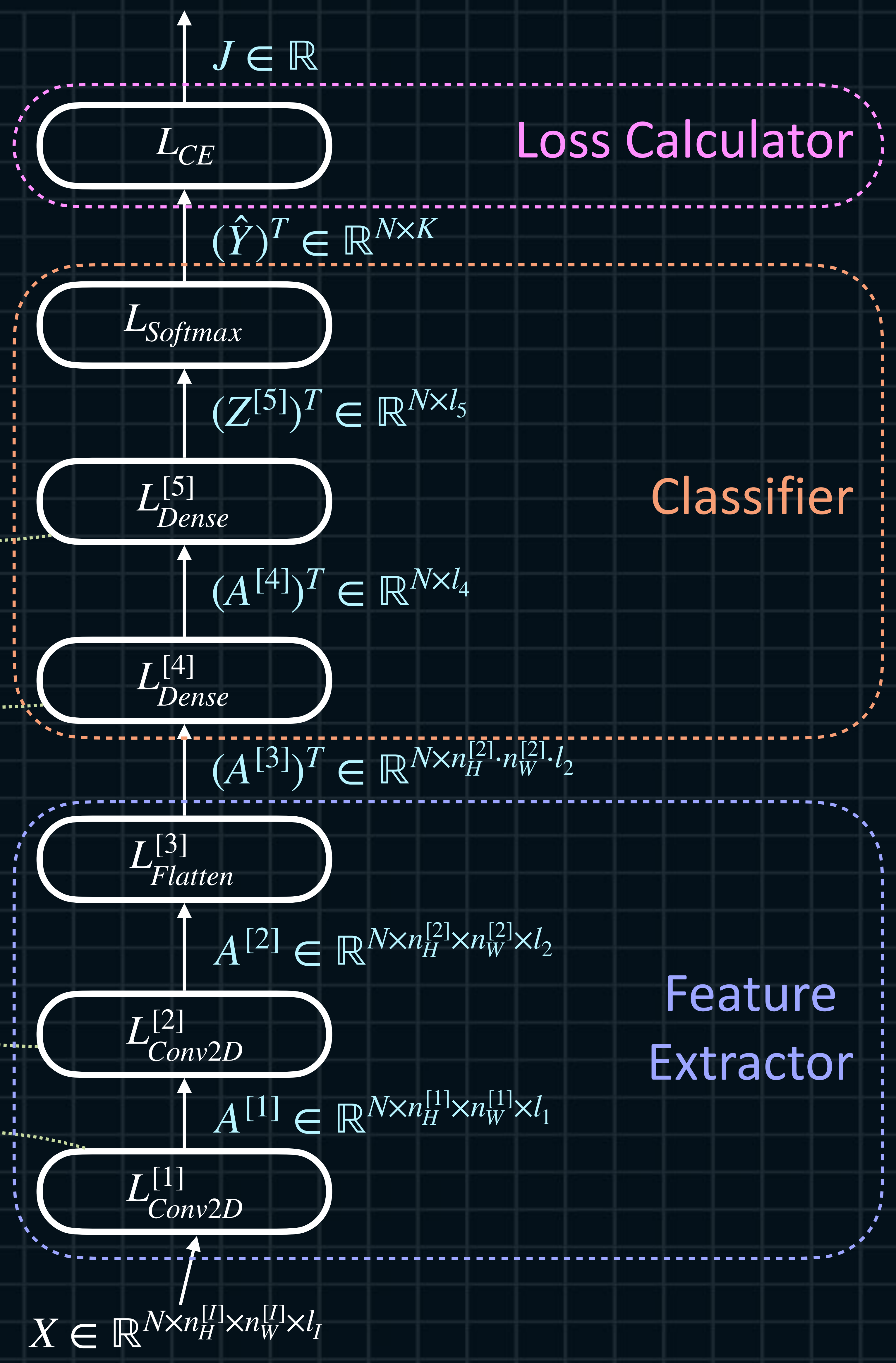
$\vec{b}^{[4]} \in \mathbb{R}^{l_4}$

$\mathscr{K}^{[2]} \in \mathbb{R}^{f^{[2]} \times f^{[2]} \times l_1 \times l_2}$

$\vec{b}^{[2]} \in \mathbb{R}^{l_2}$

$\mathscr{K}^{[1]} \in \mathbb{R}^{f^{[1]} \times f^{[1]} \times l_I \times l_1}$

$\vec{b}^{[1]} \in \mathbb{R}^{l_1}$




$$W^{[5]} := W^{[5]} - \alpha \frac{\partial J}{\partial W^{[5]}} \qquad \vec{b}^{[5]} := \vec{b}^{[5]} - \alpha \frac{\partial J}{\partial \vec{b}^{[5]}}$$

$$W^{[4]} := W^{[4]} - \alpha \frac{\partial J}{\partial W^{[4]}} \qquad \vec{b}^{[4]} := \vec{b}^{[4]} - \alpha \frac{\partial J}{\partial \vec{b}^{[4]}}$$

$$\mathscr{K}^{[2]} := \mathscr{K}^{[2]} - \alpha \frac{\partial J}{\partial \mathscr{K}^{[2]}} \qquad \vec{b}^{[2]} := \vec{b}^{[2]} - \alpha \frac{\partial J}{\partial \vec{b}^{[2]}}$$

$$\mathscr{K}^{[1]} := \mathscr{K}^{[1]} - \alpha \frac{\partial J}{\partial \mathscr{K}^{[1]}} \qquad \vec{b}^{[1]} := \vec{b}^{[1]} - \alpha \frac{\partial J}{\partial \vec{b}^{[1]}}$$

# - Theoretical and Practical Jacobians

Upgrades of Derivatives(Jacobians)

$$\frac{\partial f}{\partial x} \longrightarrow \frac{\partial f}{\partial \vec{x}} \longrightarrow \frac{\partial \vec{f}}{\partial x} \longrightarrow \frac{\partial \vec{f}}{\partial \vec{x}} \longrightarrow \frac{\partial f}{\partial X}$$

# - Theoretical and Practical Jacobians

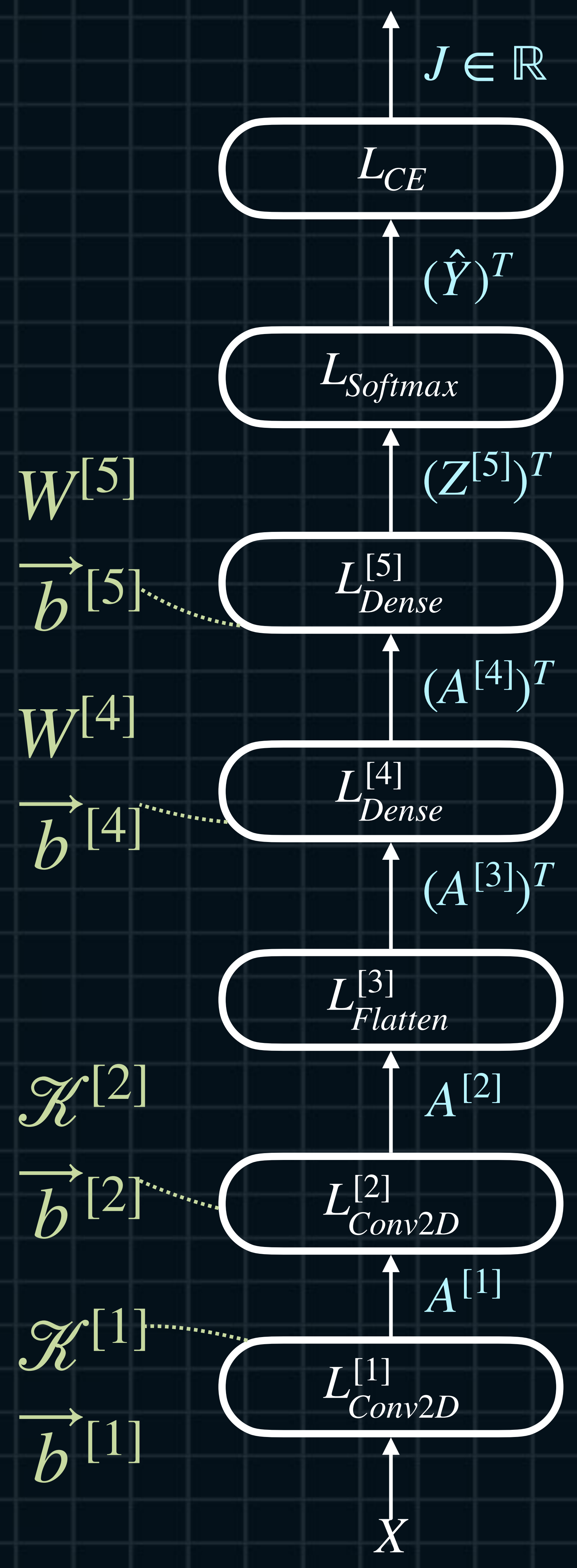
## Theoretical Jacobians

| | $x$ | $\vec{x}$ |
|---|---|---|
| $f$ | $\frac{\partial f}{\partial x}$ | $\frac{\partial f}{\partial \vec{x}}$ |
| $\vec{f}$ | $\frac{\partial \vec{f}}{\partial x}$ | $\frac{\partial \vec{f}}{\partial \vec{x}}$ |

Lecture.1
Why Backpropagation
and Jacobians?

# - Theoretical and Practical Jacobians

## Practical Jacobians




$\frac{\partial J}{\partial Z^{[5]}}$

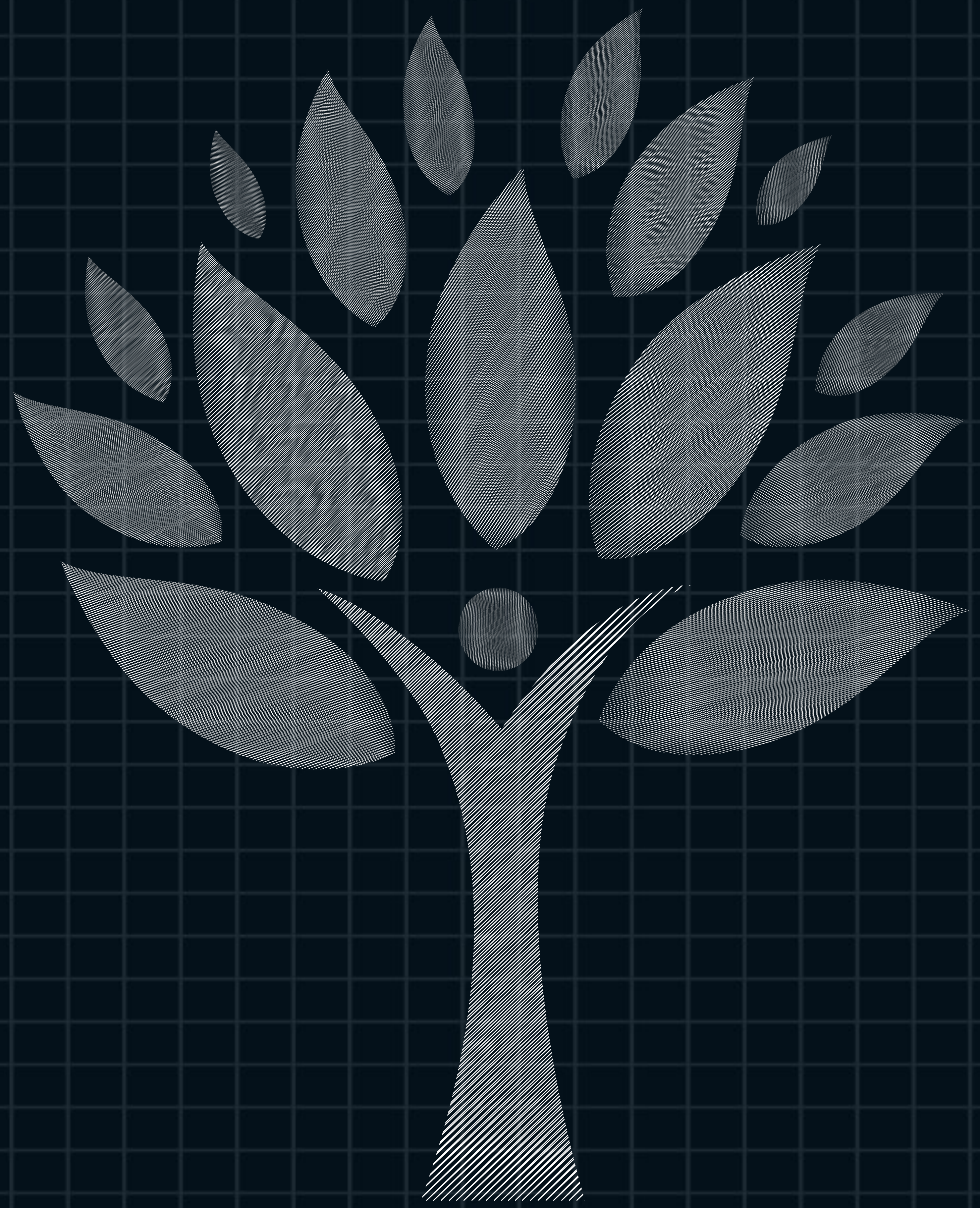
$\frac{\partial J}{\partial A^{[4]}}$ $\frac{\partial J}{\partial W^{[5]}}$ $\frac{\partial J}{\partial \vec{b}^{[5]}}$

$\frac{\partial J}{\partial A^{[3]}}$ $\frac{\partial J}{\partial W^{[4]}}$ $\frac{\partial J}{\partial \vec{b}^{[4]}}$

$\frac{\partial J}{\partial A^{[2]}}$

$\frac{\partial J}{\partial A^{[1]}}$ $\frac{\partial J}{\partial \mathscr{K}^{[2]}}$ $\frac{\partial J}{\partial \vec{b}^{[2]}}$

$\frac{\partial J}{\partial \mathscr{K}^{[1]}}$ $\frac{\partial J}{\partial \vec{b}^{[1]}}$



# Backpropagation and Jacobian Matrices

## Lecture.1 Why Backpropagation and Jacobians?