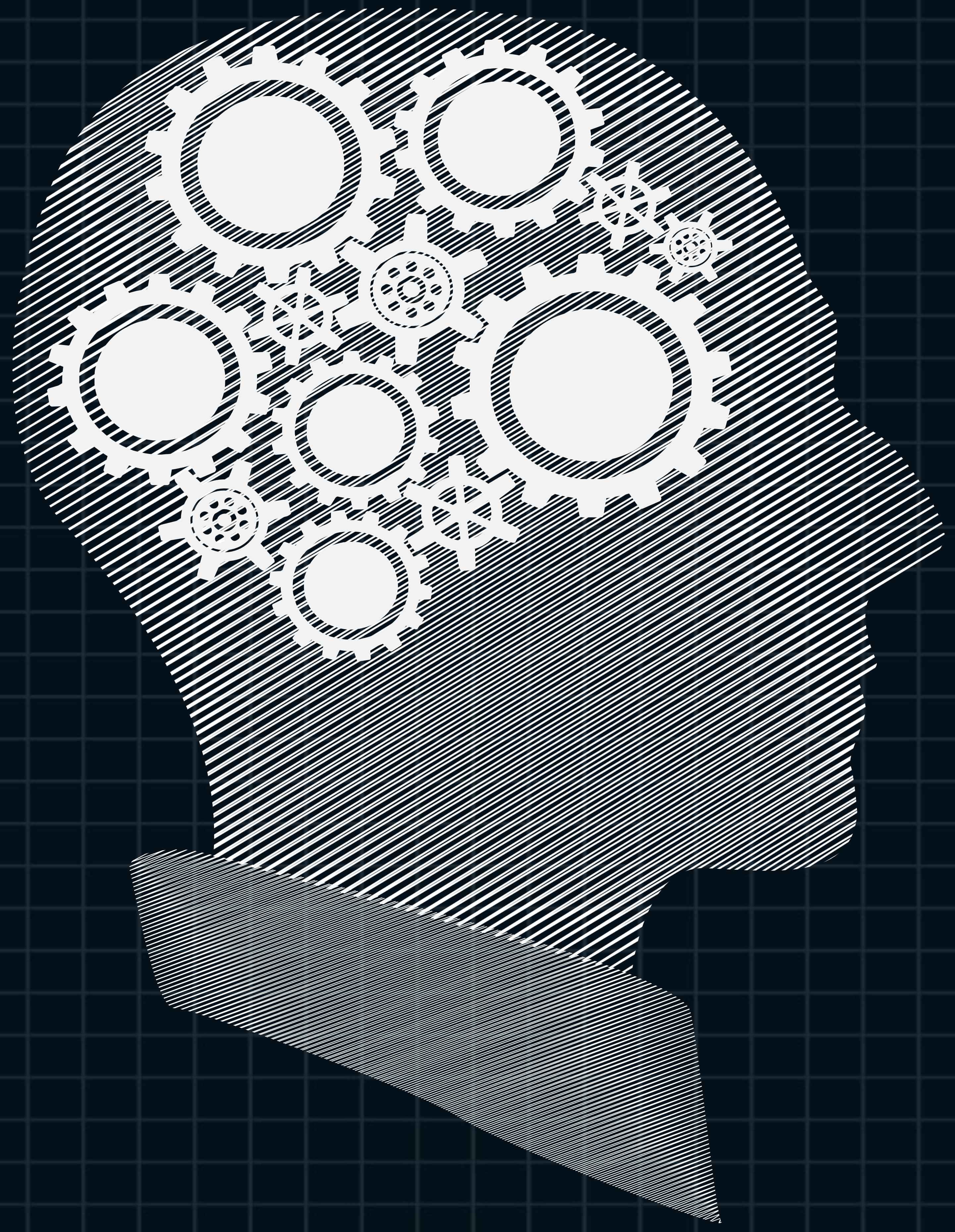


Backpropagation and Jacobian Matrices

Lecture.6 Element-wise Operations and Jacobians

Diagonal Matrices

main diagonal

$$M = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{pmatrix}$$

$$M = \begin{pmatrix} m_{11} & 0 & \dots & 0 \\ 0 & m_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_{nn} \end{pmatrix}$$

$$D = (m_{ij}),\ 1 \le i, j \le n$$

$$i \neq j \implies d_{ij} = 0$$

$$\forall D \in \mathbb{D}$$

Diagonal Matrices Notation

$$\begin{pmatrix} m_{11} & 0 & \dots & 0 \\ 0 & m_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_{nn} \end{pmatrix} = diag(m_{11}, m_{22}, \dots, m_{nn}) = diag(\dots, m_i, \ \dots)$$

Properties of Diagonal Matrix

$$A, B \in \mathbb{R}^{n\times n}$$

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} + \begin{pmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_n \end{pmatrix} = \begin{pmatrix} a_1+b_1 & 0 & \dots & 0 \\ 0 & a_2+b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n+b_n \end{pmatrix}$$

$$diag(a_1, a_2, \dots, a_n) + diag(b_1, b_2, \dots, b_n) = diag(a_1+b_1, a_1+b_2, \dots, a_n+b_n)$$

Lecture.6 Element-wise Operations and Jacobians - Diagonal Matrices

Properties of Diagonal Matrix

$$A, B \in \mathbb{R}^{n \times n}$$

$$\begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 & 0 & \dots & 0 \\ 0 & b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_n \end{pmatrix} = \begin{pmatrix} a_1 b_1 & 0 & \dots & 0 \\ 0 & a_2 b_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n b_n \end{pmatrix}$$

$$diag(a_1, a_2, \dots, a_n) \cdot diag(b_1, b_2, \dots, b_n) = diag(a_1 b_1, a_1 b_2, \dots, a_n b_n)$$

Identity Matrices

$$M \cdot M^{-1} = I$$

$$I = (i_{\alpha\beta}),\ 1 \le \alpha, \beta \le n$$

$$i_{\alpha\beta} = \begin{cases} 1, & if\ \alpha = \beta \\ 0, & if\ \alpha \ne \beta \end{cases}$$

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

$$M \cdot I = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{pmatrix} = M$$

Properties of Diagonal Matrix

$$A, B \in \mathbb{R}^{n\times n}$$

$$diag(a_1, a_2, \ldots, a_n) \cdot diag(a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1})$$

$$\begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{a_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{a_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{a_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

$$diag(a_1, a_2, \ldots, a_n)^{-1} = diag(a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1})$$

- Unary Element-wise Operations

General Case

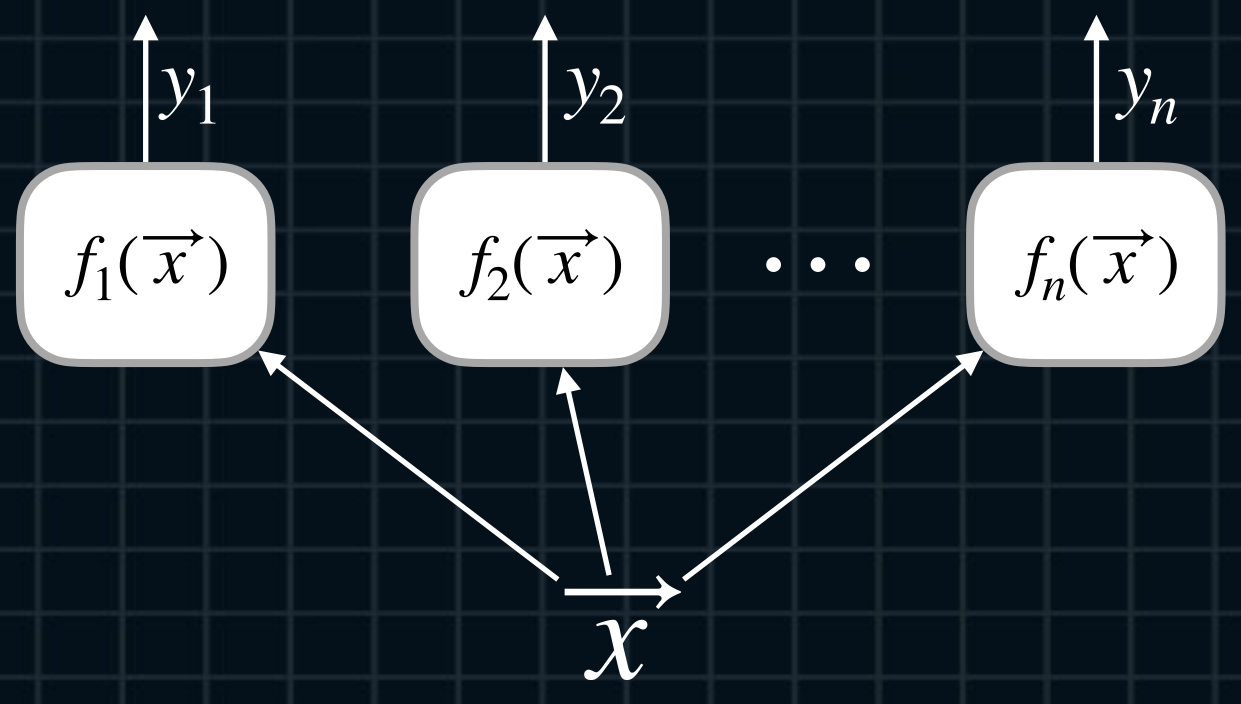
$$\vec{f}(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_n(\vec{x}) \end{pmatrix}$$

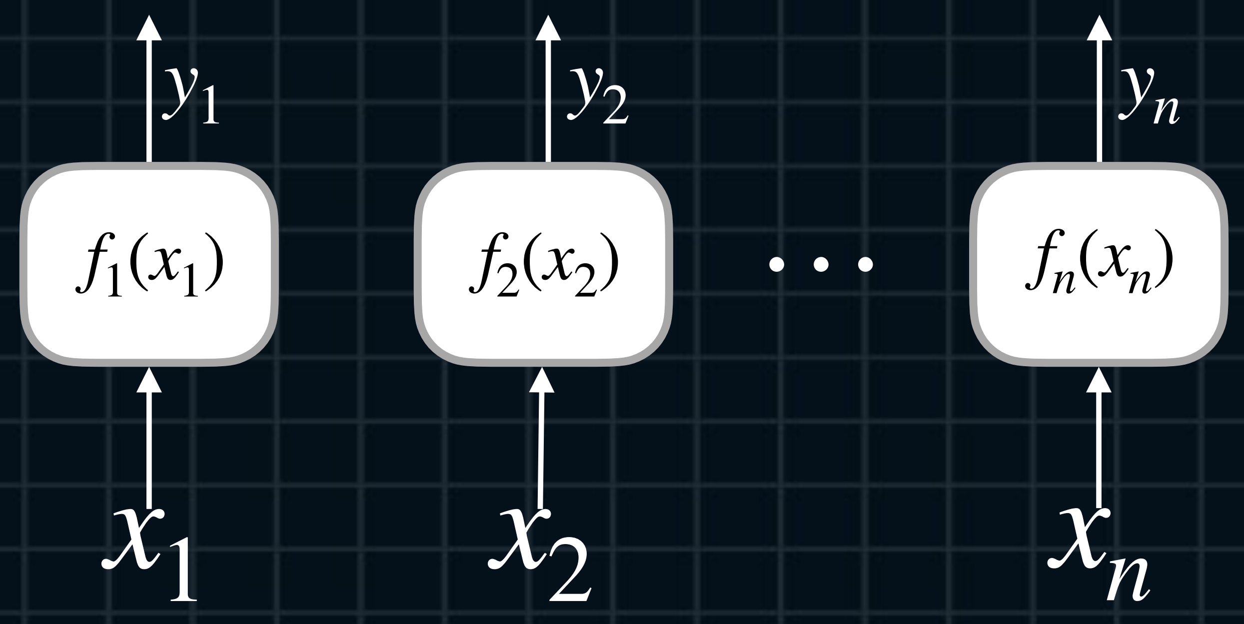
$$f_i(\vec{x}) = f_i(x_i)$$

$$\vec{f}(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_n(\vec{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1) \\ f_2(x_2) \\ \vdots \\ f_n(x_n) \end{pmatrix}$$

Lecture.6 Element-wise Operations and Jacobians - Unary Element-wise Operations

General Case





Example

$$f_i(\vec{x}) = x_i \qquad \vec{f}(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_n(\vec{x}) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$\frac{\partial f_1(\vec{x})}{\partial \vec{x}} = \frac{\partial f_1(x_1)}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial f_1(x_1)}{\partial x_1} & \frac{\partial f_1(x_1)}{\partial x_2} & \dots & \frac{\partial f_1(x_1)}{\partial x_n} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial [x_1]}{\partial x_1} & \frac{\partial [x_1]}{\partial x_2} & \dots & \frac{\partial [x_1]}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial f_1(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial f_2(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & 1 & \dots & 0 \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial f_n(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & 1 \end{pmatrix}$$

$$\frac{\partial f_i(\vec{x})}{\partial \vec{x}} = \frac{\partial f_i(x_i)}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}$$

Example

$$\frac{\partial f_1(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial f_2(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & 1 & \dots & 0 \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial f_n(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & 1 \end{pmatrix}$$

$$\frac{\partial f_i(\vec{x})}{\partial \vec{x}} = \frac{\partial f_i(x_i)}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial f_1(\vec{x})}{\partial \vec{x}} \\ \frac{\partial f_2(\vec{x})}{\partial \vec{x}} \\ \vdots \\ \frac{\partial f_n(\vec{x})}{\partial \vec{x}} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(\vec{x})}{\partial x_1} & 0 & \dots & 0 \\ 0 & \frac{\partial f_2(\vec{x})}{\partial x_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial f_n(\vec{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} = diag\left(\frac{\partial f_1}{\partial x_1}, \frac{\partial f_2}{\partial x_2}, \dots, \frac{\partial f_n}{\partial x_n}\right) = diag\left(1, 1, \dots, 1\right)$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} \in \mathbb{D}$$

- Unary Element-wise Operations

General Case

$$f_i(\vec{x}) = f_i(x_i)$$

$$\vec{f}(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_n(\vec{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1) \\ f_2(x_2) \\ \vdots \\ f_n(x_n) \end{pmatrix}$$

$$\frac{\partial f_1(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial f_1(x_1)}{\partial x_1} & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial f_2(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & \frac{\partial f_2(x_2)}{\partial x_2} & \dots & 0 \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial f_n(\vec{x})}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & \frac{\partial f_n(x_n)}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial f_i(\vec{x})}{\partial \vec{x}} = \frac{\partial f_i(x_i)}{\partial \vec{x}} = \begin{pmatrix} 0 & 0 & \dots & \frac{\partial f_i(x_i)}{\partial x_i} & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial f_1(\vec{x})}{\partial \vec{x}} \\ \frac{\partial f_2(\vec{x})}{\partial \vec{x}} \\ \vdots \\ \frac{\partial f_n(\vec{x})}{\partial \vec{x}} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x_1)}{\partial x_1} & 0 & \dots & 0 \\ 0 & \frac{\partial f_2(x_2)}{\partial x_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial f_n(x_n)}{\partial x_n} \end{pmatrix}$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} = diag\left(\frac{\partial f_1}{\partial x_1}, \frac{\partial f_2}{\partial x_2}, \dots, \frac{\partial f_n}{\partial x_n}\right)$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} \in \mathbb{D}$$

Exercise

$$f_i(\overrightarrow{x}) = ln(x_i) + e^{x_i}$$

$$\vec{f}(\overrightarrow{x}) = \begin{pmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \\ \vdots \\ f_n(\overrightarrow{x}) \end{pmatrix} = \begin{pmatrix} ln(x_1) + e^{x_1} \\ ln(x_2) + e^{x_2} \\ \vdots \\ ln(x_n) + e^{x_n} \end{pmatrix}$$

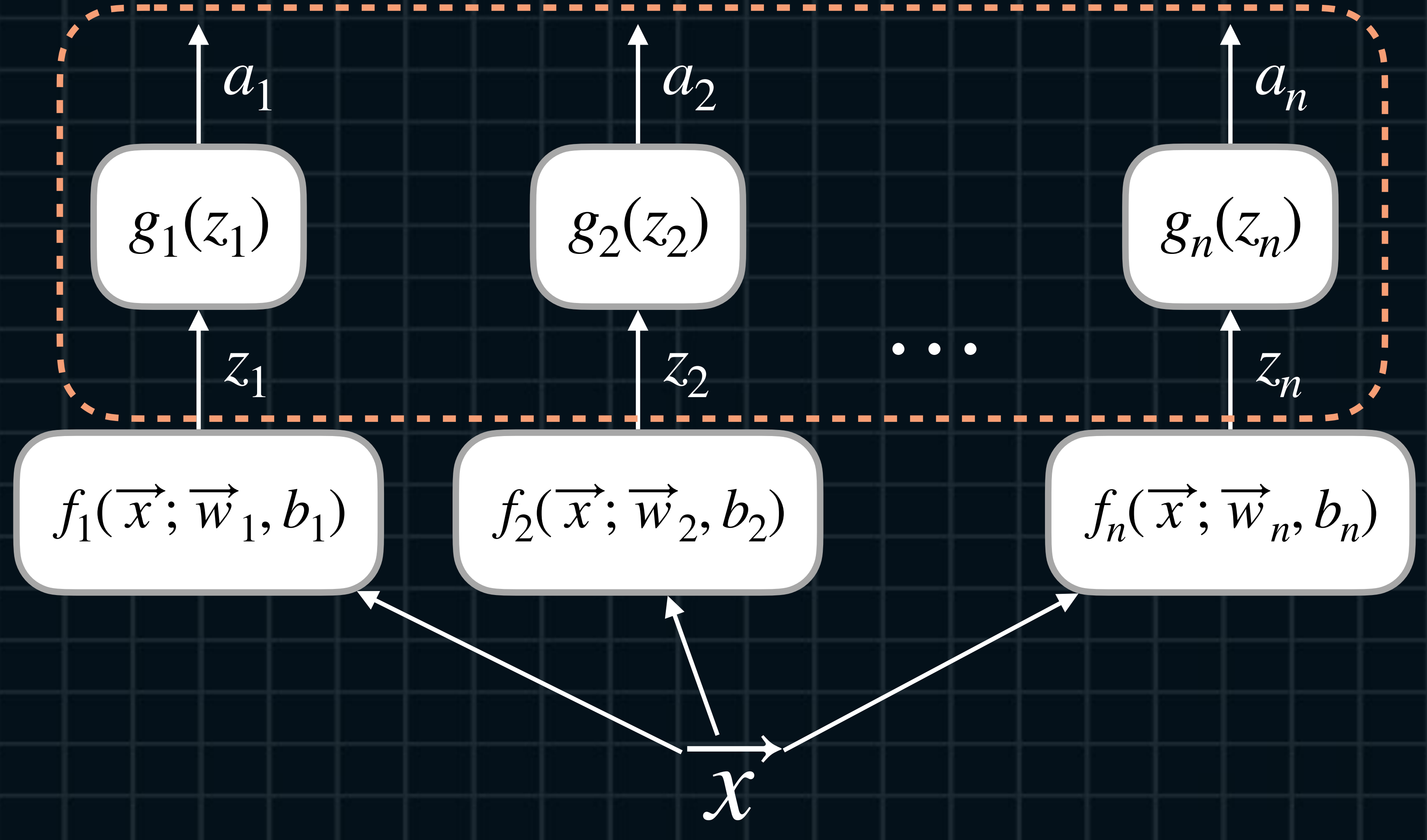
$$\frac{\partial f_i(\overrightarrow{x})}{\partial x_j} = \begin{cases} 1/x_i + e^{x_i}, & i = j \\ 0, & i \neq j \end{cases}$$

$$\frac{\partial \vec{f}}{\partial \overrightarrow{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial \overrightarrow{x}} \\ \frac{\partial f_2}{\partial \overrightarrow{x}} \\ \vdots \\ \frac{\partial f_n}{\partial \overrightarrow{x}} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}\left[ln(x_1) + e^{x_1}\right] & \frac{\partial}{\partial x_2}\left[ln(x_1) + e^{x_1}\right] & \dots & \frac{\partial}{\partial x_n}\left[ln(x_1) + e^{x_1}\right] \\ \frac{\partial}{\partial x_1}\left[ln(x_2) + e^{x_2}\right] & \frac{\partial}{\partial x_2}\left[ln(x_2) + e^{x_2}\right] & \dots & \frac{\partial}{\partial x_n}\left[ln(x_2) + e^{x_2}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\left[ln(x_n) + e^{x_n}\right] & \frac{\partial}{\partial x_2}\left[ln(x_n) + e^{x_n}\right] & \dots & \frac{\partial}{\partial x_n}\left[ln(x_n) + e^{x_n}\right] \end{pmatrix} = \begin{pmatrix} 1/x_1 + e^{x_1} & 0 & \dots & 0 \\ 0 & 1/x_2 + e^{x_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/x_n + e^{x_n} \end{pmatrix}$$

Review

$$\frac{\partial \sigma(z)}{\partial z} = a(1-a) \qquad \frac{\partial tanh(z)}{\partial z} = (1+a)(1-a) \qquad \frac{\partial ReLU(z)}{\partial z} = \begin{cases} 1, \ z \geq 0 \\ 0, \ otherwise \end{cases}$$

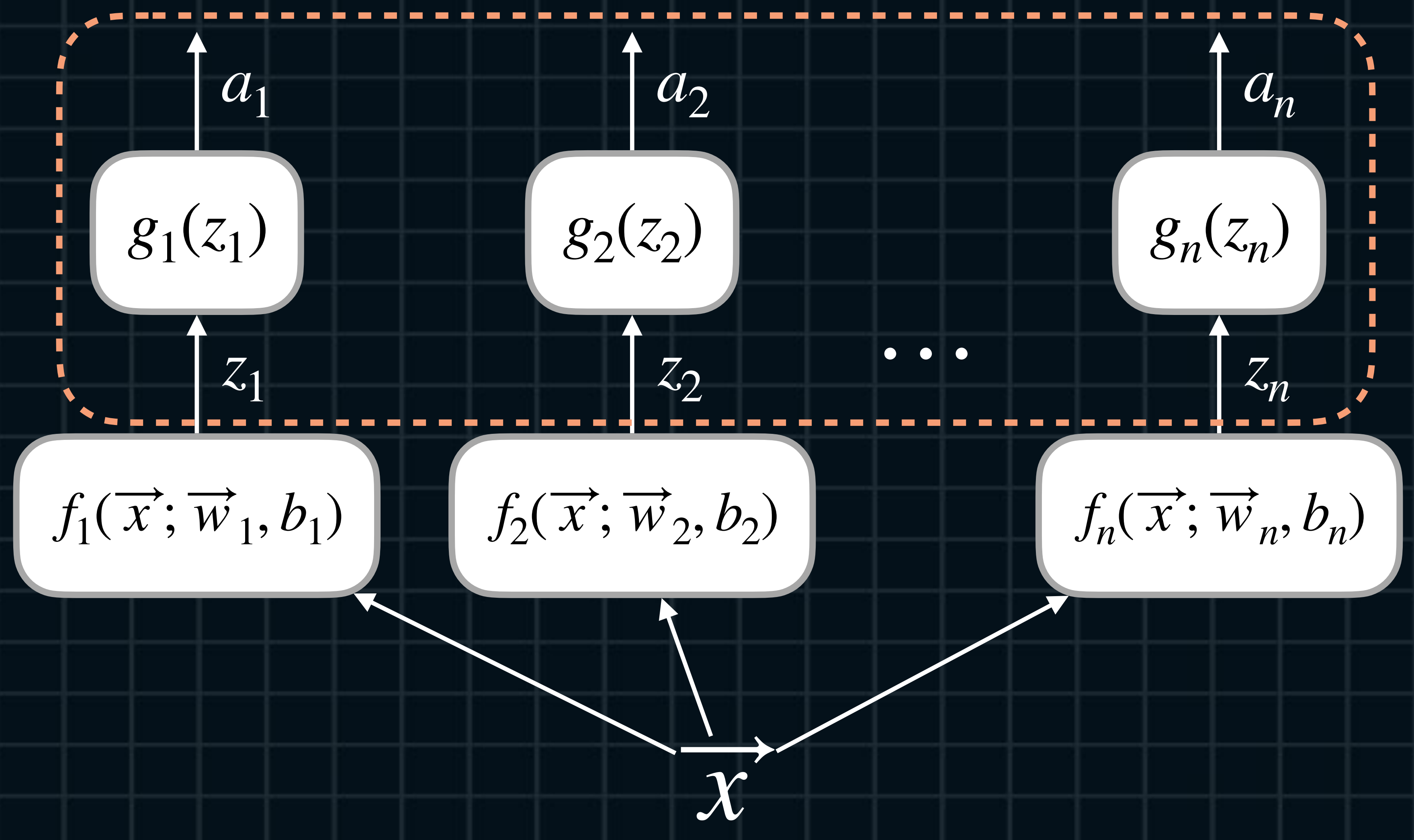
Activation Function and Element-wise Operations



$$g_i(\vec{z}) = g_i(z_i)$$

$$\vec{g}(\vec{z}) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} g_1(\vec{z}) \\ g_2(\vec{z}) \\ \vdots \\ g_n(\vec{z}) \end{pmatrix} = \begin{pmatrix} g_1(z_1) \\ g_2(z_2) \\ \vdots \\ g_n(z_n) \end{pmatrix}$$

Activation Function and Element-wise Operations



$$g_i(\vec{z}) = g_i(z_i)$$

$$\vec{g}(\vec{z}) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} g_1(\vec{z}) \\ g_2(\vec{z}) \\ \vdots \\ g_n(\vec{z}) \end{pmatrix} = \begin{pmatrix} g_1(z_1) \\ g_2(z_2) \\ \vdots \\ g_n(z_n) \end{pmatrix} \longrightarrow$$

$$\frac{\partial g_i}{\partial z_j} = \begin{cases} \frac{\partial g_i}{\partial z_i}, & i = j \\ 0, & i \neq j \end{cases} \longrightarrow$$

$$\frac{\partial g_i}{\partial \vec{z}} = \begin{pmatrix} 0 & 0 & \dots & \frac{\partial g_i}{\partial z_i} & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix}$$

Lecture.6 Element-wise Operations and Jacobians - Jacobians of Dense Layers

Sigmoid Activation Functions

$$a = g(z) = \sigma(z)$$

$$\vec{g}(\vec{z}) = \begin{pmatrix} g_1(\vec{z}) \\ g_2(\vec{z}) \\ \vdots \\ g_n(\vec{z}) \end{pmatrix} = \begin{pmatrix} g_1(z_1) \\ g_2(z_2) \\ \vdots \\ g_n(z_n) \end{pmatrix} = \begin{pmatrix} 1/(1+e^{-z_1}) \\ 1/(1+e^{-z_2}) \\ \vdots \\ 1/(1+e^{-z_n}) \end{pmatrix} \qquad \frac{\partial g_i(\vec{x})}{\partial z_j} = \begin{cases} a_i(1-a_i), & i = j \\ 0, & i \neq j \end{cases}$$

$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} a_1(1-a_1) & 0 & \dots & 0 \\ 0 & a_2(1-a_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n(1-a_n) \end{pmatrix}$$

Tanh Activation Functions

$$a = g(z) = tanh(z)$$

$$\vec{g}(\vec{z}) = \begin{pmatrix} g_1(\vec{z}) \\ g_2(\vec{z}) \\ \vdots \\ g_n(\vec{z}) \end{pmatrix} = \begin{pmatrix} g_1(z_1) \\ g_2(z_2) \\ \vdots \\ g_n(z_n) \end{pmatrix} = \begin{pmatrix} (e^{z_1} - e^{-z_1})/(e^{z_1} + e^{-z_1}) \\ (e^{z_2} - e^{-z_2})/(e^{z_2} + e^{-z_2}) \\ \vdots \\ (e^{z_n} - e^{-z_n})/(e^{z_n} + e^{-z_n}) \end{pmatrix} \quad \frac{\partial g_i(\vec{x})}{\partial z_j} = \begin{cases} (1 + a_i)(1 - a_i), & i = j \\ 0, & i \neq j \end{cases}$$

$$\frac{\partial \vec{g}}{\partial \vec{z}} = \begin{pmatrix} (1 + a_1)(1 - a_1) & 0 & \dots & 0 \\ 0 & (1 + a_2)(1 - a_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1 + a_n)(1 - a_n) \end{pmatrix}$$

ReLU Activation Functions

$$a = g(z) = ReLU(z)$$

$$\vec{g}(\vec{z}) = \begin{pmatrix} g_1(\vec{z}) \\ g_2(\vec{z}) \\ \vdots \\ g_n(\vec{z}) \end{pmatrix} = \begin{pmatrix} g_1(z_1) \\ g_2(z_2) \\ \vdots \\ g_n(z_n) \end{pmatrix} = \begin{pmatrix} max(0, z_1) \\ max(0, z_2) \\ \vdots \\ max(0, z_n) \end{pmatrix}$$

$$\frac{\partial g_i(\vec{x})}{\partial z_j} = \begin{cases} 1, & i = j \ \& \ z_i \geq 0 \\ 0, & i \neq j \ \| \ z_i < 0 \end{cases}$$

$$\frac{\partial \vec{g}}{\partial \vec{z}} = \begin{pmatrix} \begin{cases} 1, & z_1 \geq 0 \\ 0, & z_1 < 0 \end{cases} & 0 & \cdots & 0 \\ 0 & \begin{cases} 1, & z_2 \geq 0 \\ 0, & z_2 < 0 \end{cases} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{cases} 1, & z_n \geq 0 \\ 0, & z_n < 0 \end{cases} \end{pmatrix}$$

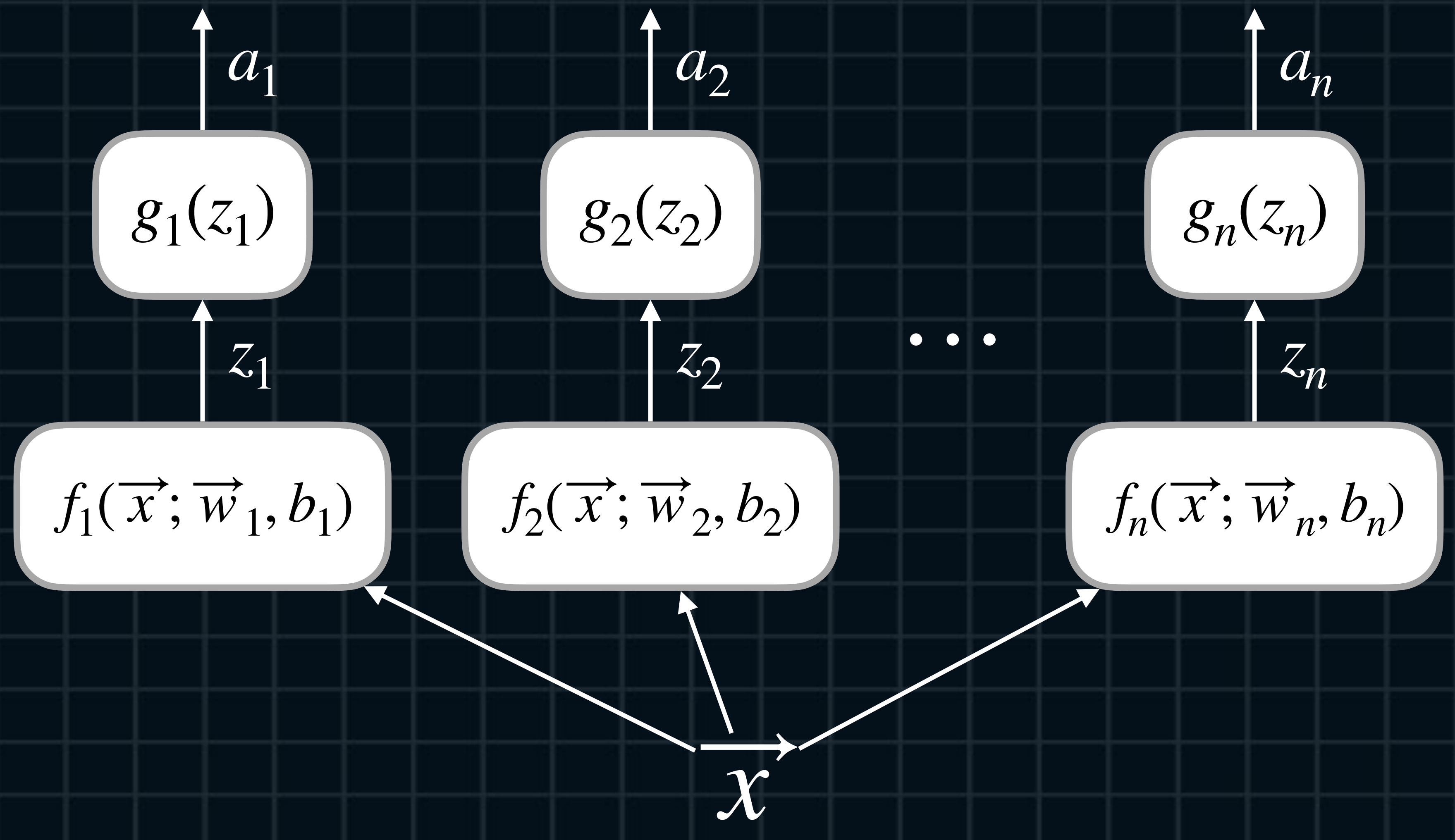
Jacobians of Dense Layers

$$\frac{\partial \vec{z}}{\partial \vec{w}_1} = \begin{pmatrix} x_1 & x_2 & \dots & x_{l_I} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{w}_2} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ x_1 & x_2 & \dots & x_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial \vec{z}}{\partial \vec{w}_n} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \dots & x_{l_I} \end{pmatrix}$$



$$\frac{\partial \vec{z}}{\partial \vec{x}} = \begin{pmatrix} \longleftarrow & (\vec{w}_1)^T & \longrightarrow \\ \longleftarrow & (\vec{w}_2)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{w}_n)^T & \longrightarrow \end{pmatrix} = W^T$$

$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{b}} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Backpropagation within Dense Layers

$$\frac{\partial \vec{a}}{\partial \vec{w}_1} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{w}_1} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots & x_{l_I} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1}\cdot x_1 & \frac{\partial a_1}{\partial z_1}\cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1}\cdot x_{l_I} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{w}_2} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{w}_2} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 0 \\ x_1 & x_2 & \dots & x_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \frac{\partial a_1}{\partial z_1}\cdot x_1 & \frac{\partial a_1}{\partial z_1}\cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1}\cdot x_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial \vec{a}}{\partial \vec{w}_n} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{w}_n} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix} \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \dots & x_{l_I} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_1}{\partial z_1}\cdot x_1 & \frac{\partial a_1}{\partial z_1}\cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1}\cdot x_{l_I} \end{pmatrix}$$

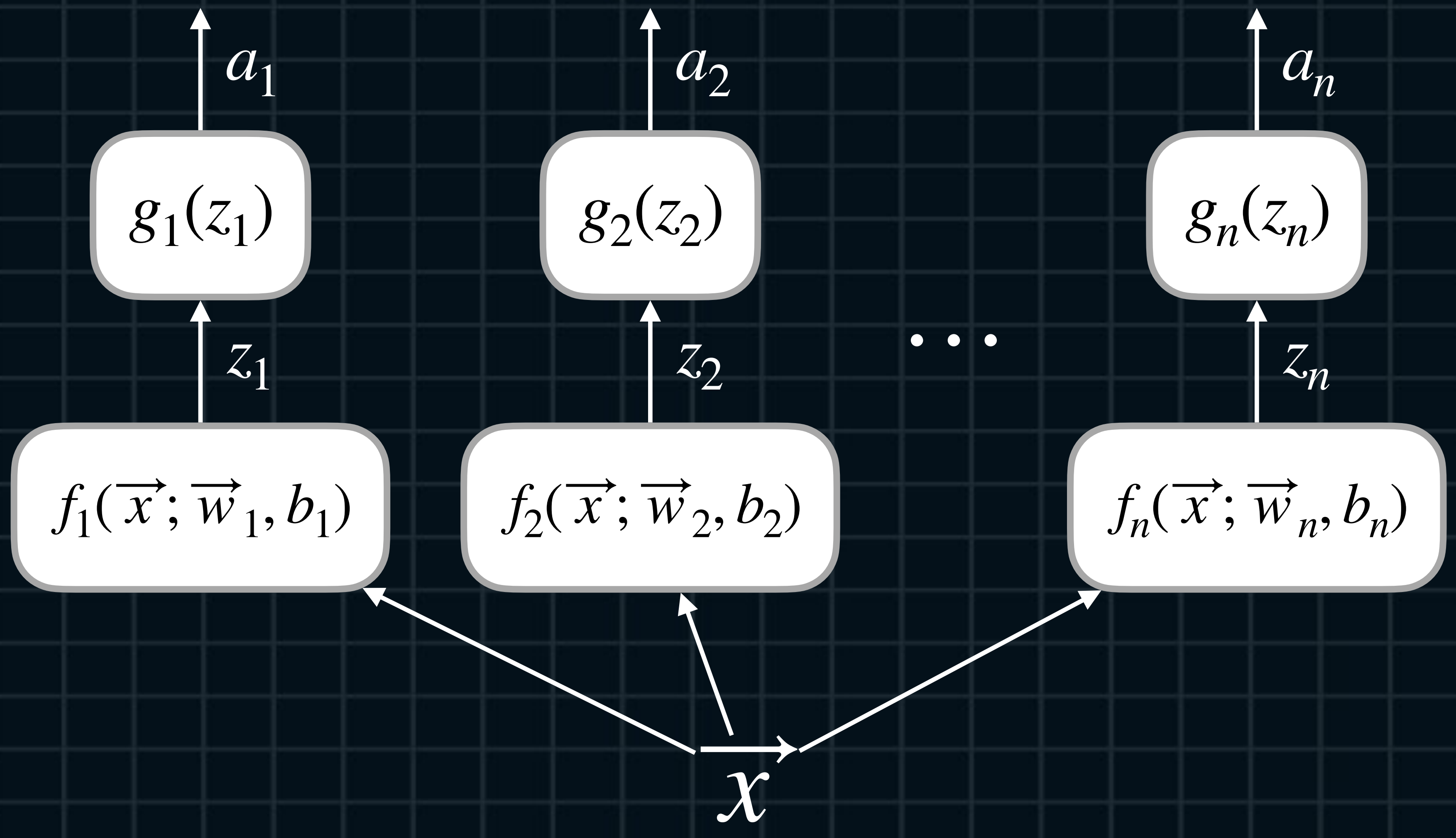
$$\frac{\partial \vec{a}}{\partial \vec{b}} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{b}} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix}$$

Backpropagation within Dense Layers

$$\frac{\partial \vec{a}}{\partial \vec{x}} = \frac{\partial \vec{a}}{\partial \vec{z}} \frac{\partial \vec{z}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix} \begin{pmatrix} \longleftarrow & (\vec{w}_1)^T & \longrightarrow \\ \longleftarrow & (\vec{w}_2)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{w}_n)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} \longleftarrow & \frac{\partial a_1}{\partial z_1} \cdot (\vec{w}_1)^T & \longrightarrow \\ \longleftarrow & \frac{\partial a_2}{\partial z_2} \cdot (\vec{w}_2)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \frac{\partial a_n}{\partial z_n} \cdot (\vec{w}_n)^T & \longrightarrow \end{pmatrix}$$

Lecture.6 Element-wise Operations and Jacobians - Jacobians of Activation Functions

Backpropagation within Dense Layers



$$\frac{\partial \vec{a}}{\partial \vec{w}_1} = \begin{pmatrix} \nabla_{\vec{w}_1} a_1 \\ \nabla_{\vec{w}_1} a_2 \\ \vdots \\ \nabla_{\vec{w}_1} a_n \end{pmatrix} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} \cdot x_1 & \frac{\partial a_1}{\partial z_1} \cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1} \cdot x_{l_I} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{w}_2} = \begin{pmatrix} \nabla_{\vec{w}_2} a_1 \\ \nabla_{\vec{w}_2} a_2 \\ \vdots \\ \nabla_{\vec{w}_2} a_n \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \frac{\partial a_1}{\partial z_1} \cdot x_1 & \frac{\partial a_1}{\partial z_1} \cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1} \cdot x_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

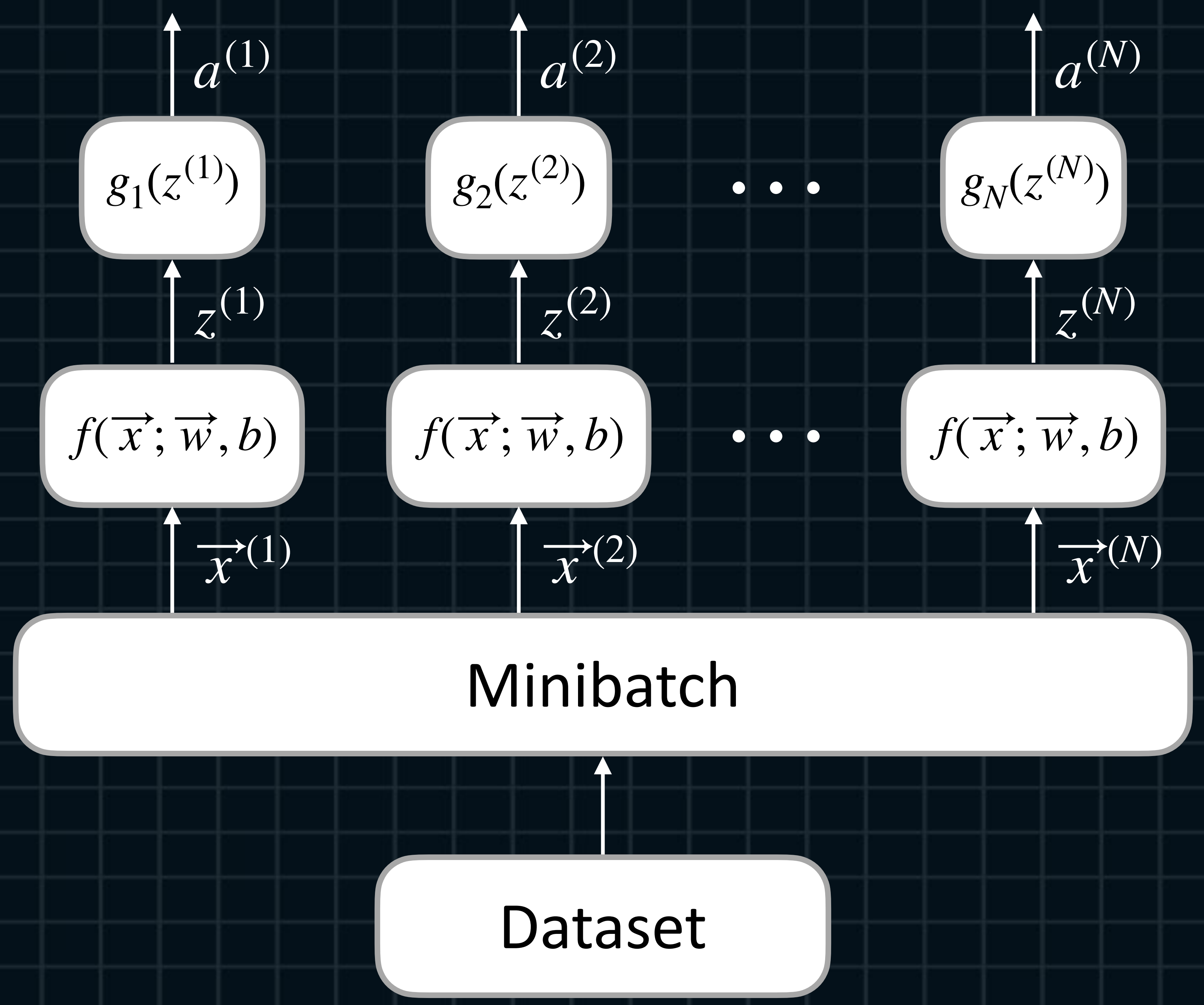
$$\vdots$$

$$\frac{\partial \vec{a}}{\partial \vec{w}_n} = \begin{pmatrix} \nabla_{\vec{w}_n} a_1 \\ \nabla_{\vec{w}_n} a_2 \\ \vdots \\ \nabla_{\vec{w}_n} a_n \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_1}{\partial z_1} \cdot x_1 & \frac{\partial a_1}{\partial z_1} \cdot x_2 & \dots & \frac{\partial a_1}{\partial z_1} \cdot x_{l_I} \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{b}} = \begin{pmatrix} \frac{\partial a_1}{\partial b_1} & \frac{\partial a_1}{\partial b_2} & \dots & \frac{\partial a_1}{\partial b_n} \\ \frac{\partial a_2}{\partial b_1} & \frac{\partial a_2}{\partial b_2} & \dots & \frac{\partial a_2}{\partial b_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_n}{\partial b_1} & \frac{\partial a_n}{\partial b_2} & \dots & \frac{\partial a_n}{\partial b_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial a_1}{\partial z_1} & 0 & \dots & 0 \\ 0 & \frac{\partial a_2}{\partial z_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\partial a_n}{\partial z_n} \end{pmatrix}$$

$$\begin{pmatrix} \nabla_{\vec{x}} a_1 \\ \nabla_{\vec{x}} a_2 \\ \vdots \\ \nabla_{\vec{x}} a_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & \frac{\partial a_1}{\partial z_1} \cdot (\vec{w}_1)^T & \longrightarrow \\ \longleftarrow & \frac{\partial a_2}{\partial z_2} \cdot (\vec{w}_2)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \frac{\partial a_n}{\partial z_n} \cdot (\vec{w}_n)^T & \longrightarrow \end{pmatrix}$$

Activation Functions and Mini-batches



$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix} \qquad \vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}$$

$$g_i(\vec{z}) = g_i(z_i) = g_i(z^{(i)})$$

$$\frac{\partial g_i}{\partial z_j} = \begin{cases} \frac{\partial a^{(i)}}{\partial z^{(i)}}, & i = j \\ 0, & i \neq j \end{cases}$$

Activation Functions and Mini-batches

$$\vec{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix} \qquad \vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}$$

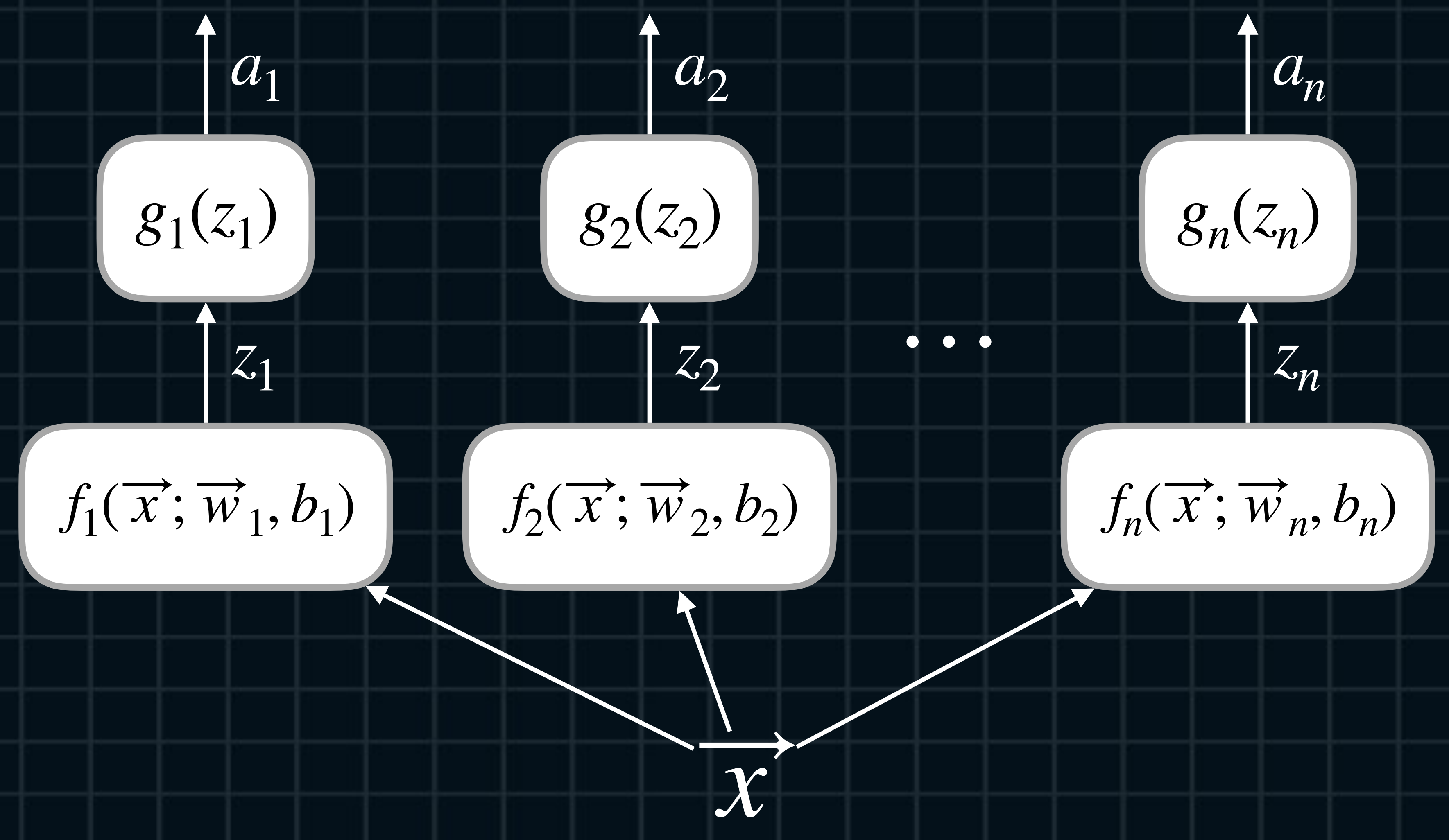
$$g_i(\vec{z}) = g_i(z_i) = g_i(z^{(i)})$$

$$\frac{\partial g_i}{\partial z_j} = \begin{cases} \frac{\partial a^{(i)}}{\partial z^{(i)}}, & i = j \\ 0, & i \neq j \end{cases}$$

$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} \partial a_1 / \partial \vec{z} \\ \partial a_2 / \partial \vec{z} \\ \vdots \\ \partial a_N / \partial \vec{z} \end{pmatrix} = \begin{pmatrix} \partial a^{(1)}/\partial z^{(1)} & \partial a^{(1)}/\partial z^{(2)} & \dots & \partial a^{(1)}/\partial z^{(N)} \\ \partial a^{(2)}/\partial z^{(1)} & \partial a^{(2)}/\partial z^{(2)} & \dots & \partial a^{(2)}/\partial z^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \partial a^{(N)}/\partial z^{(1)} & \partial a^{(N)}/\partial z^{(2)} & \dots & \partial a^{(N)}/\partial z^{(N)} \end{pmatrix}$$

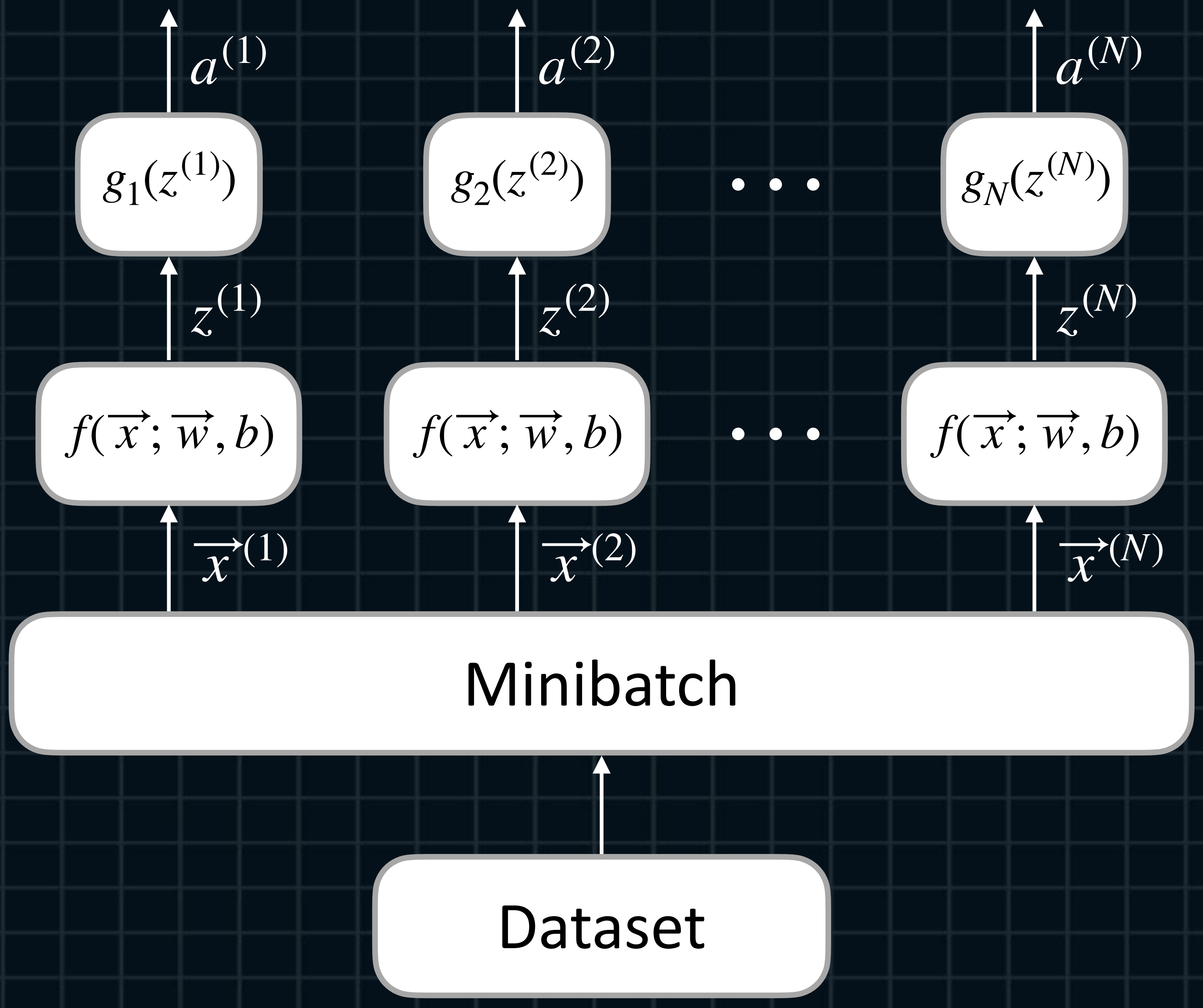
$$= \begin{pmatrix} a^{(1)}(1 - a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1 - a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1 - a^{(N)}) \end{pmatrix}, if \;\; g = \sigma$$

Comparison



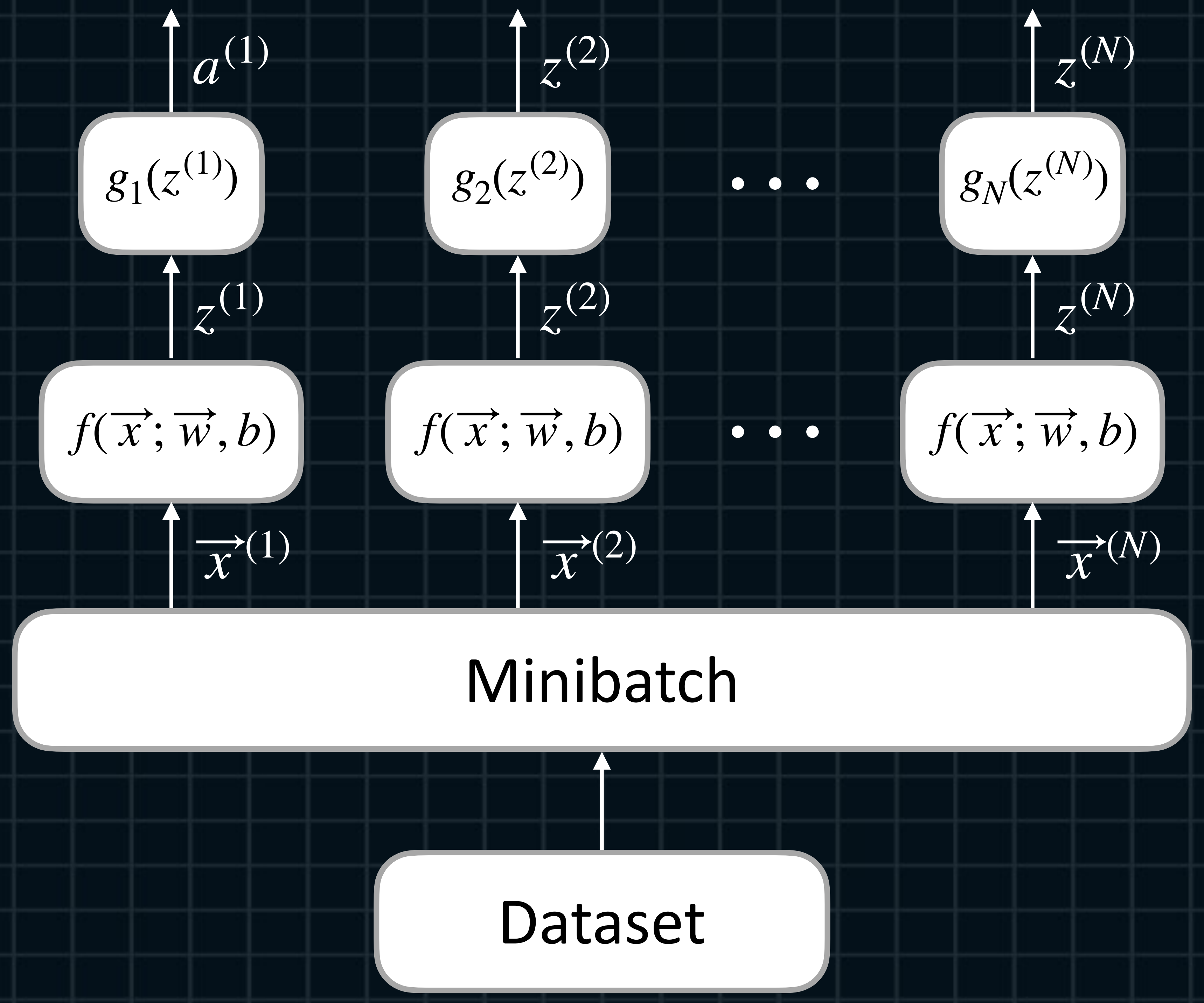
$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} a_1(1-a_1) & 0 & \dots & 0 \\ 0 & a_2(1-a_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n(1-a_n) \end{pmatrix}$$

VS



$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix}$$

Jacobians of Artificial Neuron



$$\frac{\partial \vec{a}}{\partial \vec{z}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{w}} = \begin{pmatrix} \nabla_{\vec{w}}\, z^{(1)} \\ \nabla_{\vec{w}}\, z^{(2)} \\ \vdots \\ \nabla_{\vec{w}}\, z^{(N)} \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = X^T$$

$$\frac{\partial \vec{z}}{\partial b} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial \vec{x}^{(1)}} = \begin{pmatrix} \longleftarrow & \vec{w}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} \quad \frac{\partial \vec{z}}{\partial \vec{x}^{(2)}} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{w}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} \quad \cdots \quad \frac{\partial \vec{z}}{\partial \vec{x}^{(N)}} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{w}^T & \longrightarrow \end{pmatrix}$$

Backpropagation within Artificial Neuron

$$\frac{\partial \vec{a}}{\partial \vec{w}} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{w}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix} \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} \longleftarrow & a^{(1)}(1-a^{(1)}) \cdot \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & a^{(2)}(1-a^{(2)}) \cdot \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & a^{(N)}(1-a^{(N)}) \cdot \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial b} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial b} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) \\ a^{(2)}(1-a^{(2)}) \\ \vdots \\ a^{(N)}(1-a^{(N)}) \end{pmatrix}$$

Backpropagation within Artificial Neuron

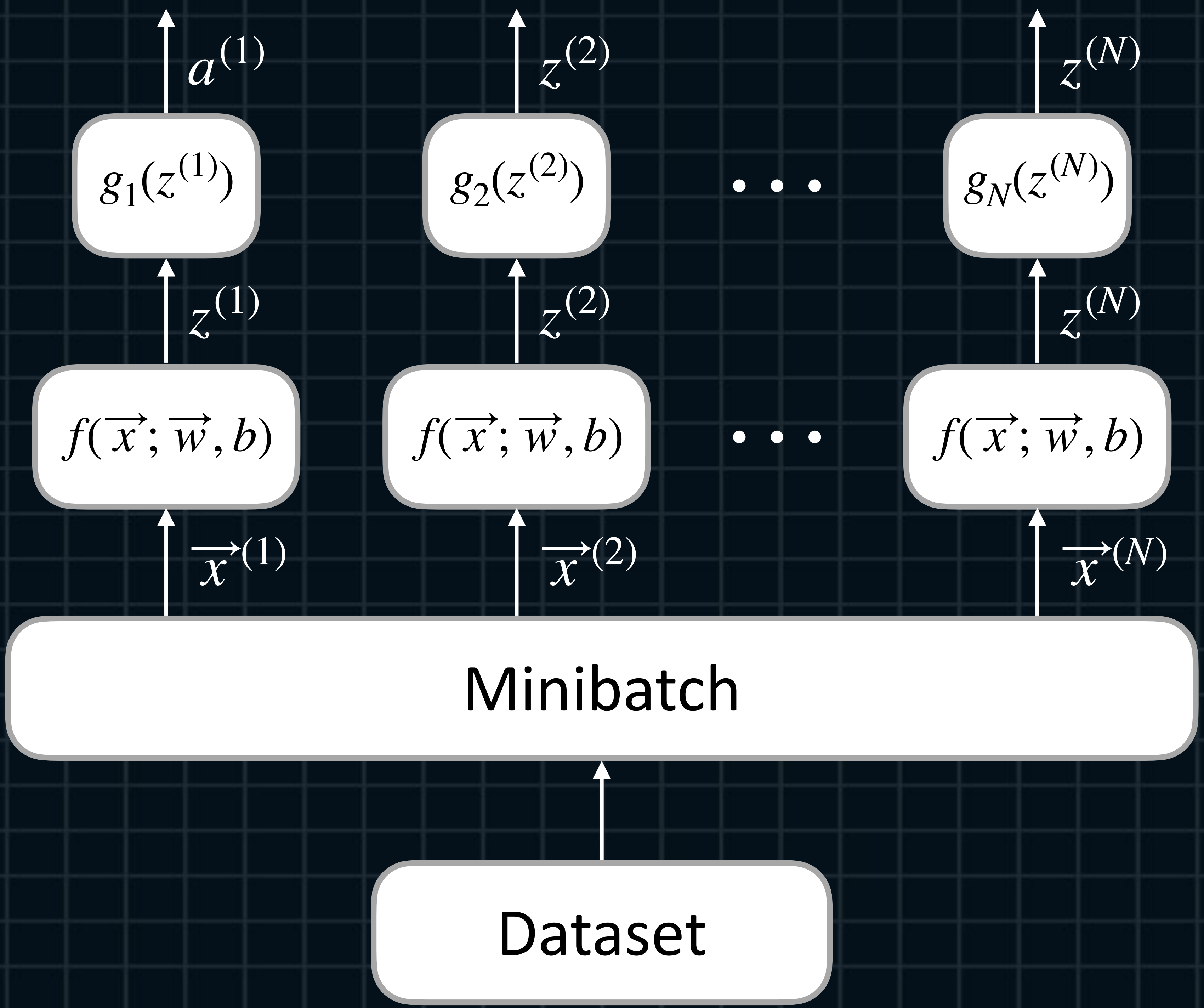
$$\frac{\partial \vec{a}}{\partial \vec{x}^{(1)}} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{x}^{(1)}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix} \begin{pmatrix} \longleftarrow & \vec{w}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} \longleftarrow & a^{(1)}(1-a^{(1)}) \cdot \vec{w}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{x}^{(2)}} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{x}^{(2)}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix} \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{w}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & a^{(2)}(1-a^{(2)}) \cdot \vec{w}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial \vec{a}}{\partial \vec{x}^{(N)}} = \frac{\partial \vec{a}}{\partial \vec{z}}\frac{\partial \vec{z}}{\partial \vec{x}^{(N)}} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) & 0 & \dots & 0 \\ 0 & a^{(2)}(1-a^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a^{(N)}(1-a^{(N)}) \end{pmatrix} \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{w}^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & a^{(N)}(1-a^{(N)}) \cdot \vec{w}^T & \longrightarrow \end{pmatrix}$$

Backpropagation within Artificial Neuron



$$\frac{\partial \vec{a}}{\partial \vec{w}} = \begin{pmatrix} \nabla_{\vec{w}}\, a^{(1)} \\ \nabla_{\vec{w}}\, a^{(2)} \\ \vdots \\ \nabla_{\vec{w}}\, a^{(N)} \end{pmatrix} = \begin{pmatrix} \longleftarrow & a^{(1)}(1-a^{(1)}) \cdot \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & a^{(2)}(1-a^{(2)}) \cdot \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & a^{(N)}(1-a^{(N)}) \cdot \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial b} = \begin{pmatrix} \frac{\partial a^{(1)}}{\partial b} \\ \frac{\partial a^{(2)}}{\partial b} \\ \vdots \\ \frac{\partial a^{(N)}}{\partial b} \end{pmatrix} = \begin{pmatrix} a^{(1)}(1-a^{(1)}) \\ a^{(2)}(1-a^{(2)}) \\ \vdots \\ a^{(N)}(1-a^{(N)}) \end{pmatrix}$$

$$\frac{\partial \vec{a}}{\partial \vec{x}^{(1)}} = \begin{pmatrix} \longleftarrow & a^{(1)}(1-a^{(1)}) \cdot \vec{w}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} \quad \frac{\partial \vec{a}}{\partial \vec{x}^{(2)}} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & a^{(2)}(1-a^{(2)}) \cdot \vec{w}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \vec{0}^T & \longrightarrow \end{pmatrix} \quad \cdots \quad \frac{\partial \vec{a}}{\partial \vec{x}^{(N)}} = \begin{pmatrix} \longleftarrow & \vec{0}^T & \longrightarrow \\ \longleftarrow & \vec{0}^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & a^{(N)}(1-a^{(N)}) \cdot \vec{w}^T & \longrightarrow \end{pmatrix}$$

- Binary Element-wise Operations

General Case

$$\vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$f_i(\vec{x}, \vec{y}) = f_i(x_i, y_i)$$

$$\vec{f}(\vec{x}, \vec{y}) = \begin{pmatrix} f_1(\vec{x}, \vec{y}) \\ f_2(\vec{x}, \vec{y}) \\ \vdots \\ f_n(\vec{x}, \vec{y}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, y_1) \\ f_1(x_2, y_2) \\ \vdots \\ f_1(x_n, y_n) \end{pmatrix}$$

$$\frac{\partial \vec{f}}{\partial \vec{x}} \qquad \frac{\partial \vec{f}}{\partial \vec{y}}$$

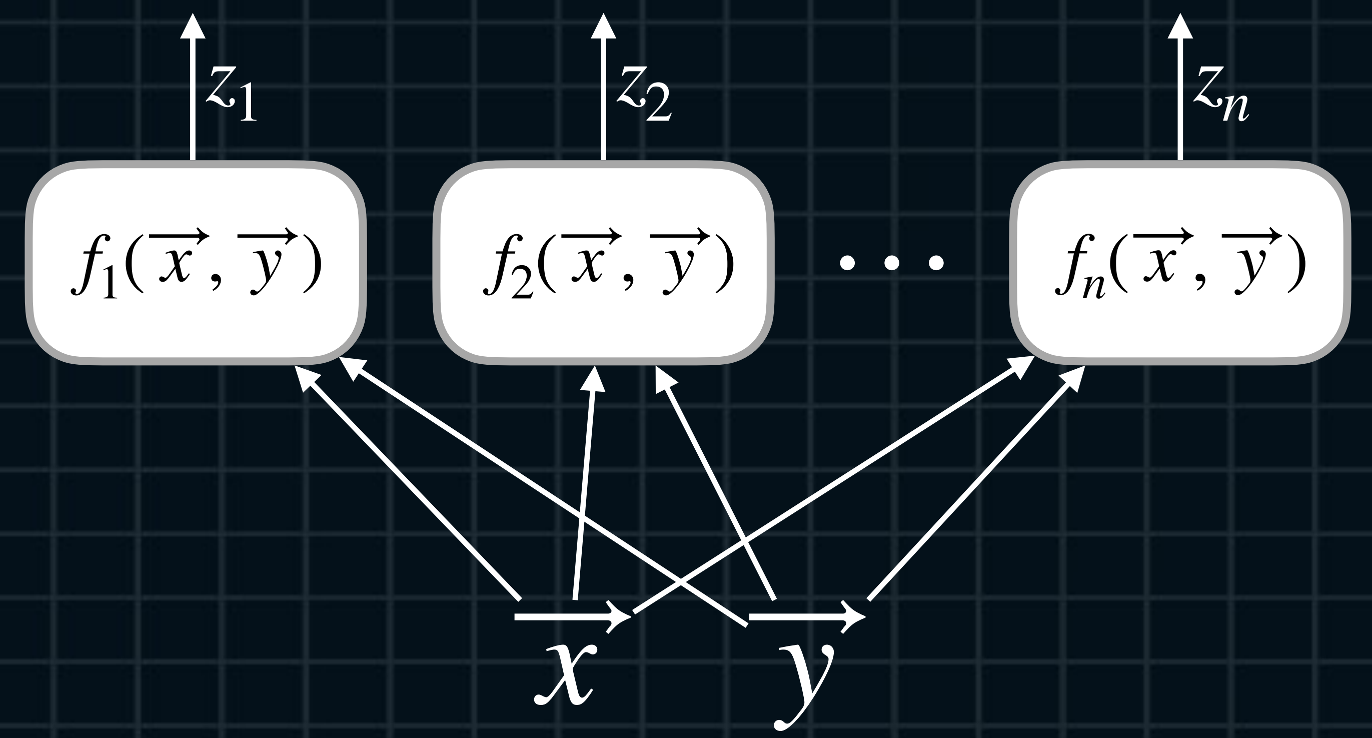
$$\Longrightarrow$$

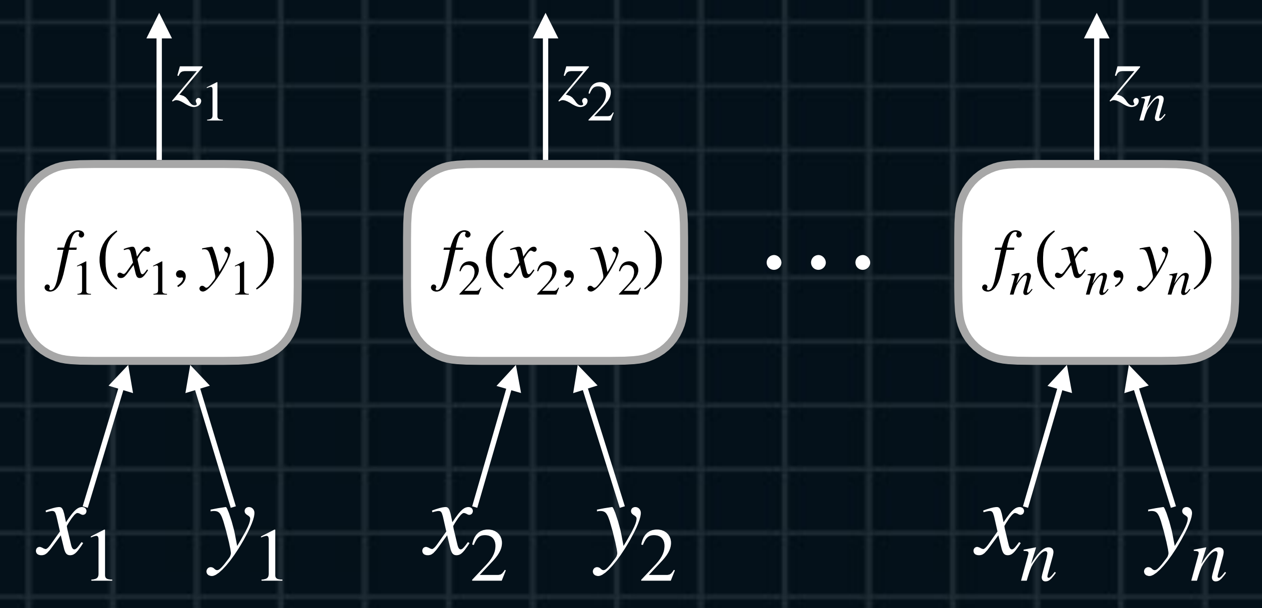
$$\frac{\partial \vec{f}}{\partial \vec{x}} = \begin{pmatrix} \nabla_{\vec{x}} f_1 \\ \nabla_{\vec{x}} f_2 \\ \vdots \\ \nabla_{\vec{x}} f_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x_1, y_1)}{\partial x_1} & \frac{\partial f_1(x_1, y_1)}{\partial x_2} & \cdots & \frac{\partial f_1(x_1, y_1)}{\partial x_n} \\ \frac{\partial f_2(x_2, y_2)}{\partial x_1} & \frac{\partial f_2(x_2, y_2)}{\partial x_2} & \cdots & \frac{\partial f_2(x_2, y_2)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(x_n, y_n)}{\partial x_1} & \frac{\partial f_n(x_n, y_n)}{\partial x_2} & \cdots & \frac{\partial f_n(x_n, y_n)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x_1, y_1)}{\partial x_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial f_2(x_2, y_2)}{\partial x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial f_n(x_n, y_n)}{\partial x_n} \end{pmatrix} \in \mathbb{D}$$

$$\frac{\partial \vec{f}}{\partial \vec{y}} = \begin{pmatrix} \nabla_{\vec{y}} f_1 \\ \nabla_{\vec{y}} f_2 \\ \vdots \\ \nabla_{\vec{y}} f_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x_1, y_1)}{\partial y_1} & \frac{\partial f_1(x_1, y_1)}{\partial y_2} & \cdots & \frac{\partial f_1(x_1, y_1)}{\partial y_n} \\ \frac{\partial f_2(x_2, y_2)}{\partial y_1} & \frac{\partial f_2(x_2, y_2)}{\partial y_2} & \cdots & \frac{\partial f_2(x_2, y_2)}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(x_n, y_n)}{\partial y_1} & \frac{\partial f_n(x_n, y_n)}{\partial y_2} & \cdots & \frac{\partial f_n(x_n, y_n)}{\partial y_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x_1, y_1)}{\partial y_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial f_2(x_2, y_2)}{\partial y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial f_n(x_n, y_n)}{\partial y_n} \end{pmatrix} \in \mathbb{D}$$

- Binary Element-wise Operations

General Case





Exercises

$$\vec{w} = \vec{u} + \vec{v} \qquad w_i = u_i + v_i$$

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}$$

$$\frac{\partial \vec{w}}{\partial \vec{u}} = \begin{pmatrix} \frac{\partial[u_1+v_1]}{\partial u_1} & \frac{\partial[u_1+v_1]}{\partial u_2} & \dots & \frac{\partial[u_1+v_1]}{\partial u_n} \\ \frac{\partial[u_2+v_2]}{\partial u_1} & \frac{\partial[u_2+v_2]}{\partial u_2} & \dots & \frac{\partial[u_2+v_2]}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial[u_n+v_n]}{\partial u_1} & \frac{\partial[u_n+v_n]}{\partial u_2} & \dots & \frac{\partial[u_n+v_n]}{\partial u_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = I$$

$$\frac{\partial \vec{w}}{\partial \vec{v}} = \begin{pmatrix} \frac{\partial[u_1+v_1]}{\partial v_1} & \frac{\partial[u_1+v_1]}{\partial v_2} & \dots & \frac{\partial[u_1+v_1]}{\partial v_n} \\ \frac{\partial[u_2+v_2]}{\partial v_1} & \frac{\partial[u_2+v_2]}{\partial v_2} & \dots & \frac{\partial[u_2+v_2]}{\partial v_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial[u_n+v_n]}{\partial v_1} & \frac{\partial[u_n+v_n]}{\partial v_2} & \dots & \frac{\partial[u_n+v_n]}{\partial v_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = I$$

Exercises

$$\vec{w} = \vec{u} - \vec{v} \qquad w_i = u_i - v_i$$

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}$$

$$\frac{\partial \vec{w}}{\partial \vec{u}} = \begin{pmatrix} \frac{\partial [u_1 - v_1]}{\partial u_1} & \frac{\partial [u_1 - v_1]}{\partial u_2} & \cdots & \frac{\partial [u_1 - v_1]}{\partial u_n} \\ \frac{\partial [u_2 - v_2]}{\partial u_1} & \frac{\partial [u_2 - v_2]}{\partial u_2} & \cdots & \frac{\partial [u_2 - v_2]}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [u_n - v_n]}{\partial u_1} & \frac{\partial [u_n - v_n]}{\partial u_2} & \cdots & \frac{\partial [u_n - v_n]}{\partial u_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I$$

$$\frac{\partial \vec{w}}{\partial \vec{v}} = \begin{pmatrix} \frac{\partial [u_1 - v_1]}{\partial v_1} & \frac{\partial [u_1 - v_1]}{\partial v_2} & \cdots & \frac{\partial [u_1 - v_1]}{\partial v_n} \\ \frac{\partial [u_2 - v_2]}{\partial v_1} & \frac{\partial [u_2 - v_2]}{\partial v_2} & \cdots & \frac{\partial [u_2 - v_2]}{\partial v_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [u_n - v_n]}{\partial v_1} & \frac{\partial [u_n - v_n]}{\partial v_2} & \cdots & \frac{\partial [u_n - v_n]}{\partial v_n} \end{pmatrix} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix} = -I$$

Exercises

$$\vec{w} = \vec{u} \bigcirc \vec{v} \qquad w_i = u_i v_i$$

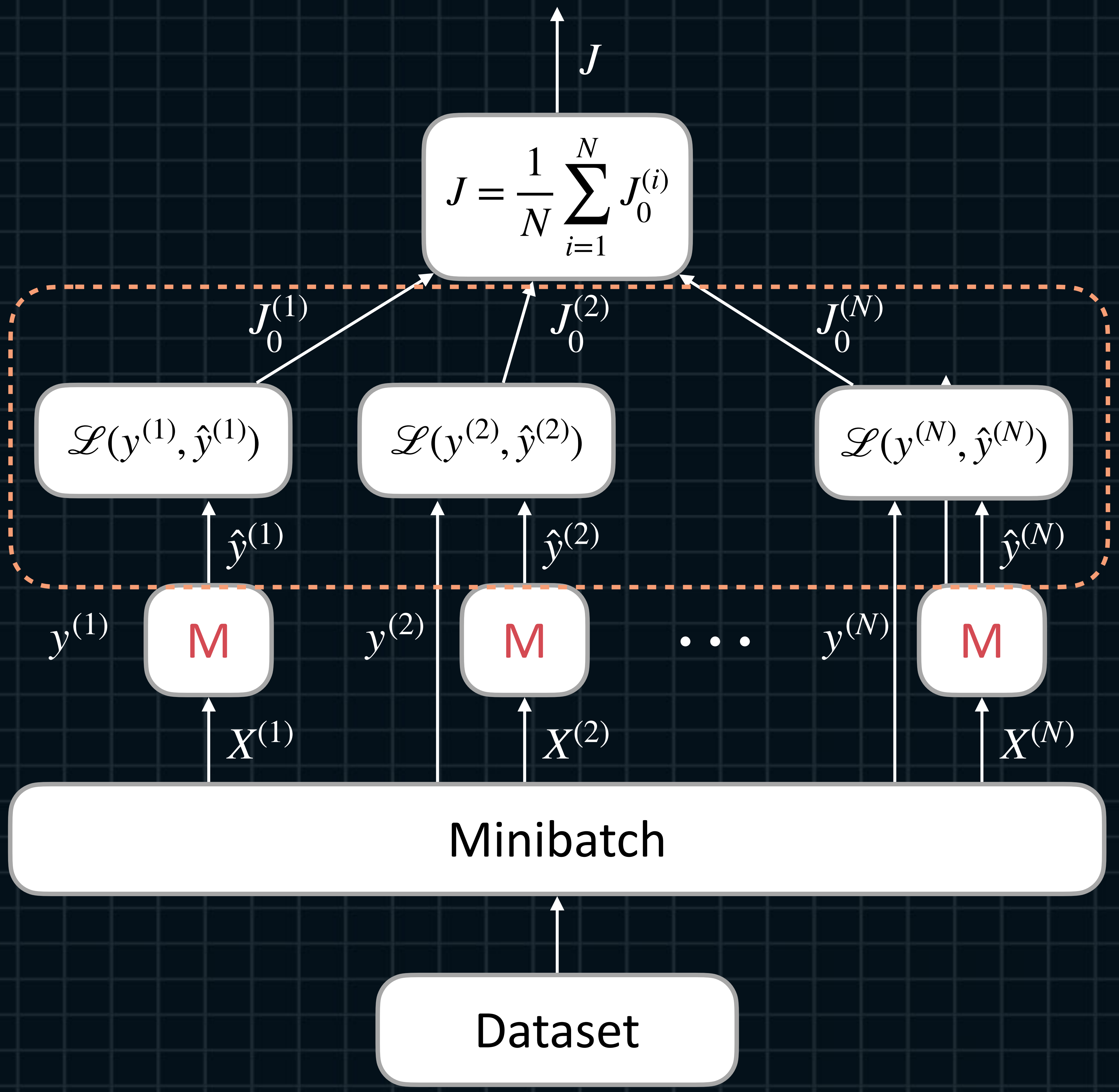
$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ u_2 v_2 \\ \vdots \\ u_n v_n \end{pmatrix}$$

$$\frac{\partial \vec{w}}{\partial \vec{u}} = \begin{pmatrix} \frac{\partial [u_1 v_1]}{\partial u_1} & \frac{\partial [u_1 v_1]}{\partial u_2} & \dots & \frac{\partial [u_1 v_1]}{\partial u_n} \\ \frac{\partial [u_2 v_2]}{\partial u_1} & \frac{\partial [u_2 v_2]}{\partial u_2} & \dots & \frac{\partial [u_2 v_2]}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [u_n v_n]}{\partial u_1} & \frac{\partial [u_n v_n]}{\partial u_2} & \dots & \frac{\partial [u_n v_n]}{\partial u_n} \end{pmatrix} = \begin{pmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_n \end{pmatrix} = diag(\vec{v})$$

$$\frac{\partial \vec{w}}{\partial \vec{v}} = \begin{pmatrix} \frac{\partial [u_1 v_1]}{\partial v_1} & \frac{\partial [u_1 v_1]}{\partial v_2} & \dots & \frac{\partial [u_1 v_1]}{\partial v_n} \\ \frac{\partial [u_2 v_2]}{\partial v_1} & \frac{\partial [u_2 v_2]}{\partial v_2} & \dots & \frac{\partial [u_2 v_2]}{\partial v_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [u_n v_n]}{\partial v_1} & \frac{\partial [u_n v_n]}{\partial v_2} & \dots & \frac{\partial [u_n v_n]}{\partial v_n} \end{pmatrix} = \begin{pmatrix} u_1 & 0 & \dots & 0 \\ 0 & u_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_n \end{pmatrix} = diag(\vec{u})$$

Lecture.6 Element-wise Operations and Jacobians - Loss Functions and Mini-batches

General Case



$$\vec{J}_0 = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix} \qquad \vec{\hat{y}} = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}$$

$$J_0^{(i)} = \mathscr{L}_i(\vec{y}, \vec{\hat{y}}) = \mathscr{L}_i(y^{(i)}, \hat{y}^{(i)})$$

$$\vec{J}_0 = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix} = \begin{pmatrix} \mathscr{L}_1(y^{(1)}, \hat{y}^{(1)}) \\ \mathscr{L}_2(y^{(2)}, \hat{y}^{(2)}) \\ \vdots \\ \mathscr{L}_N(y^{(N)}, \hat{y}^{(N)}) \end{pmatrix}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \vec{\hat{y}}} \\ \frac{\partial J_0^{(2)}}{\partial \vec{\hat{y}}} \\ \vdots \\ \frac{\partial J_0^{(N)}}{\partial \vec{\hat{y}}} \end{pmatrix} = \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(N)}} \\ \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(N)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{\partial J}{\partial \vec{J}_0}\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix} \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & 0 & \cdots & 0 \\ 0 & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{N}\frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & \frac{1}{N}\frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{1}{N}\frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix}$$

Mean Squared Error

$$J_0^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2 \quad \frac{\partial J}{\partial J_0^{(i)}} = -2(y^{(i)} - \hat{y}^{(i)})$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(N)}} \\ \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(N)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix} = \begin{pmatrix} -2(y^{(1)} - \hat{y}^{(i)}) & 0 & \cdots & 0 \\ 0 & -2(y^{(2)} - \hat{y}^{(2)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2(y^{(N)} - \hat{y}^{(N)}) \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{\partial J}{\partial \vec{J}_0} \frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix} \begin{pmatrix} -2(y^{(1)} - \hat{y}^{(i)}) & 0 & \cdots & 0 \\ 0 & -2(y^{(2)} - \hat{y}^{(2)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -2(y^{(N)} - \hat{y}^{(N)}) \end{pmatrix}$$

$$= \frac{1}{N} \begin{pmatrix} -2(y^{(1)} - \hat{y}^{(1)}) & -2(y^{(2)} - \hat{y}^{(2)}) & \cdots & -2(y^{(N)} - \hat{y}^{(N)}) \end{pmatrix}$$

$$= \frac{-2}{N}(\vec{y} - \vec{\hat{y}})$$

Binary Cross Entropy Error

$$J_0^{(i)} = -\left[y^{(i)}log\left(\hat{y}^{(i)}\right) + (1 - y^{(i)})log\left(1 - \hat{y}^{(i)}\right)\right] \qquad \frac{\partial J}{\partial J_0^{(i)}} = \frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right)}$$

$$\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(N)}} \\ \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(N)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(1)}} & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \end{pmatrix} = \begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}\left(1 - \hat{y}^{(1)}\right)} & 0 & \cdots & 0 \\ 0 & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}\left(1 - \hat{y}^{(2)}\right)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}\left(1 - \hat{y}^{(N)}\right)} \end{pmatrix}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{\partial J}{\partial \vec{J}_0}\frac{\partial \vec{J}_0}{\partial \vec{\hat{y}}} = \frac{1}{N}\left(\frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}\left(1 - \hat{y}^{(1)}\right)} \quad \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}\left(1 - \hat{y}^{(2)}\right)} \quad \cdots \quad \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}\left(1 - \hat{y}^{(N)}\right)}\right)$$

$$= \frac{-1}{N}(\vec{y} - \vec{\hat{y}}) \,/\, \vec{\hat{y}}\,(1 - \vec{\hat{y}})$$

Backpropagation and Jacobian Matrices

Lecture.6 Element-wise Operations and Jacobians