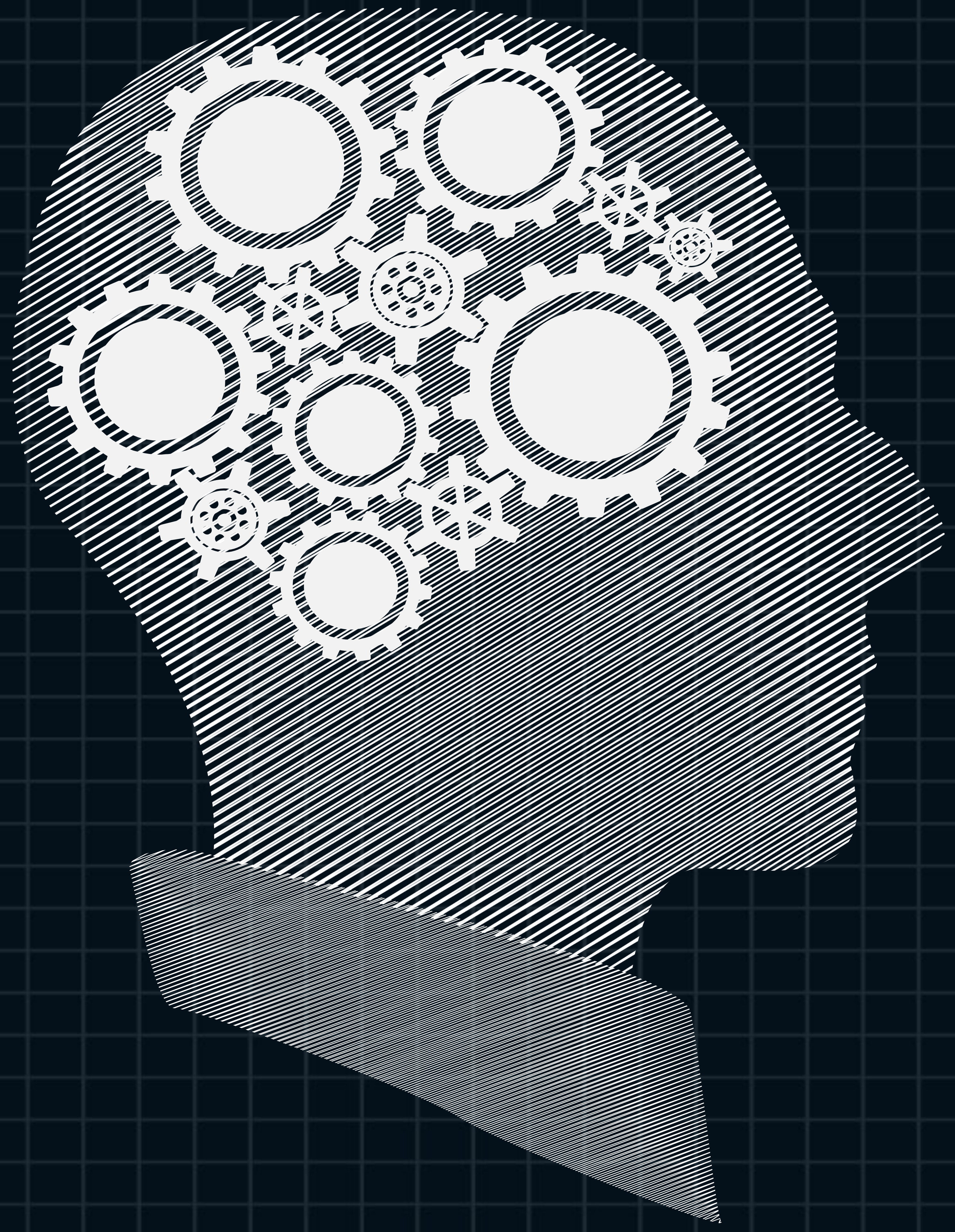
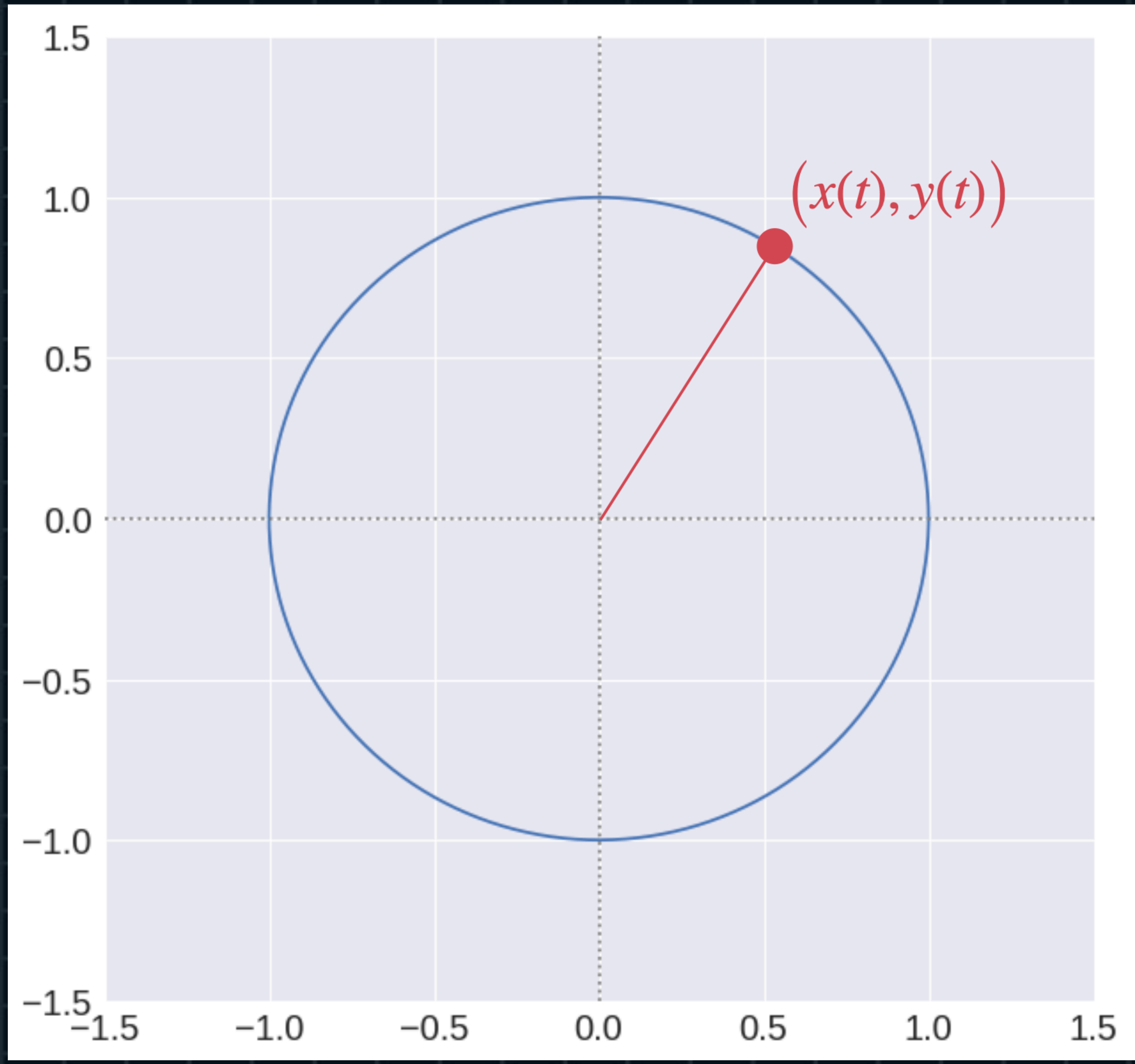


Backpropagation and Jacobian Matrices

Lecture.5 Vector Functions and Jacobians

Dots On a Circle



$$x(t) = cos(2\pi t) \qquad y(t) = sin(2\pi t)$$

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} cos(2\pi t) \\ sin(2\pi t) \end{pmatrix}$$

$$\vec{r}(0) = \begin{pmatrix} x(0) \\ y(0) \end{pmatrix} = \begin{pmatrix} cos(2\pi \cdot 0) \\ sin(2\pi \cdot 0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\vec{r}(0.25) = \begin{pmatrix} x(0.25) \\ y(0.25) \end{pmatrix} = \begin{pmatrix} cos(2\pi \cdot 0.25) \\ sin(2\pi \cdot 0.25) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\vec{r}(0.5) = \begin{pmatrix} x(0.5) \\ y(0.5) \end{pmatrix} = \begin{pmatrix} cos(2\pi \cdot 0.5) \\ sin(2\pi \cdot 0.5) \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\vec{r}(0.75) = \begin{pmatrix} x(0.75) \\ y(0.75) \end{pmatrix} = \begin{pmatrix} cos(2\pi \cdot 0.75) \\ sin(2\pi \cdot 0.75) \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Dots On a Circle

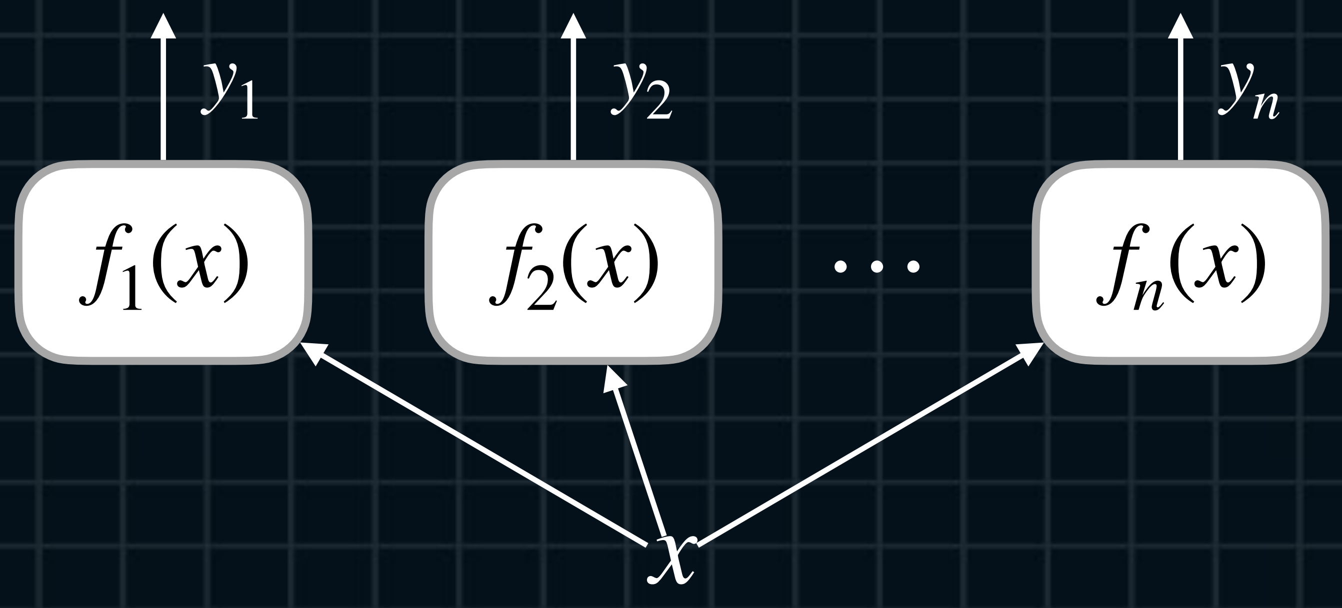
$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} cos(2\pi t) \\ sin(2\pi t) \end{pmatrix}$$

$$\vec{r}(t) : \quad \mathbb{R} \longrightarrow \mathbb{R}^2$$

$$\begin{tikzcd}
x & & y \\
\boxed{x(t) = cos(t)} \arrow[u] & & \boxed{y(t) = sin(t)} \arrow[u] \\
& t \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Vector Functions

$$\vec{f}(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix} \qquad \vec{f}(x): \quad \mathbb{R} \longrightarrow \mathbb{R}^n$$



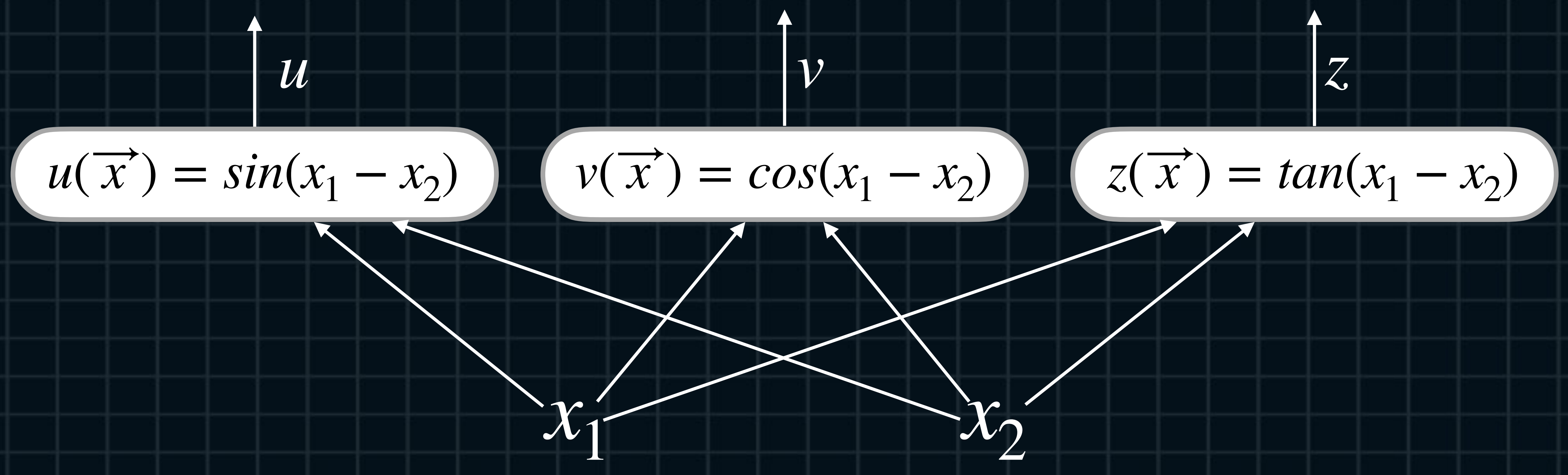
Vector Functions with Multiple Inputs

$$\left(\vec{x}\right)^T = (x_1 \quad x_2)$$

$$u(\vec{x}) = sin(x_1 - x_2)$$
$$v(\vec{x}) = cos(x_1 - x_2)$$
$$z(\vec{x}) = tan(x_1 - x_2)$$

$$\vec{r}(\vec{x}) = \begin{pmatrix} u(\vec{x}) \\ v(\vec{x}) \\ z(\vec{x}) \end{pmatrix} = \begin{pmatrix} sin(x_1 - x_2) \\ cos(x_1 - x_2) \\ tan(x_1 - x_2) \end{pmatrix}$$

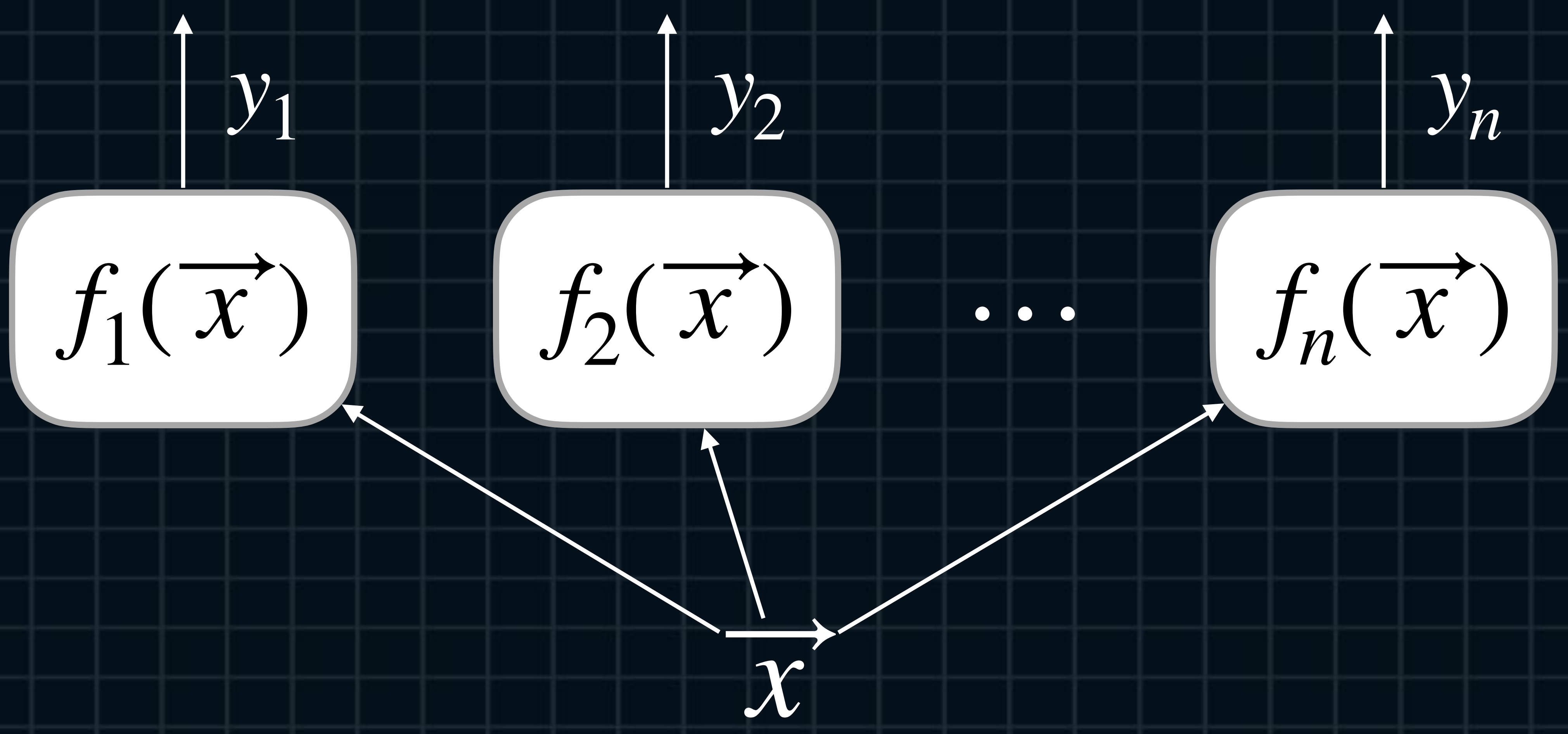
$$\vec{f}(t): \quad \mathbb{R}^2 \longrightarrow \mathbb{R}^3$$



General Vector Functions

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

$$\vec{f}(t): \quad \mathbb{R}^m \longrightarrow \mathbb{R}^n$$



Four Cases of Jacobians

$$\begin{array}{c|c|c} & x & \vec{x} \\ \hline f & \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial \vec{x}} \\ \hline \vec{f} & \dfrac{\partial \vec{f}}{\partial x} & \dfrac{\partial \vec{f}}{\partial \vec{x}} \end{array}$$

Vector Functions and Scalar Inputs

$x(t) = cos(2\pi t)$ $y(t) = sin(2\pi t)$

$$\vec{r} = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} cos(2\pi t) \\ sin(2\pi t) \end{pmatrix}$$

$$\frac{\partial x}{\partial t} = \frac{\partial cos(2\pi t)}{\partial t} = -2\pi \cdot sin(2\pi t) \qquad \frac{\partial y}{\partial t} = \frac{\partial sin(2\pi t)}{\partial t} = 2\pi \cdot cos(2\pi t)$$

$$\frac{\partial \vec{r}}{\partial t} = \begin{pmatrix} \partial x/\partial t \\ \partial y/\partial t \end{pmatrix} = \begin{pmatrix} -2\pi \cdot sin(2\pi t) \\ 2\pi \cdot cos(2\pi t) \end{pmatrix}$$

Vector Functions and Scalar Inputs

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

$$\frac{d\vec{f}}{dx} \in \mathbb{R}^{n\times 1}$$

- Jacobians of Vector Functions

Vector Functions and Vector Inputs

$$\overrightarrow{x}^T = (x_1, x_2)$$

$$u(\overrightarrow{x}) = sin(x_1 - x_2)$$
$$v(\overrightarrow{x}) = cos(x_1 - x_2)$$
$$z(\overrightarrow{x}) = tan(x_1 - x_2)$$

$$\nabla_{\overrightarrow{x}}\, u = \begin{pmatrix} \frac{\partial sin(x_1 - x_2)}{\partial x_1} & \frac{\partial sin(x_1 - x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} cos(x_1 - x_2) & -cos(x_1 - x_2) \end{pmatrix}$$

$$\nabla_{\overrightarrow{x}}\, v = \begin{pmatrix} \frac{\partial cos(x_1 - x_2)}{\partial x_1} & \frac{\partial cos(x_1 - x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} -sin(x_1 - x_2) & cos(x_1 - x_2) \end{pmatrix}$$

$$\nabla_{\overrightarrow{x}}\, z = \begin{pmatrix} \frac{\partial tan(x_1 - x_2)}{\partial x_1} & \frac{\partial tan(x_1 - x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} sec^2(x_1 - x_2) & -sec^2(x_1 - x_2) \end{pmatrix}$$

Vector Functions and Vector Inputs

$$\overrightarrow{x}^T = (x_1, x_2)$$

$$u(\overrightarrow{x}) = sin(x_1 - x_2)$$
$$v(\overrightarrow{x}) = cos(x_1 - x_2)$$
$$z(\overrightarrow{x}) = tan(x_1 - x_2)$$

$$\vec{r}(\overrightarrow{x}) = \begin{pmatrix} u(\overrightarrow{x}) \\ v(\overrightarrow{x}) \\ z(\overrightarrow{x}) \end{pmatrix}$$

$$\frac{\partial \vec{r}(\overrightarrow{x})}{\partial \overrightarrow{x}} = \begin{pmatrix} \frac{\partial u(\overrightarrow{x})}{\partial \overrightarrow{x}} \\ \frac{\partial v(\overrightarrow{x})}{\partial \overrightarrow{x}} \\ \frac{\partial z(\overrightarrow{x})}{\partial \overrightarrow{x}} \end{pmatrix}$$

Vector Functions and Vector Inputs

$$\vec{x}^T = (x_1, x_2) \qquad \begin{aligned} u(\vec{x}) &= sin(x_1 - x_2) \\ v(\vec{x}) &= cos(x_1 - x_2) \\ z(\vec{x}) &= tan(x_1 - x_2) \end{aligned}$$

$$\begin{pmatrix} \frac{\partial u(\vec{x})}{\partial \vec{x}} \\ \frac{\partial v(\vec{x})}{\partial \vec{x}} \\ \frac{\partial z(\vec{x})}{\partial \vec{x}} \end{pmatrix} = \begin{pmatrix} \nabla_{\vec{x}}\, u \\ \nabla_{\vec{x}}\, v \\ \nabla_{\vec{x}}\, z \end{pmatrix} = \begin{pmatrix} \frac{\partial sin(x_1 - x_2)}{\partial x_1} & \frac{\partial sin(x_1 - x_2)}{\partial x_2} \\ \frac{\partial cos(x_1 - x_2)}{\partial x_1} & \frac{\partial cos(x_1 - x_2)}{\partial x_2} \\ \frac{\partial tan(x_1 - x_2)}{\partial x_1} & \frac{\partial tan(x_1 - x_2)}{\partial x_2} \end{pmatrix}$$

$$= \begin{pmatrix} cos(x_1 - x_2) & -cos(x_1 - x_2) \\ -sin(x_1 - x_2) & cos(x_1 - x_2) \\ sec^2(x_1 - x_2) & -sec^2(x_1 - x_2) \end{pmatrix}$$

- Jacobians of Vector Functions

Vector Functions and Vector Inputs

$$\overrightarrow{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

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

$$\begin{pmatrix} \nabla_{\overrightarrow{x}} f_1(\overrightarrow{x}) \\ \nabla_{\overrightarrow{x}} f_2(\overrightarrow{x}) \\ \vdots \\ \nabla_{\overrightarrow{x}} f_n(\overrightarrow{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(\overrightarrow{x})}{\partial x_1} & \frac{\partial f_1(\overrightarrow{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\overrightarrow{x})}{\partial x_m} \\ \frac{\partial f_2(\overrightarrow{x})}{\partial x_1} & \frac{\partial f_2(\overrightarrow{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\overrightarrow{x})}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\overrightarrow{x})}{\partial x_1} & \frac{\partial f_n(\overrightarrow{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\overrightarrow{x})}{\partial x_m} \end{pmatrix}$$

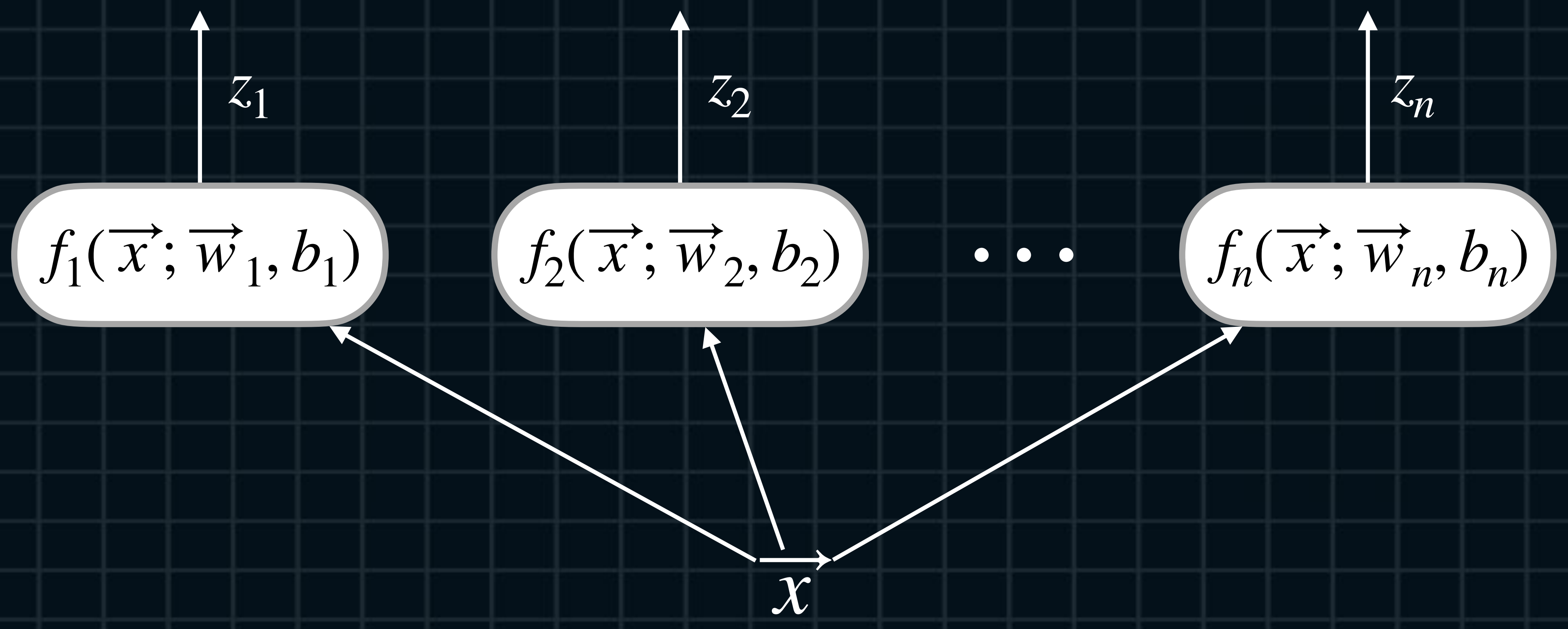
Lecture.5 Vector Functions and Jacobians - Jacobians of Vector Functions

Review

		$x \in \mathbb{R}$	$\vec{x} \in \mathbb{R}^m$	
$f \in \mathbb{R}$	$\frac{\partial f}{\partial x}$	$\boxed{\frac{\partial f}{\partial x}}$ $(1, 1)$	$\boxed{\frac{\partial f}{\partial \vec{x}}}$ $(1, m)$	$\begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_m} \end{pmatrix}$
$\vec{f} \in \mathbb{R}^n$	$\begin{pmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \\ \vdots \\ \frac{\partial f_n}{\partial x} \end{pmatrix}$	$\boxed{\frac{\partial \vec{f}}{\partial x}}$ $(n, 1)$	$\boxed{\frac{\partial \vec{f}}{\partial \vec{x}}}$ (n, m)	$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_m} \end{pmatrix}$

- Affine Function as a Vector Function

One Sample and Neurons

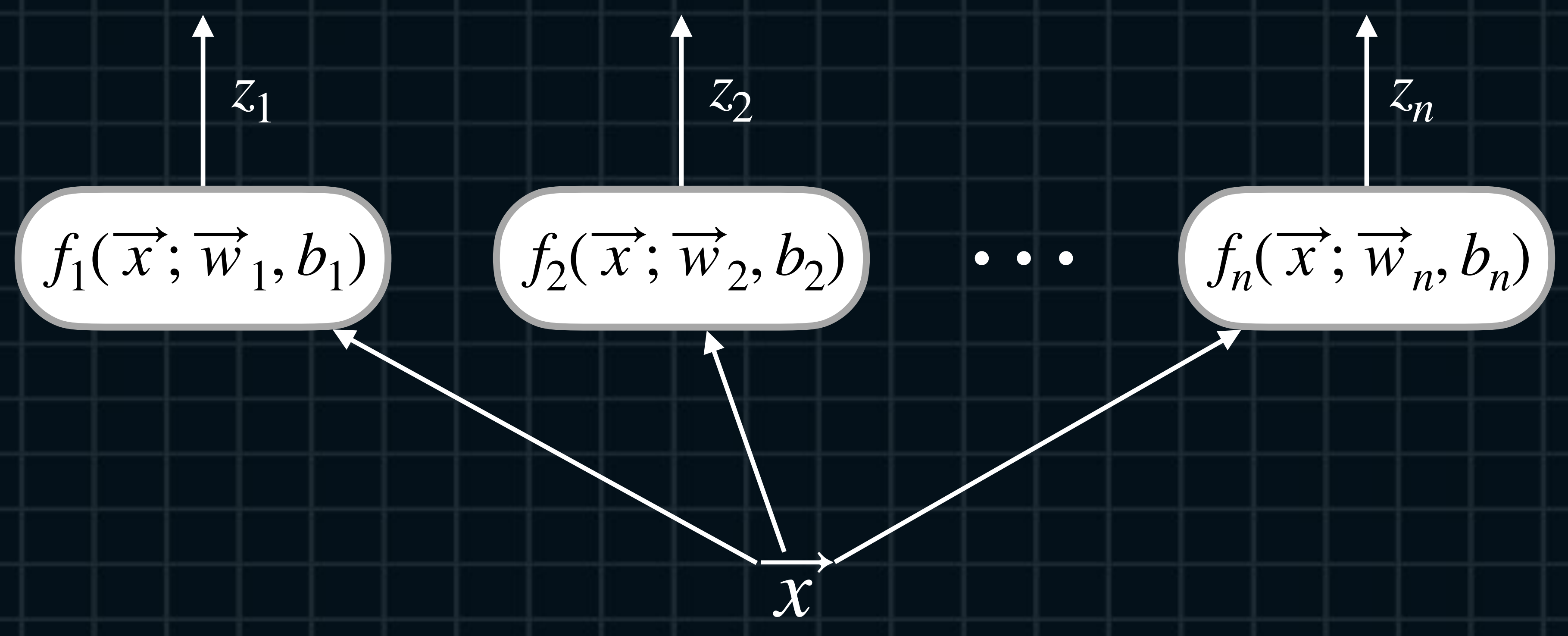


$$
\begin{aligned}
z_1 &= \vec{x}^T \cdot \vec{w}_1 + b_1 \\
z_2 &= \vec{x}^T \cdot \vec{w}_2 + b_2 \\
&\vdots \\
z_n &= \vec{x}^T \cdot \vec{w}_n + b_n
\end{aligned}
$$

$$
\begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & (\vec{x})^T & \longrightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} + \begin{pmatrix} b_1 & b_2 & \dots & b_n \end{pmatrix}
$$

- Affine Function as a Vector Function

One Sample and Neurons



$$z_1 = \vec{x}^T \cdot \vec{w}_1 + b_1 \qquad \frac{\partial z_1}{\partial \vec{x}} = \left(\vec{w}_1\right)^T$$

$$z_2 = \vec{x}^T \cdot \vec{w}_2 + b_2 \qquad \frac{\partial z_2}{\partial \vec{x}} = \left(\vec{w}_2\right)^T$$

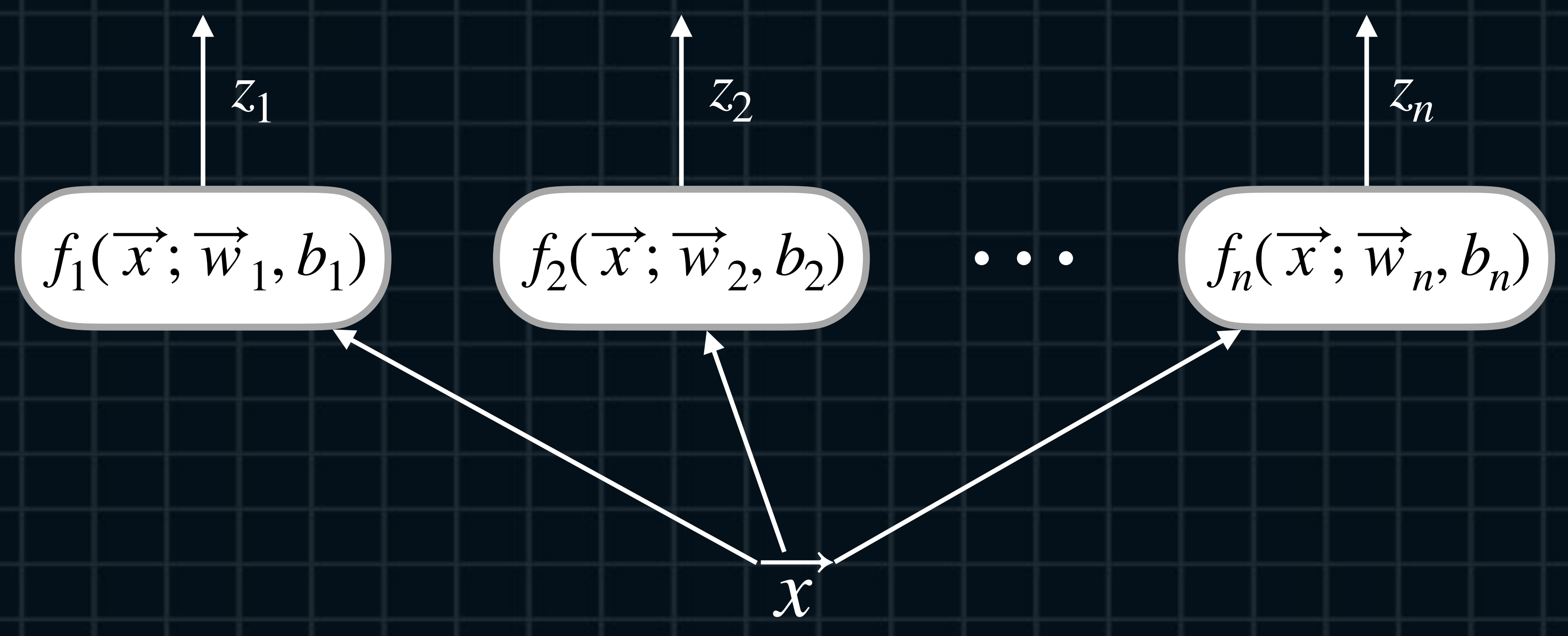
$$\vdots \qquad \vdots$$

$$z_n = \vec{x}^T \cdot \vec{w}_n + b_n \qquad \frac{\partial z_n}{\partial \vec{x}} = \left(\vec{w}_n\right)^T$$

$$\frac{\partial \vec{z}}{\partial \vec{x}} = \begin{pmatrix} \nabla_{\vec{x}}\, z_1 \\ \nabla_{\vec{x}}\, z_2 \\ \vdots \\ \nabla_{\vec{x}}\, z_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{w}_1\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{w}_2\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{w}_n\right)^T & \longrightarrow \end{pmatrix} = W^T$$

- Affine Function as a Vector Function

One Sample and Neurons



$$z_1 = \vec{x}^T \cdot \vec{w}_1 + b_1 \qquad \frac{\partial z_1}{\partial \vec{w}_1} = \left(\vec{x}\right)^T$$

$$z_2 = \vec{x}^T \cdot \vec{w}_2 + b_2 \qquad \frac{\partial z_1}{\partial \vec{w}_2} = \left(\vec{x}\right)^T$$

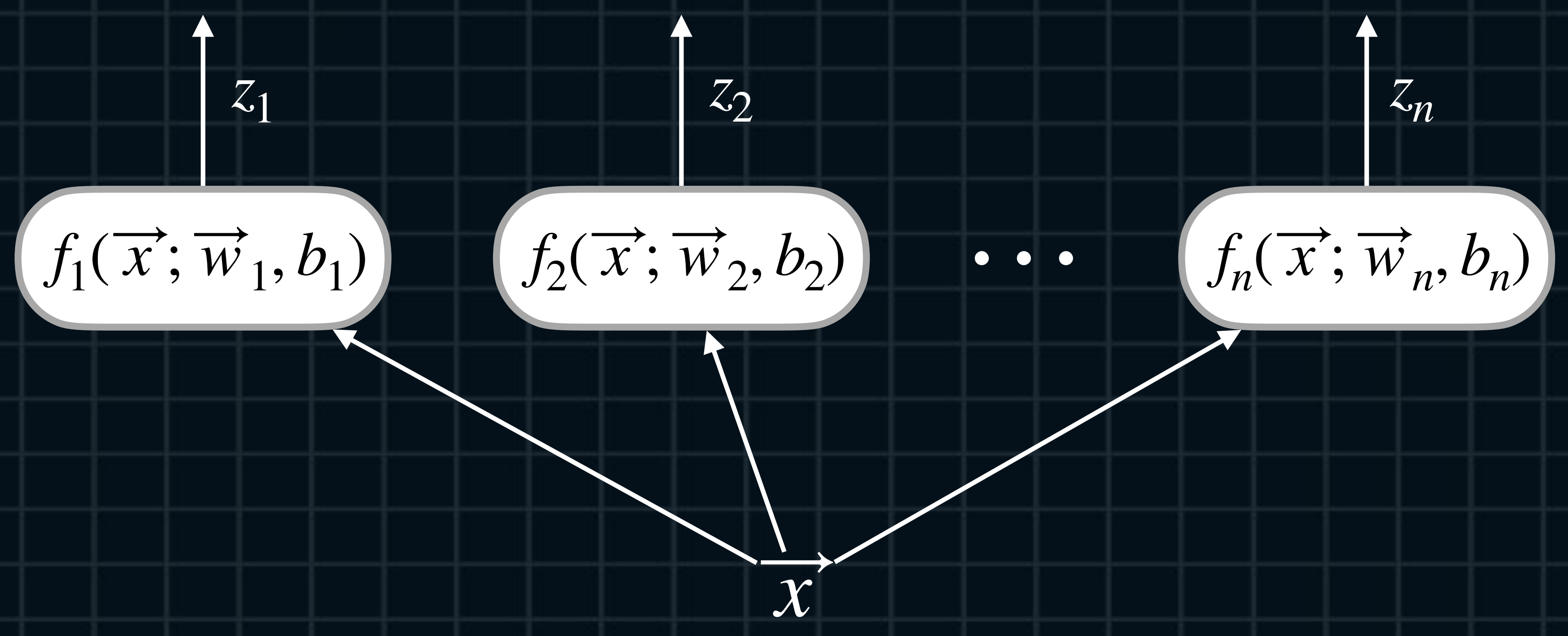
$$\vdots \qquad \vdots$$

$$z_n = \vec{x}^T \cdot \vec{w}_n + b_n \qquad \frac{\partial z_1}{\partial \vec{w}_n} = \left(\vec{x}\right)^T$$

$$\frac{\partial \vec{z}}{\partial \vec{w}_1} = \begin{pmatrix} \nabla_{\vec{w}_1} z_1 \\ \nabla_{\vec{w}_1} z_2 \\ \vdots \\ \nabla_{\vec{w}_1} z_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{x}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \end{pmatrix} \quad \frac{\partial \vec{z}}{\partial \vec{w}_2} = \begin{pmatrix} \nabla_{\vec{w}_2} z_1 \\ \nabla_{\vec{w}_2} z_2 \\ \vdots \\ \nabla_{\vec{w}_2} z_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \end{pmatrix} \cdots \frac{\partial \vec{z}}{\partial \vec{w}_n} = \begin{pmatrix} \nabla_{\vec{w}_n} z_1 \\ \nabla_{\vec{w}_n} z_2 \\ \vdots \\ \nabla_{\vec{w}_n} z_n \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{0}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}\right)^T & \longrightarrow \end{pmatrix}$$

- Affine Function as a Vector Function

One Sample and Neurons



$$z_1 = \vec{x}^T \cdot \vec{w}_1 + b_1$$

$$z_2 = \vec{x}^T \cdot \vec{w}_2 + b_2$$

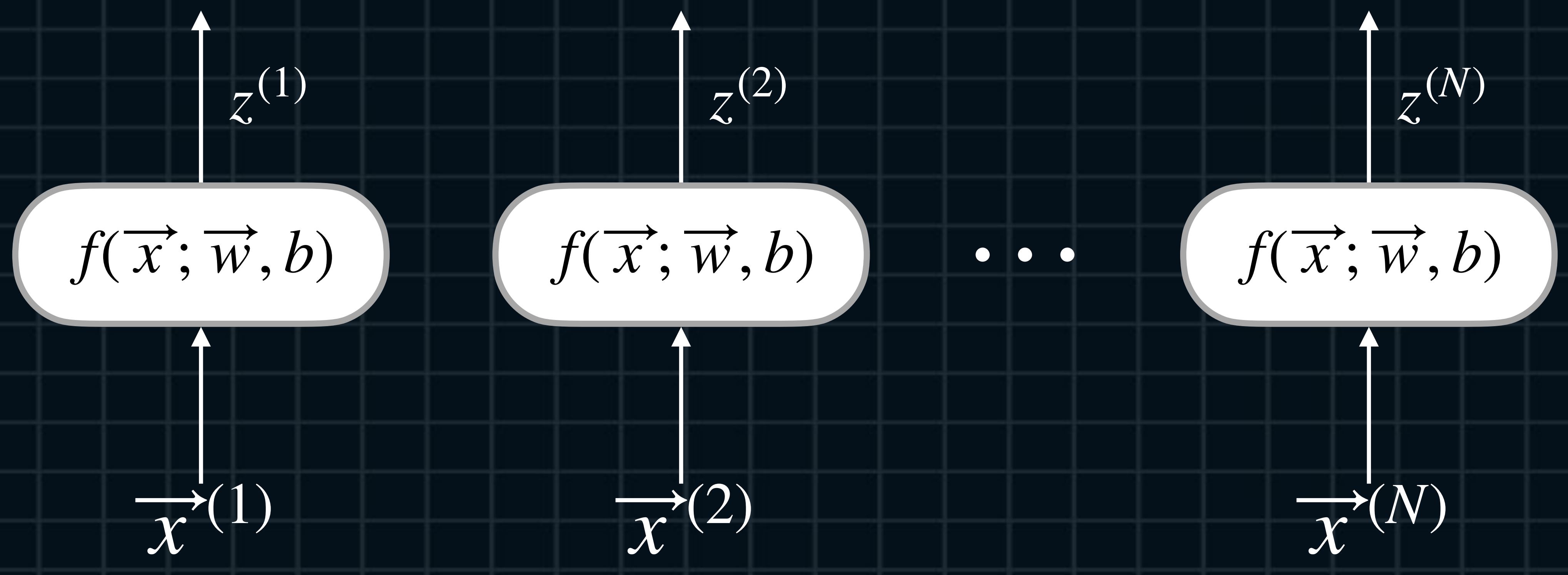
$$\vdots$$

$$z_n = \vec{x}^T \cdot \vec{w}_n + b_n$$

$$\frac{\partial \vec{z}}{\partial \vec{b}} = \begin{pmatrix} \nabla_{\vec{b}}\, z_1 \\ \nabla_{\vec{b}}\, z_2 \\ \vdots \\ \nabla_{\vec{b}}\, z_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Lecture.5 Vector Functions and Jacobians - Affine Function as a Vector Function

Mini-batch and One Neuron



$$z^{(1)} = \left(\vec{x}^{(1)}\right)^T \cdot \vec{w} + b$$

$$z^{(2)} = \left(\vec{x}^{(2)}\right)^T \cdot \vec{w} + b$$

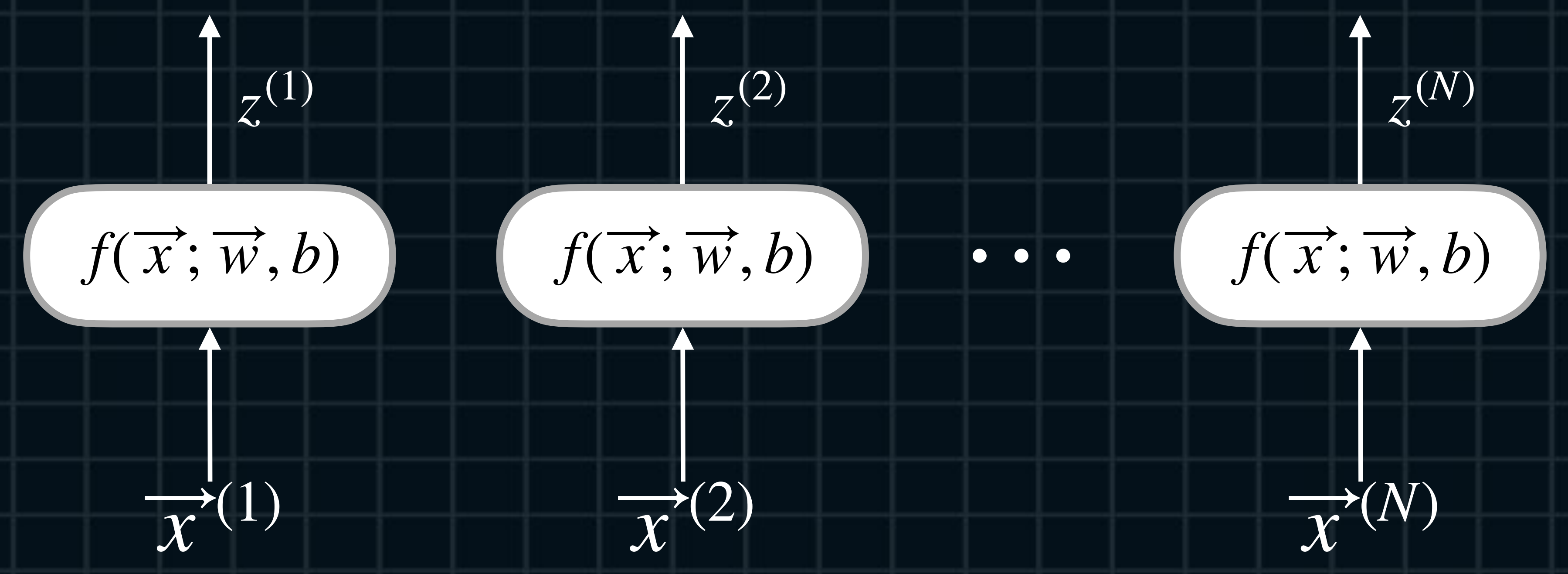
$$\vdots$$

$$z^{(N)} = \left(\vec{x}^{(N)}\right)^T \cdot \vec{w} + b$$

$$\begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix} = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} \begin{pmatrix} \uparrow \\ \vec{w} \\ \downarrow \end{pmatrix} + b$$

- Affine Function as a Vector Function

Mini-batch and One Neuron



$$z^{(1)} = \left(\vec{x}^{(1)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(1)}}{\partial \vec{x}^{(1)}} = \vec{w}^T$$

$$z^{(2)} = \left(\vec{x}^{(2)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(2)}}{\partial \vec{x}^{(2)}} = \vec{w}^T$$

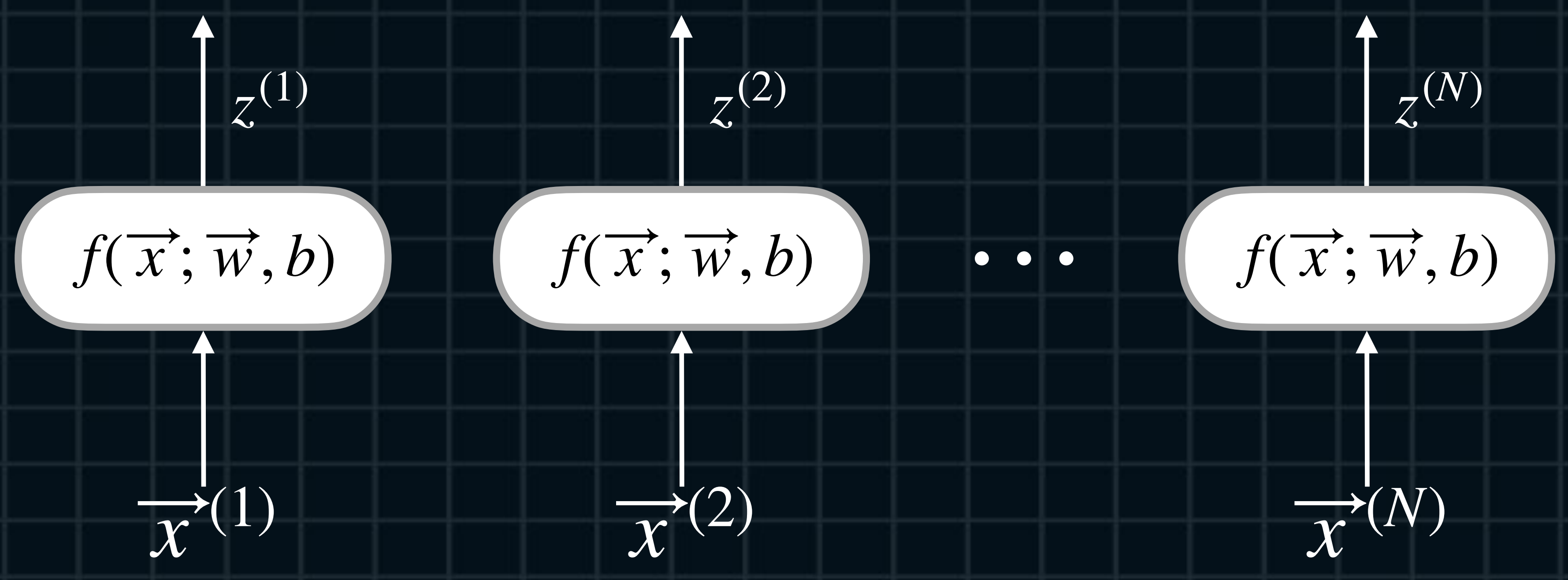
$$\vdots \qquad \vdots$$

$$z^{(N)} = \left(\vec{x}^{(N)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(N)}}{\partial \vec{x}^{(N)}} = \vec{w}^T$$

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

- Affine Function as a Vector Function

Mini-batch and One Neuron



$$z^{(1)} = \left(\vec{x}^{(1)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(1)}}{\partial \vec{w}} = \left(\vec{x}^{(1)}\right)^T$$

$$z^{(2)} = \left(\vec{x}^{(2)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(2)}}{\partial \vec{w}} = \left(\vec{x}^{(2)}\right)^T$$

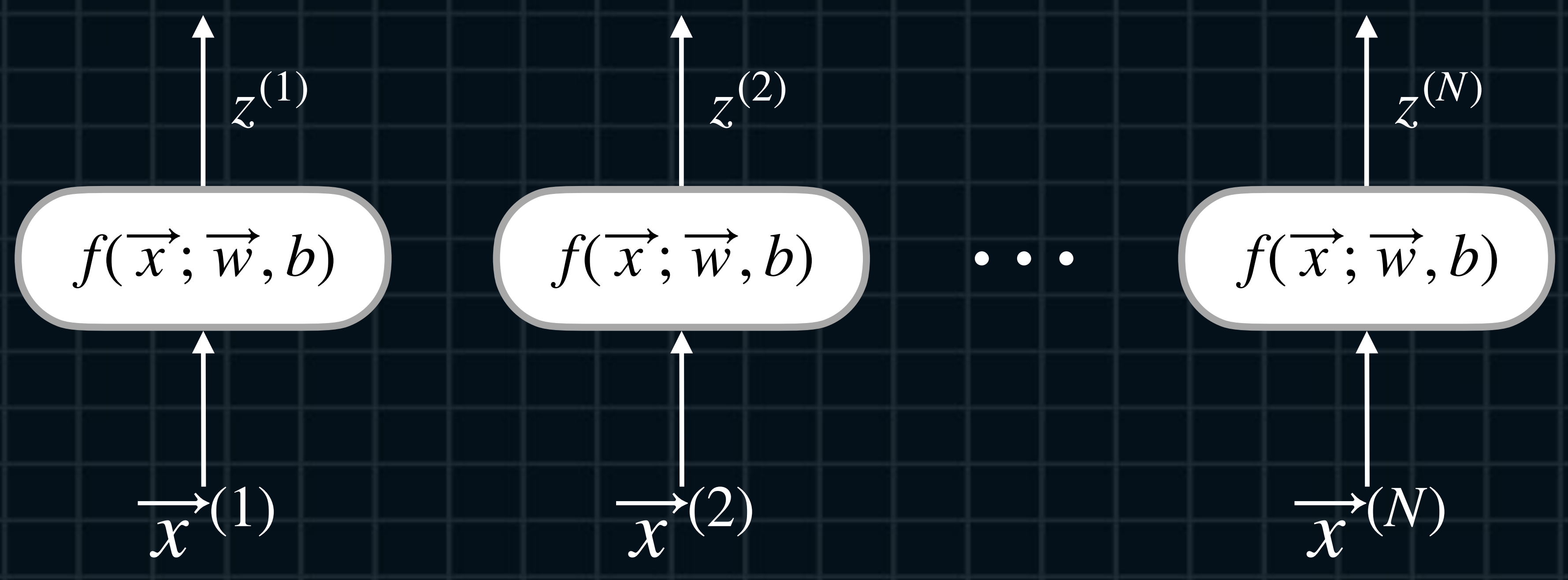
$$\vdots \qquad\qquad \vdots$$

$$z^{(N)} = \left(\vec{x}^{(N)}\right)^T \cdot \vec{w} + b \qquad \frac{\partial z^{(N)}}{\partial \vec{w}} = \left(\vec{x}^{(N)}\right)^T$$

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

- Affine Function as a Vector Function

Mini-batch and One Neuron



$$z^{(1)} = \left(\vec{x}^{(1)}\right)^T \cdot \vec{w} + b$$

$$z^{(2)} = \left(\vec{x}^{(2)}\right)^T \cdot \vec{w} + b$$

$$\vdots$$

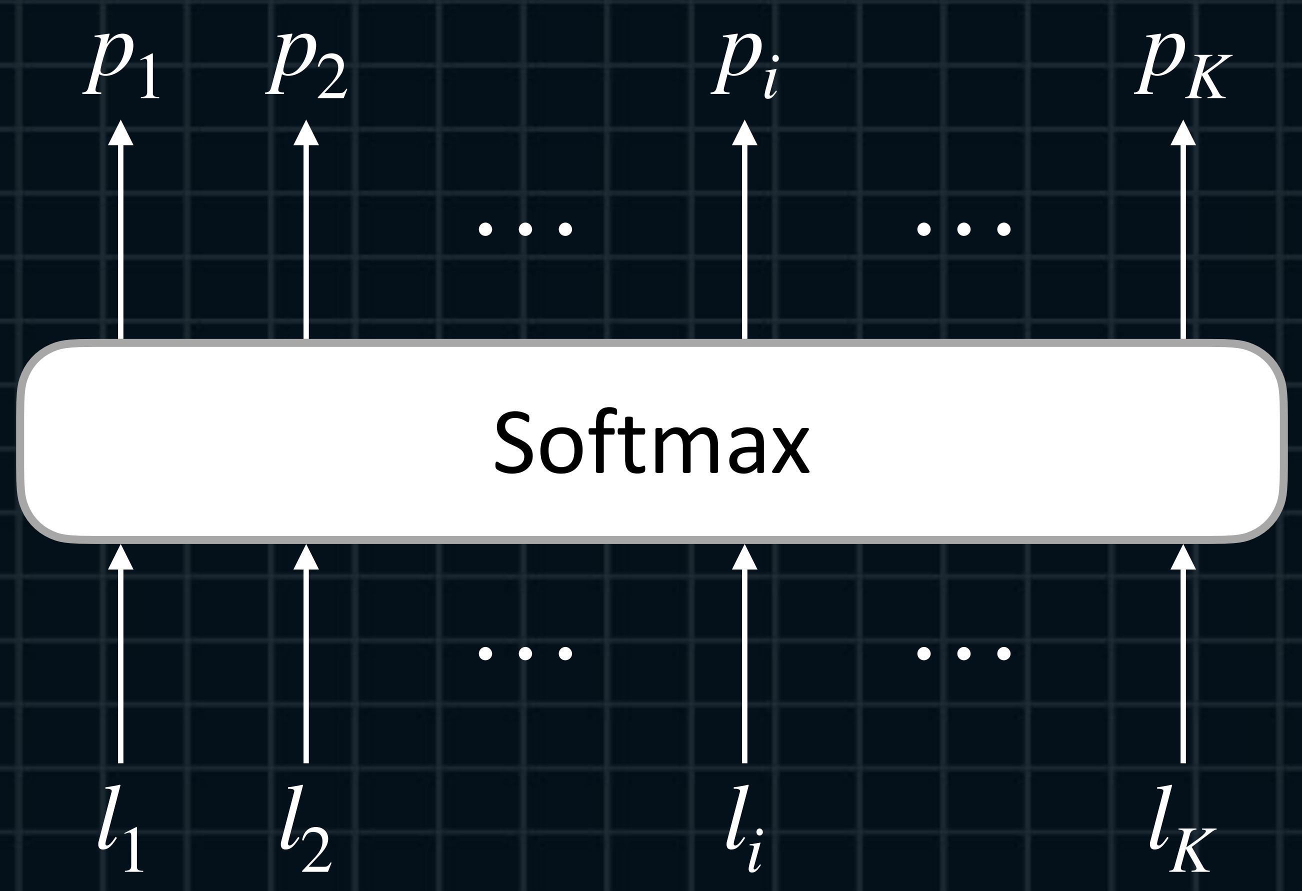
$$z^{(N)} = \left(\vec{x}^{(N)}\right)^T \cdot \vec{w} + b$$

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

Softmax as a Vector Function

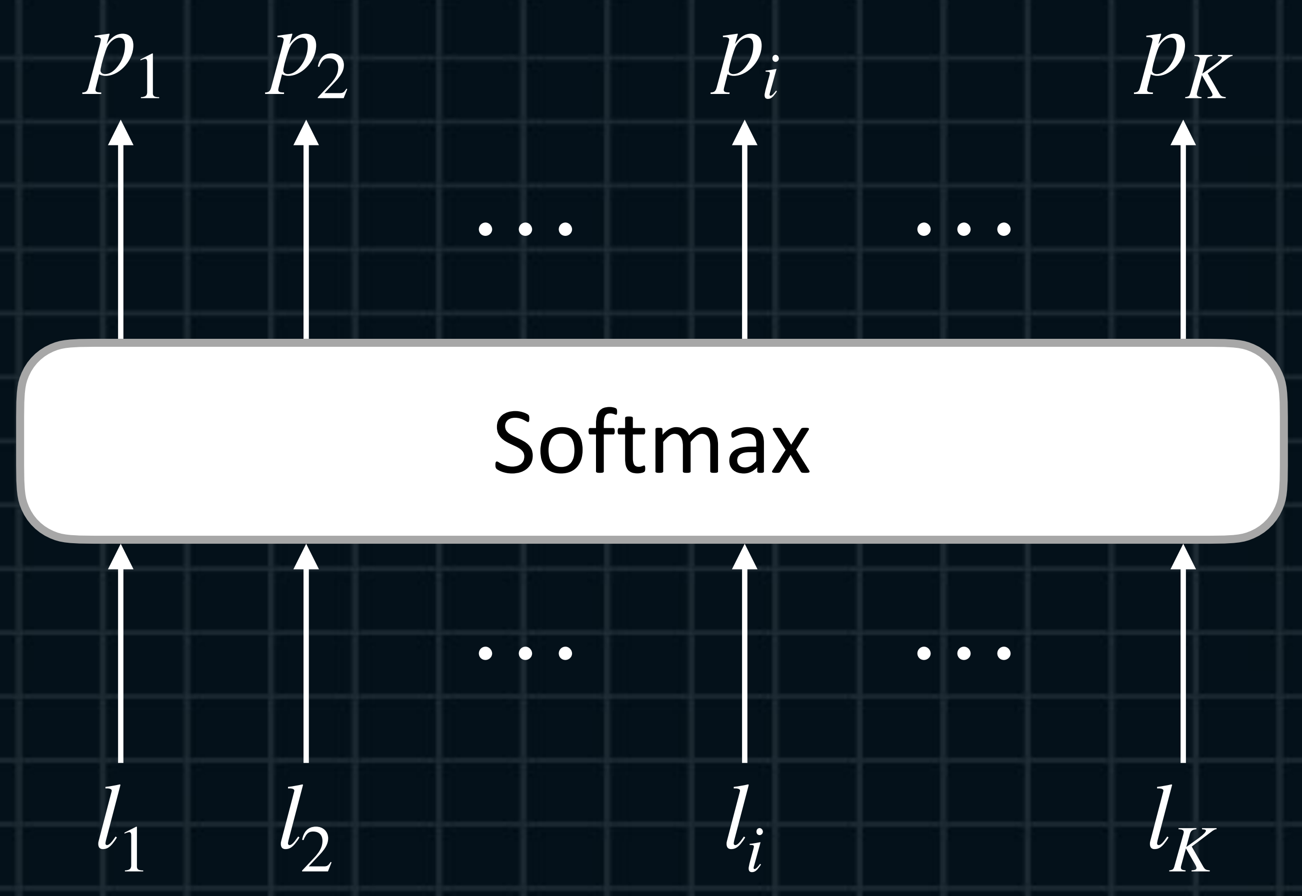
$$p_i = \frac{e^{l_i}}{\sum_{k=1}^{K} e^{l_k}}$$

$$\vec{l} \in \mathbb{R}^K, \ \ \vec{p} \in \mathbb{R}^K$$



$$p_1 = S_1(\vec{l}) = e^{l_1}/S$$
$$p_2 = S_2(\vec{l}) = e^{l_2}/S$$
$$\vdots$$
$$p_K = S_K(\vec{l}) = e^{l_K}/S$$

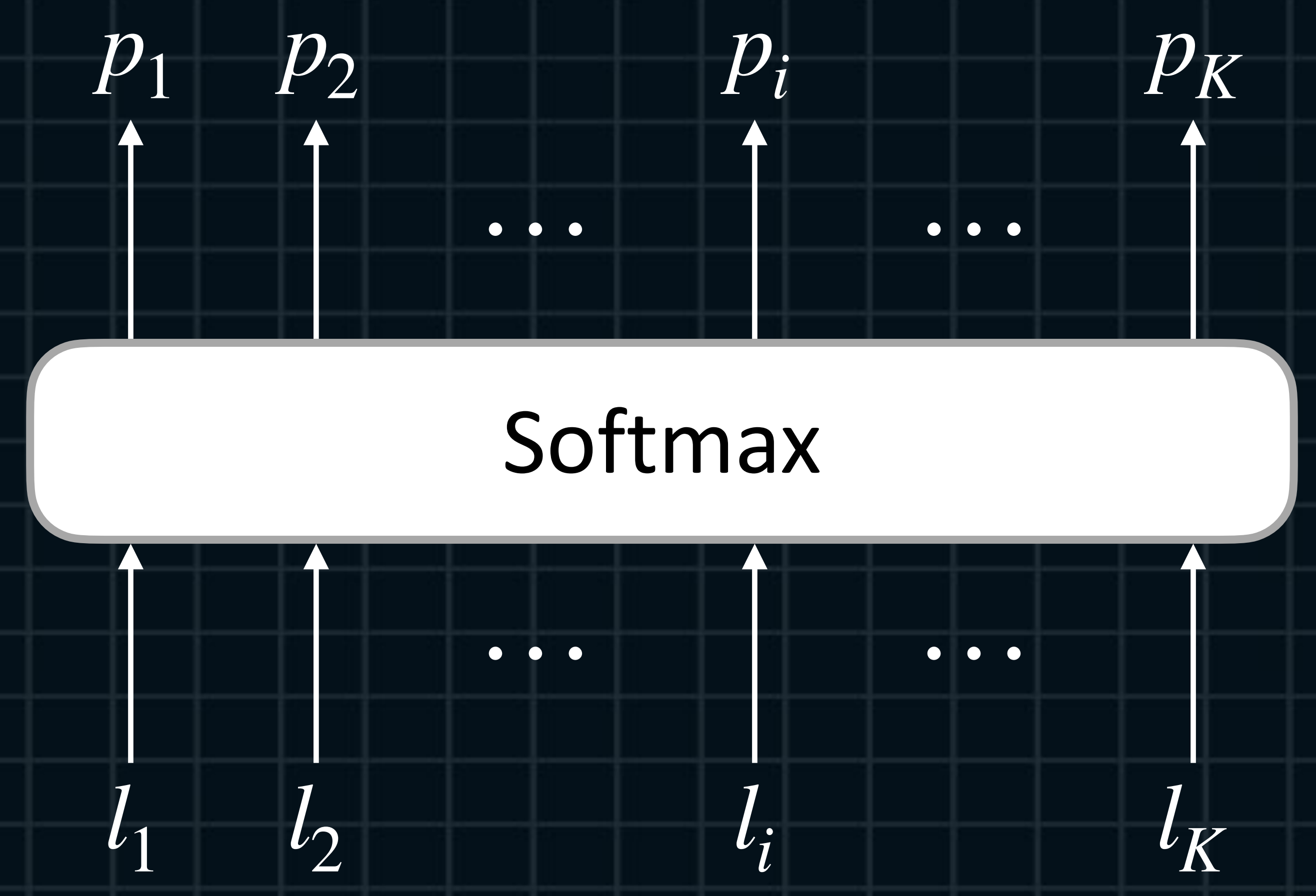
Gradients of Softmax



$$\frac{\partial p_i}{\partial l_1} = -p_i p_1, \frac{\partial p_i}{\partial l_2} = -p_i p_2, \ \dots \ , \frac{\partial p_i}{\partial l_i} = p_i(1-p_i), \ \dots \ , \frac{\partial p_i}{\partial l_K} = -p_i p_K$$

$$\frac{\partial p_i}{\partial \vec{l}} = \nabla_{\vec{l}}\, p_i = \begin{pmatrix} -p_i p_1 & -p_i p_2 & \dots & p_i(1-p_i) & \dots & -p_i p_K \end{pmatrix}$$

Gradients of Softmax



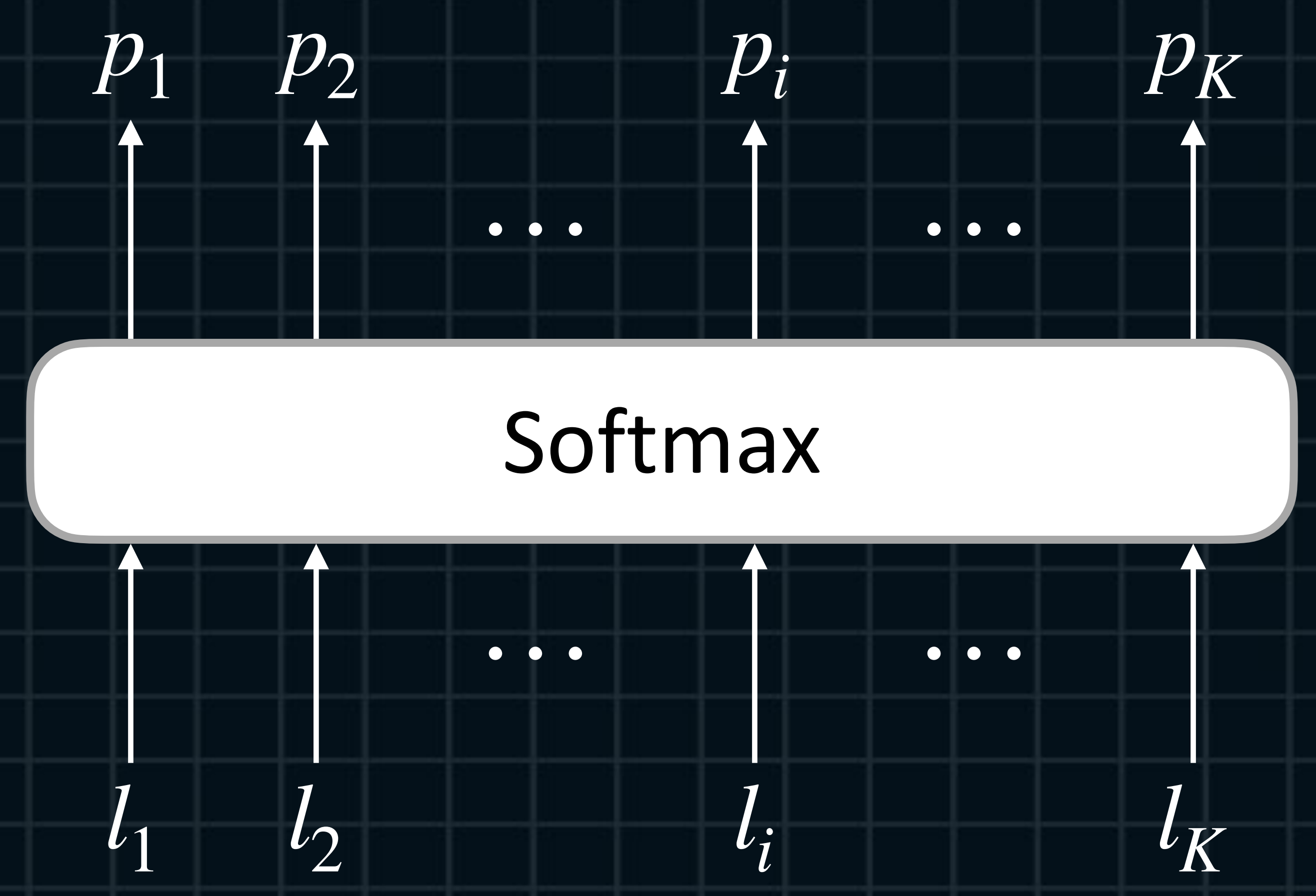
$$\frac{\partial p_1}{\partial \vec{l}} = \nabla_{\vec{l}}\, p_1 = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_K \end{pmatrix}$$

$$\frac{\partial p_2}{\partial \vec{l}} = \nabla_{\vec{l}}\, p_2 = \begin{pmatrix} -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_K \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial p_K}{\partial \vec{l}} = \nabla_{\vec{l}}\, p_K = \begin{pmatrix} -p_Kp_1 & -p_Kp_2 & \dots & p_K(1-p_K) \end{pmatrix}$$

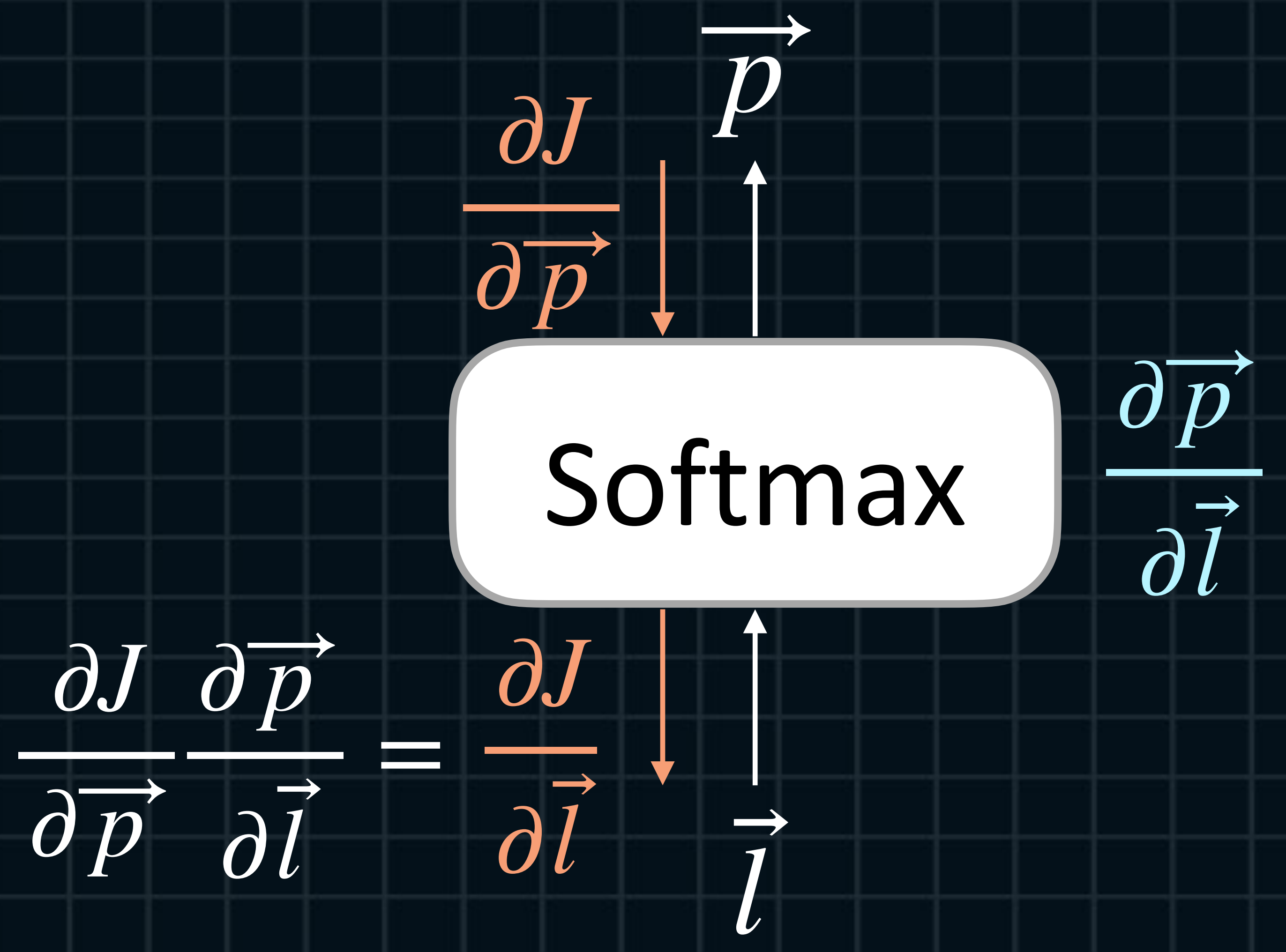
Jacobians of Softmax



$$\frac{\partial \vec{p}}{\partial \vec{l}} = \begin{pmatrix} \frac{\partial p_1}{\partial \vec{l}} \\ \frac{\partial p_2}{\partial \vec{l}} \\ \vdots \\ \frac{\partial p_K}{\partial \vec{l}} \end{pmatrix} = \begin{pmatrix} \nabla_{\vec{l}}\, p_1 \\ \nabla_{\vec{l}}\, p_2 \\ \vdots \\ \nabla_{\vec{l}}\, p_K \end{pmatrix} = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_K \\ -p_2 p_1 & p_2(1-p_2) & \cdots & -p_2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_K p_1 & -p_K p_2 & \cdots & p_K(1-p_K) \end{pmatrix} \in \mathbb{R}^{K\times K}$$

Lecture.5 Vector Functions and Jacobians - Jacobians of Softmax

Backpropagation of Softmax



$$\frac{\partial J}{\partial \vec{p}}\frac{\partial \vec{p}}{\partial \vec{l}}: \quad \mathbb{R}^{1\times K} \times \mathbb{R}^{K\times K} \longrightarrow \mathbb{R}^{1\times K}$$

Backpropagation and Jacobian Matrices

Lecture.5
Vector Functions and Jacobians