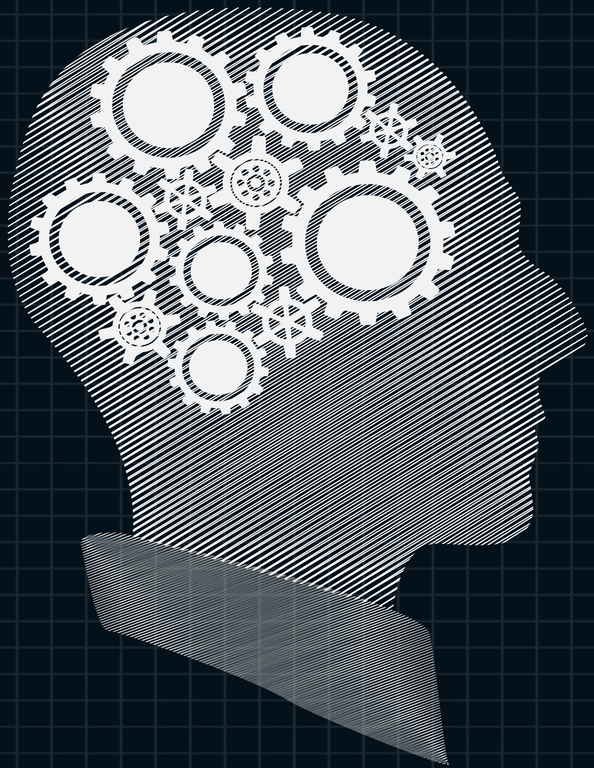


Backpropagation and Jacobian Matrices

Lecture.1
Why Backpropagation and Jacobians?

- Trainable Models and Params

Artificial Neurons

$$\hat{y} = xw + b$$

$$\hat{y} = \overrightarrow{x}^T \cdot \overrightarrow{w} + b$$

x는 n개일때,
w도 n개
b = 1

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

sigmoid...

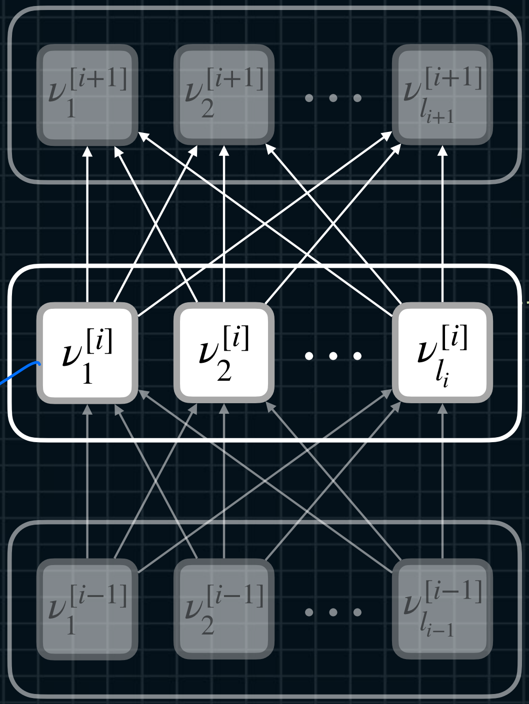
w:n
b:1

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

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

- Trainable Models and Params

Dense Layers



각각 neuron들이

l_i개 이용 weight

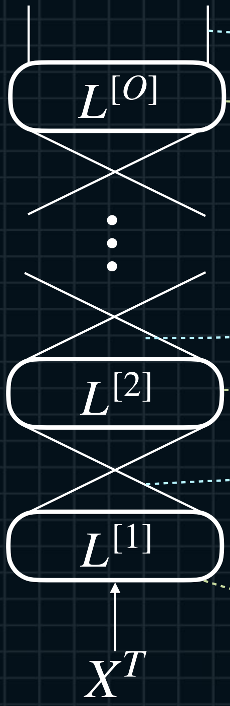
$$W^{[i]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{i-1} \times l_i}$$

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

Lecture.1 Why Backpropagation and Jacobians? - Trainable Models and Params

Dense Layers

학습해야되는 변수의 수가 기하급수적으로 늘어남



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

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

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

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

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

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

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

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

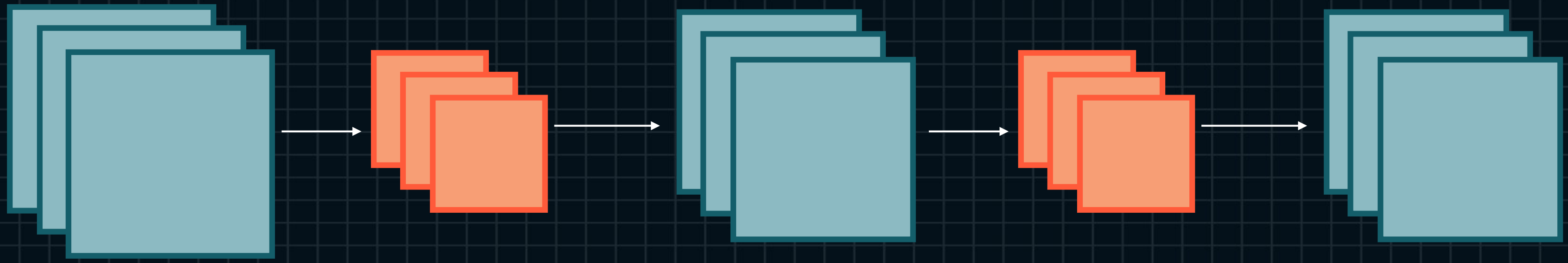
$$(\vec{b}^{[1]})^T = \begin{pmatrix} b_1^{[1]} & b_2^{[1]} & \ldots & b_{l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{1 \times l_1}$$

각 Layer마다 W와 b가 존재

Lecture.1
Why Backpropagation
and Jacobians?

- Trainable Models and Params

Conv Layers



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

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

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

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

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

Output 크기는 filter 개수에 따라

Kernel (Tensor)의 형태 또는 bias vec. 결정

- Trainable Models and Params

CNNs

$J \in \mathbb{R}$

Loss Calculator

L_{CE}

$(\hat{Y})^T \in \mathbb{R}^{N \times K}$

Classifier

$L_{Softmax}$

$(Z^{[5]})^T \in \mathbb{R}^{N \times l_5}$

$L^{[5]}_{Dense}$

$(A^{[4]})^T \in \mathbb{R}^{N \times l_4}$

$L^{[4]}_{Dense}$

$(A^{[3]})^T \in \mathbb{R}^{N \times n_H^{[2]} \cdot n_W^{[2]} \cdot l_2}$

Feature Extractor

$L^{[3]}_{Flatten}$

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

$L^{[2]}_{Conv2D}$

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

$L^{[1]}_{Conv2D}$

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

파라미터들의 형태가
Matrix | vector | 고차원의 Tensor

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

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

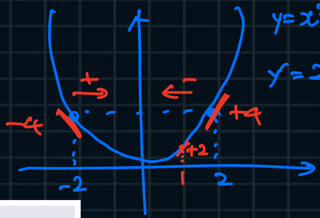
$\mathscr{K}^{[2]} \in \mathbb{R}^{f^{[2]} \times f^{[2]} \times l_1 \times l_2}$
$\vec{b}^{[2]} \in \mathbb{R}^{l_2}$

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

Lecture.1 Why Backpropagation and Jacobians?

- Gradient-based Learning

Differential Coefficient in DL



$y = x^2$

$y' = 2x$

부호 / 크기

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

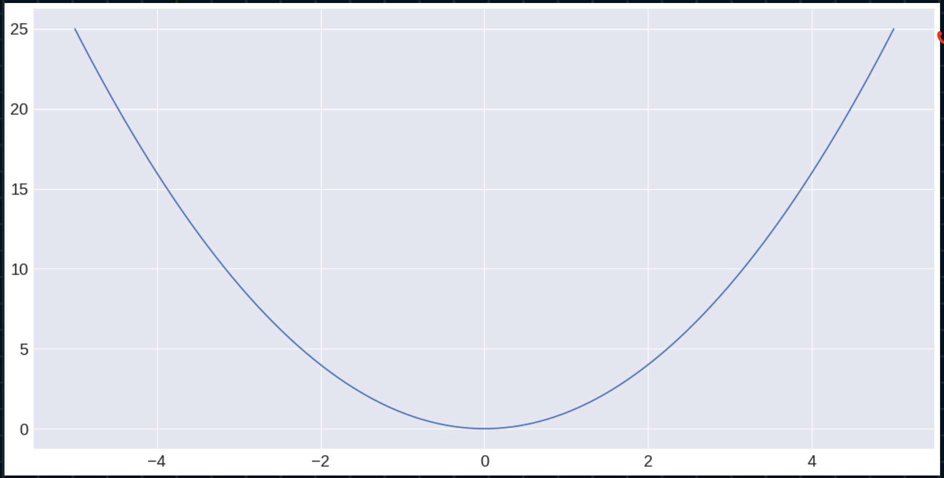
$y'|_{x=-2} = - 4$

y 값을 낮추려면?

y'(미분계수) 부호의 역순으로 진행

x값이 움직여야 하는 방향

움직여야 하는 방향 $= -\frac{dy}{dx}$



$$y' = 2x$$

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

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

$$y' = 2x$$

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

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

- Gradient-based Learning

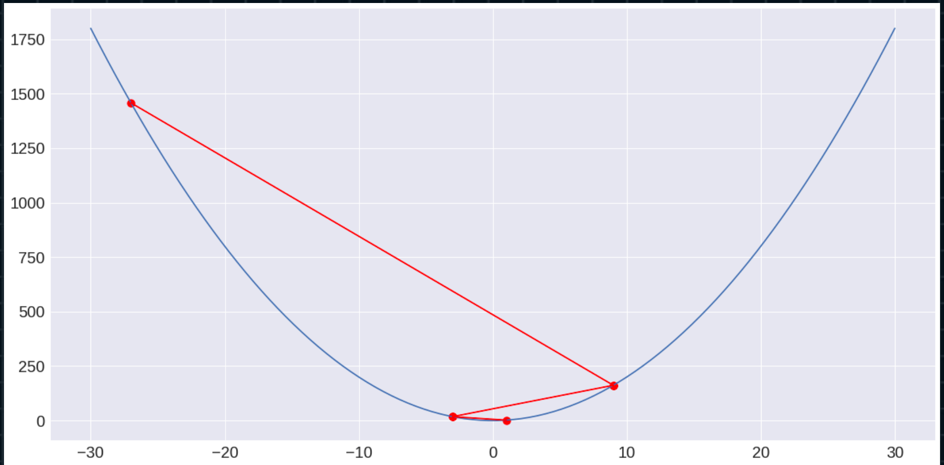
Update Notation

$$\underbrace{x :=}_{\text{update}} x + a$$

e.g $x = x + 2$

- Gradient-based Learning

Effectiveness of Gradients



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

프로그래밍 상의 문제?

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

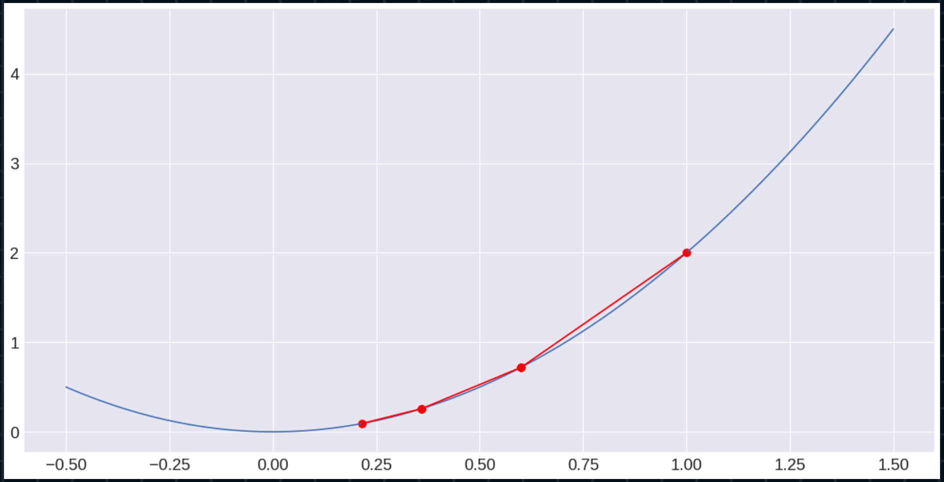
→ 오히려 반복할수록 멀어짐

- Gradient-based Learning

ex) like. 미분계수

유성에서의 한계 극복!

Learning Rate and Gradient-based Learning



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

↳ 미분계수를 완화 해주는 역할 (supressed)

$$x := x - 0.1 \cdot f'(x)$$
$$= x - 0.4x$$
$$= 0.6x$$

α = 하이퍼 파라미터 (사용자 조정)

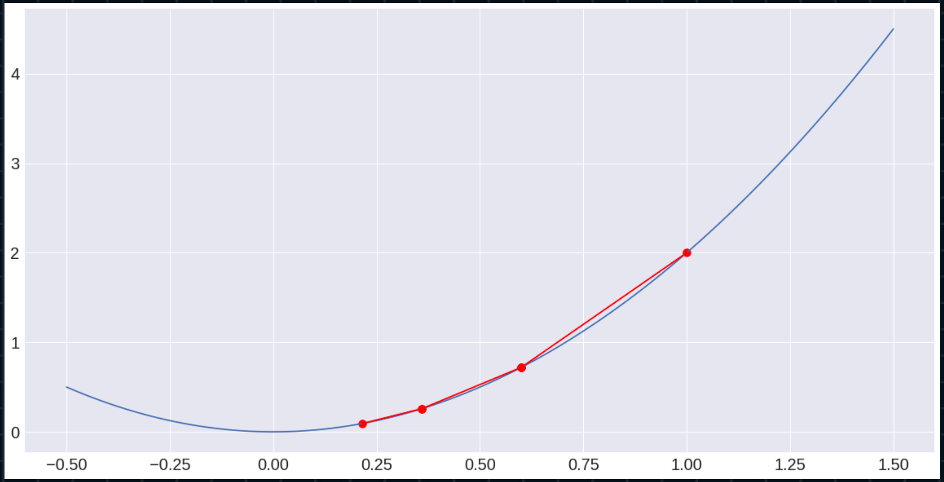
$$x := 0.6x$$

$$x := 0.6 \cdot 1 = 0.6$$
$$x := 0.6 \cdot 0.6 = 0.36$$
$$x := 0.6 \cdot 0.36 = 0.216$$

값이 작아지는 형태

- Gradient-based Learning

Descending Without a Map



실제 DL에서는 $y=x^2$과 같은 함수 형상을 알 수 X

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

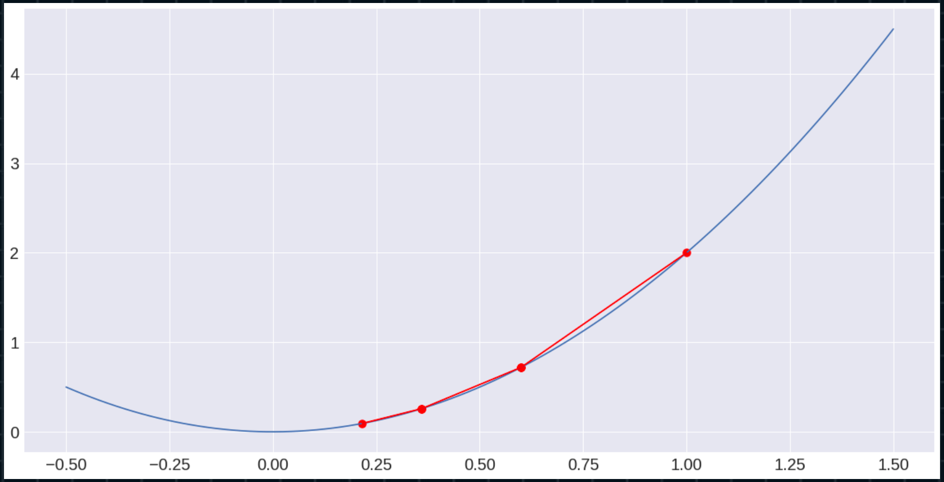
미분을 하는 이유?

- 함숫값을 최소로 하는 x 값을 찾기 위함
- 미분계수의 반대 방향으로 진행 필요

Lecture.1
Why Backpropagation and Jacobians?

- Gradient-based Learning

Target of Gradient



왜 함수 값을 줄여야 하는가?

→ Loss ↓

MSE (Mean Squ—)
BCEE (Binary Cross —)
CCEE (Categorical—)

$$\underbrace{J}_{\text{Loss}} = \mathscr{L}(y, \hat{y}) = \frac{\partial J}{\partial x}$$

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

Loss를 줄이기 위해 x 값을 어떻게 조정해야 하는가?

- Backpropagation

각각의 w.b 를 조정해가면서
J(loss)가 최소로 되는 값 [illegible]

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

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

$$y = f_4(u_3)$$

$$u_3 = f_3(u_2)$$

$$u_2 = f_2(u_1)$$

$$u_1 = f_1(x)$$

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

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

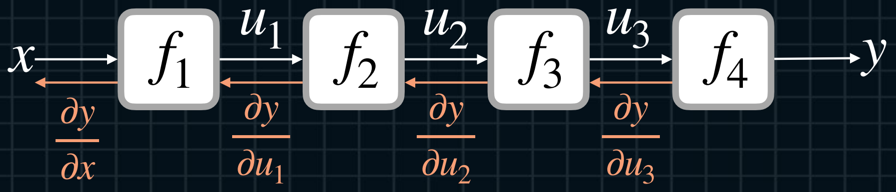
$$\frac{\partial y}{\partial u_2}$$

Forward

Backward

- Backpropagation

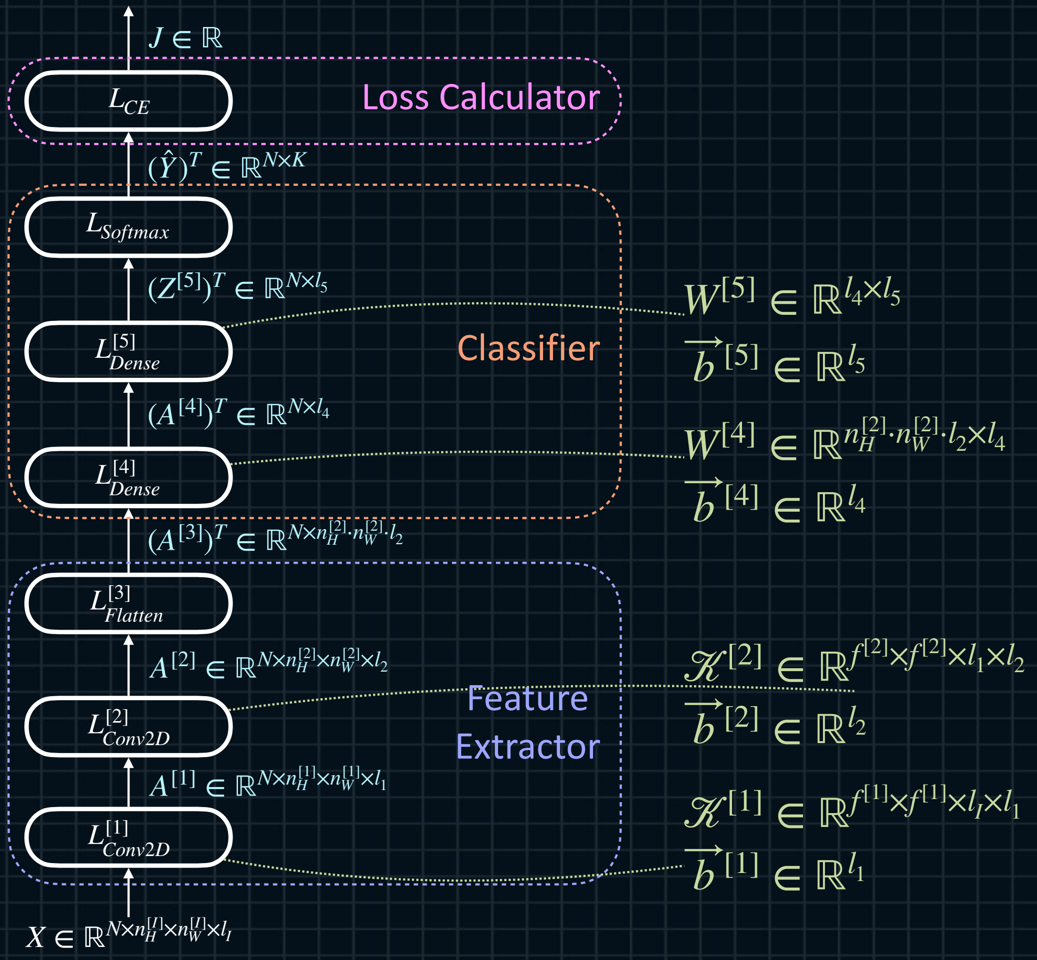
Chain Rule



Lecture.1
Why Backpropagation
and Jacobians?

- Backpropagation

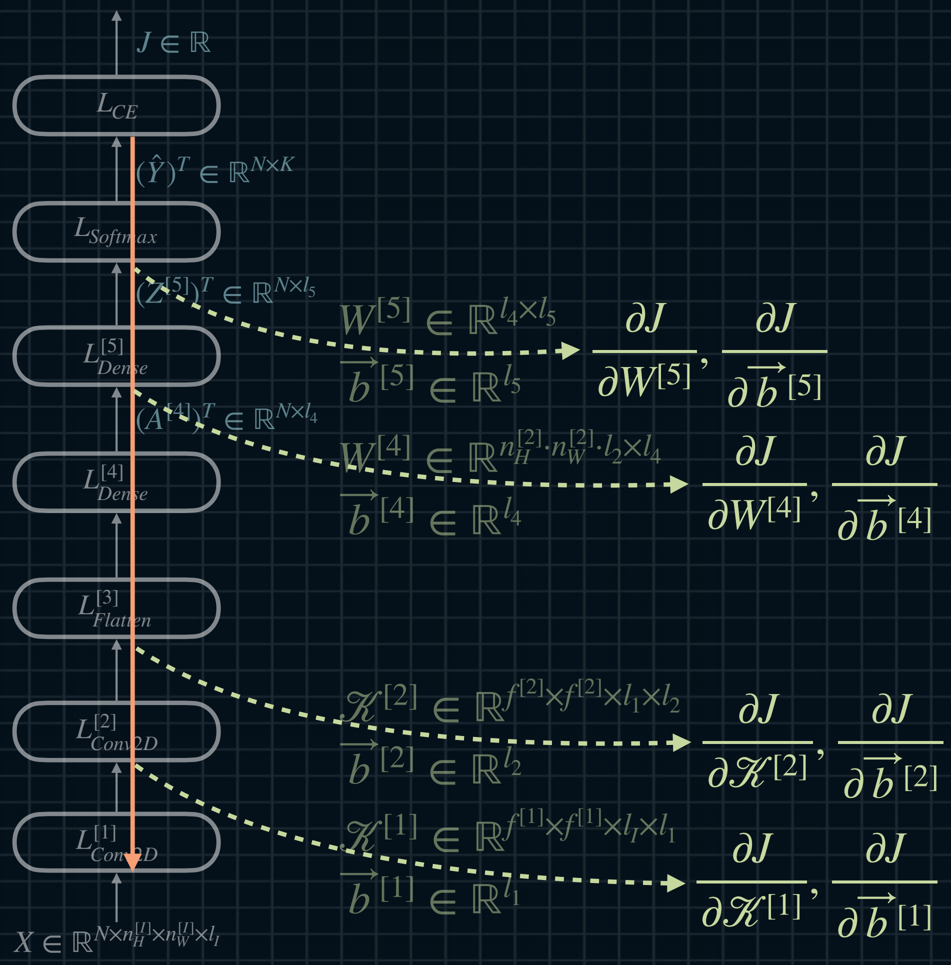
Chain Rule in Deep Learning



Lecture.1
Why Backpropagation and Jacobians?

- Backpropagation

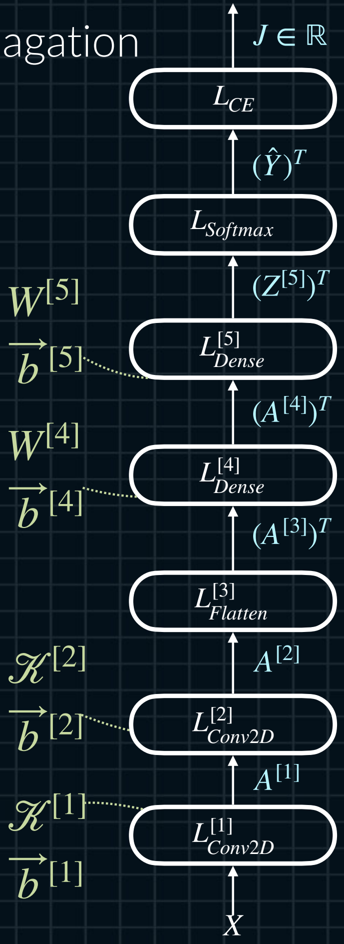
Backpropagation



Lecture.1
Why Backpropagation
and Jacobians?

- Backpropagation

Backpropagation



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

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

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

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

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

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

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

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

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

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

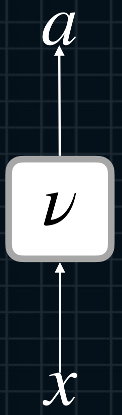
$$\frac{\partial J}{\partial A^{[3]}} \frac{\partial A^{[3]}}{\partial A^{[2]}} = \frac{\partial J}{\partial A^{[2]}}$$

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

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

- Why Jacobians?

Derivatives of Scalars



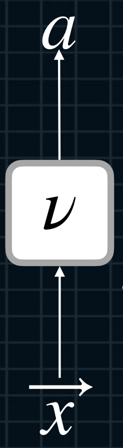
$$\nu(x; w, b) = g(xw + b)$$

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

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

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

Derivatives of Vectors



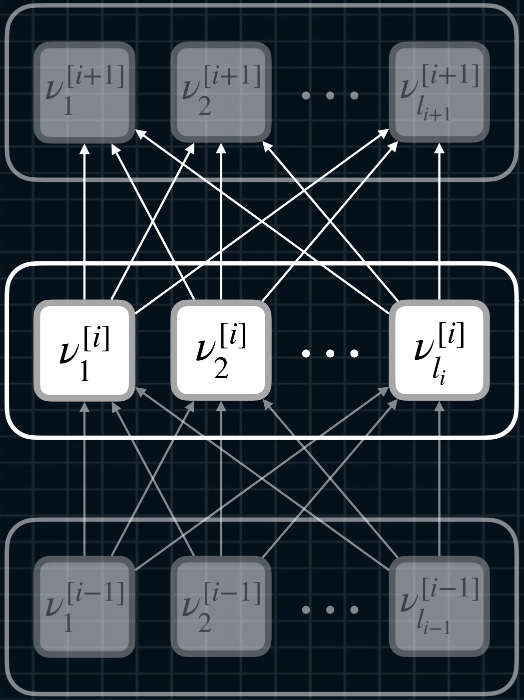
$$\nu(\vec{x}; \vec{w}, b) = g((\vec{x})^T \vec{w} + b)$$

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

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

- Why Jacobians?

Derivatives of Matrices



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

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

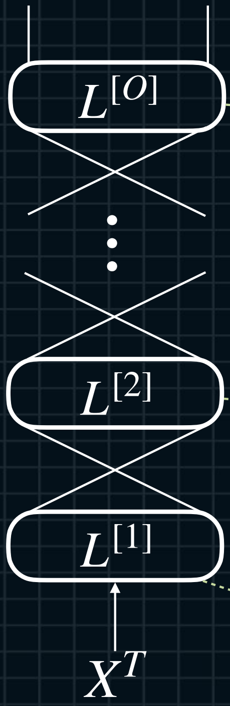
$$W^{[i]} := W^{[i]} - \alpha \frac{\partial J}{\partial W^{[i]}}$$

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

Lecture.1
Why Backpropagation
and Jacobians?

- Why Jacobians?

Derivatives of Matrices



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

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

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

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

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

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

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

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

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

Lecture.1
Why Backpropagation
and Jacobians?

- Why Jacobians?

Derivatives of Tensors

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

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

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

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

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

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

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

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

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

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

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

Lecture.1
Why Backpropagation and Jacobians?

- Why Jacobians?

Derivatives of Tensors

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

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

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

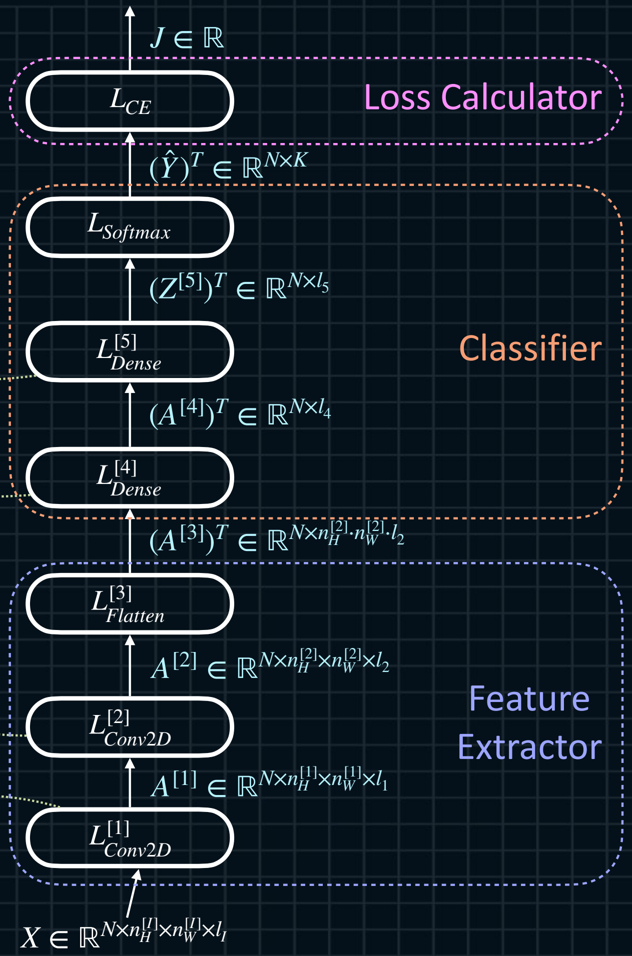
$\vec{b}^{[4]} \in \mathbb{R}^{l_4}$

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

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

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

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



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

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

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

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

- Theoretical and Practical Jacobians

Upgrades of Derivatives(Jacobians)

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

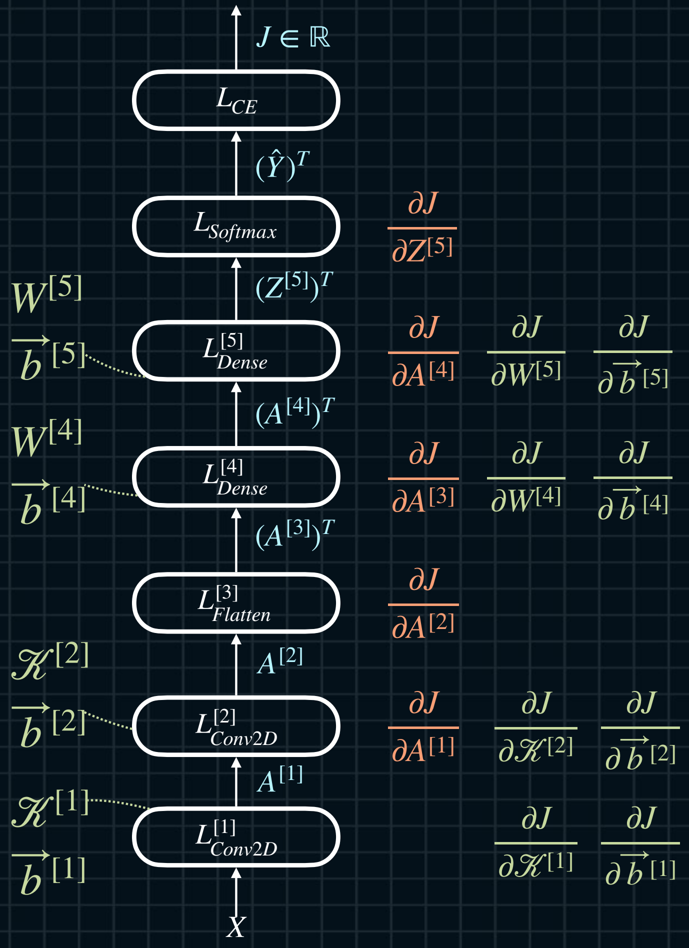
- Theoretical and Practical Jacobians

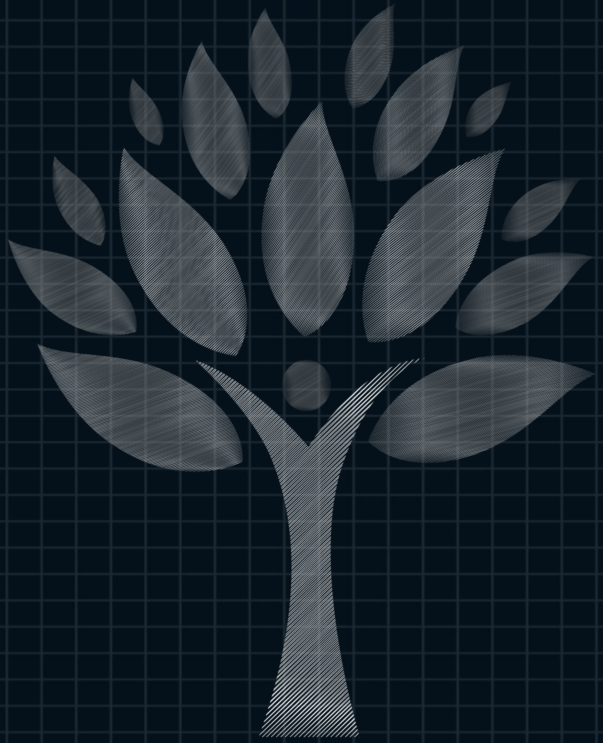
Theoretical Jacobians

	x	$\overrightarrow{x}$
f	$\frac{\partial f}{\partial x}$	$\frac{\partial f}{\partial \overrightarrow{x}}$
$\vec{f}$	$\frac{\partial \vec{f}}{\partial x}$	$\frac{\partial \vec{f}}{\partial \overrightarrow{x}}$

- Theoretical and Practical Jacobians

Practical Jacobians





Backpropagation and Jacobian Matrices

Lecture.1
Why Backpropagation and Jacobians?