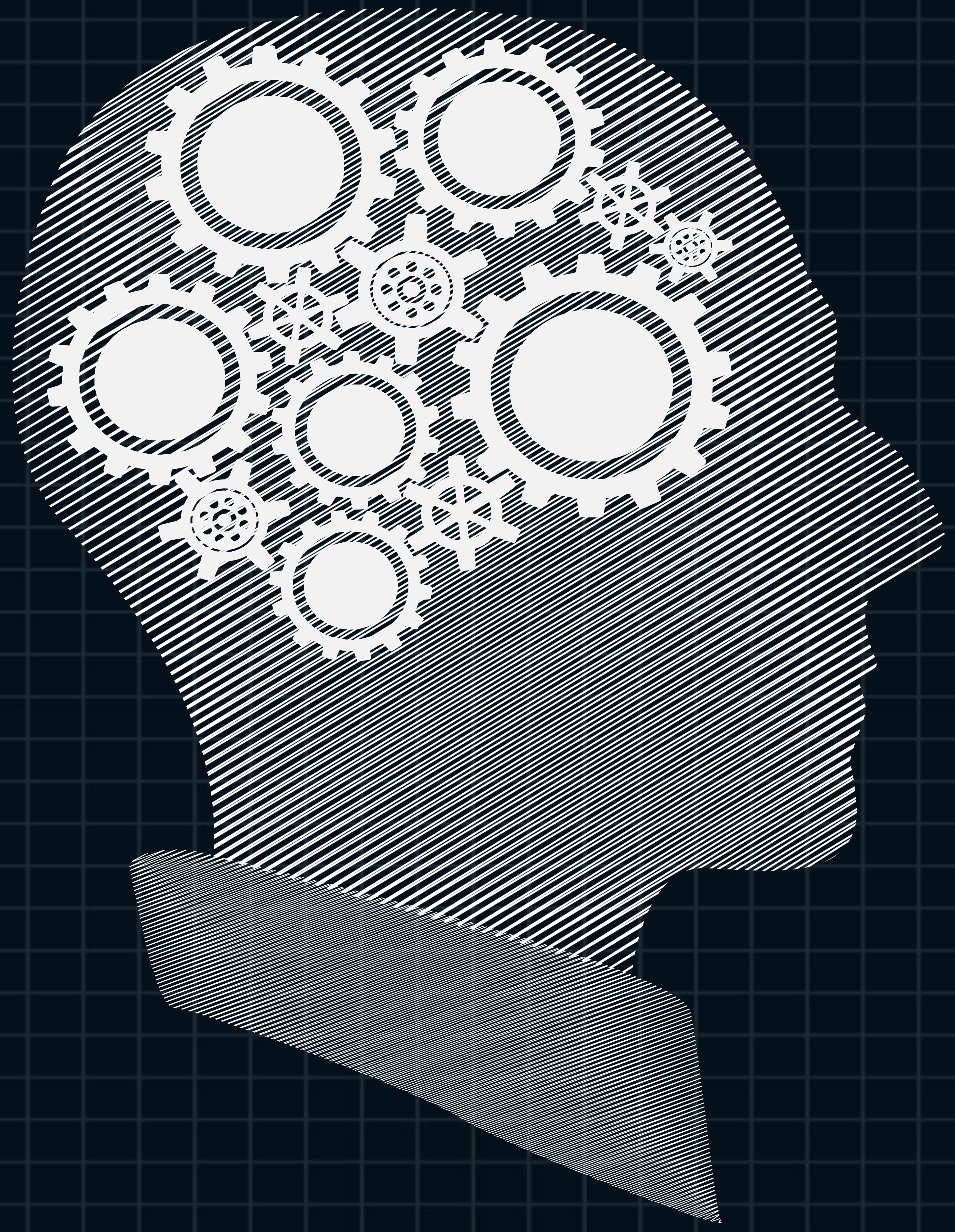
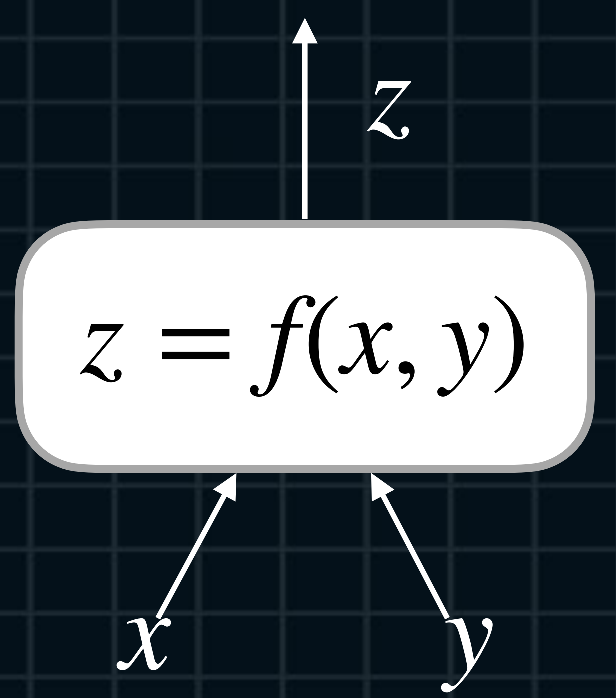
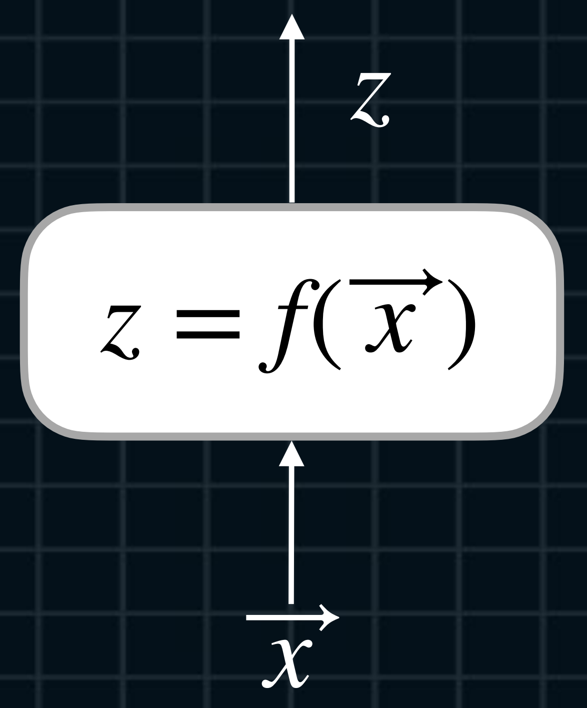


Backpropagation and Jacobian Matrices

Lecture.3 Multivariate Functions and Jacobians

Multiple Input and Single Output





$$z = f(x, y) = x + y$$

$$z = f(x, y) = x^2 + y^2$$

$$z = f(x, y) = xy$$

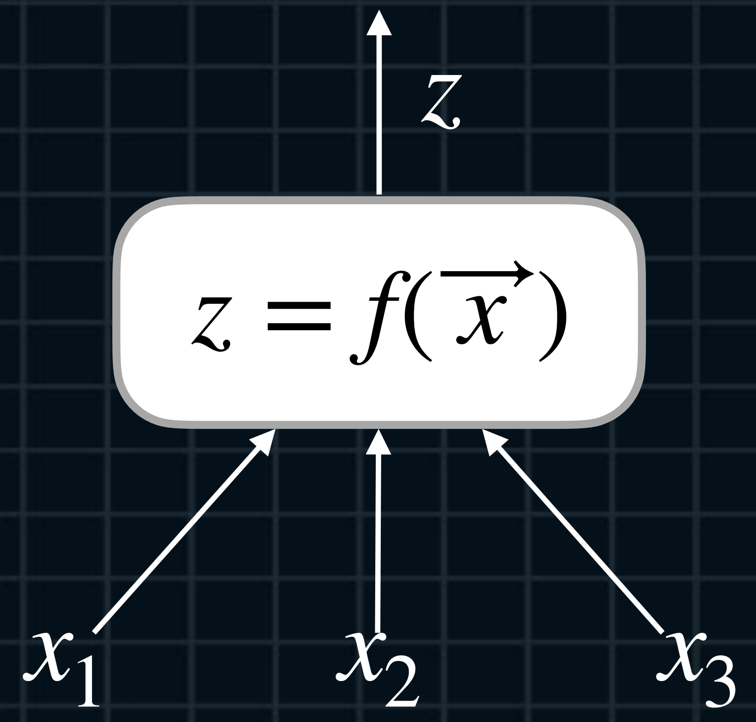
$$z = f(\overrightarrow{x}) = \frac{1}{3}\sum_{i=1}^{3} x_i$$

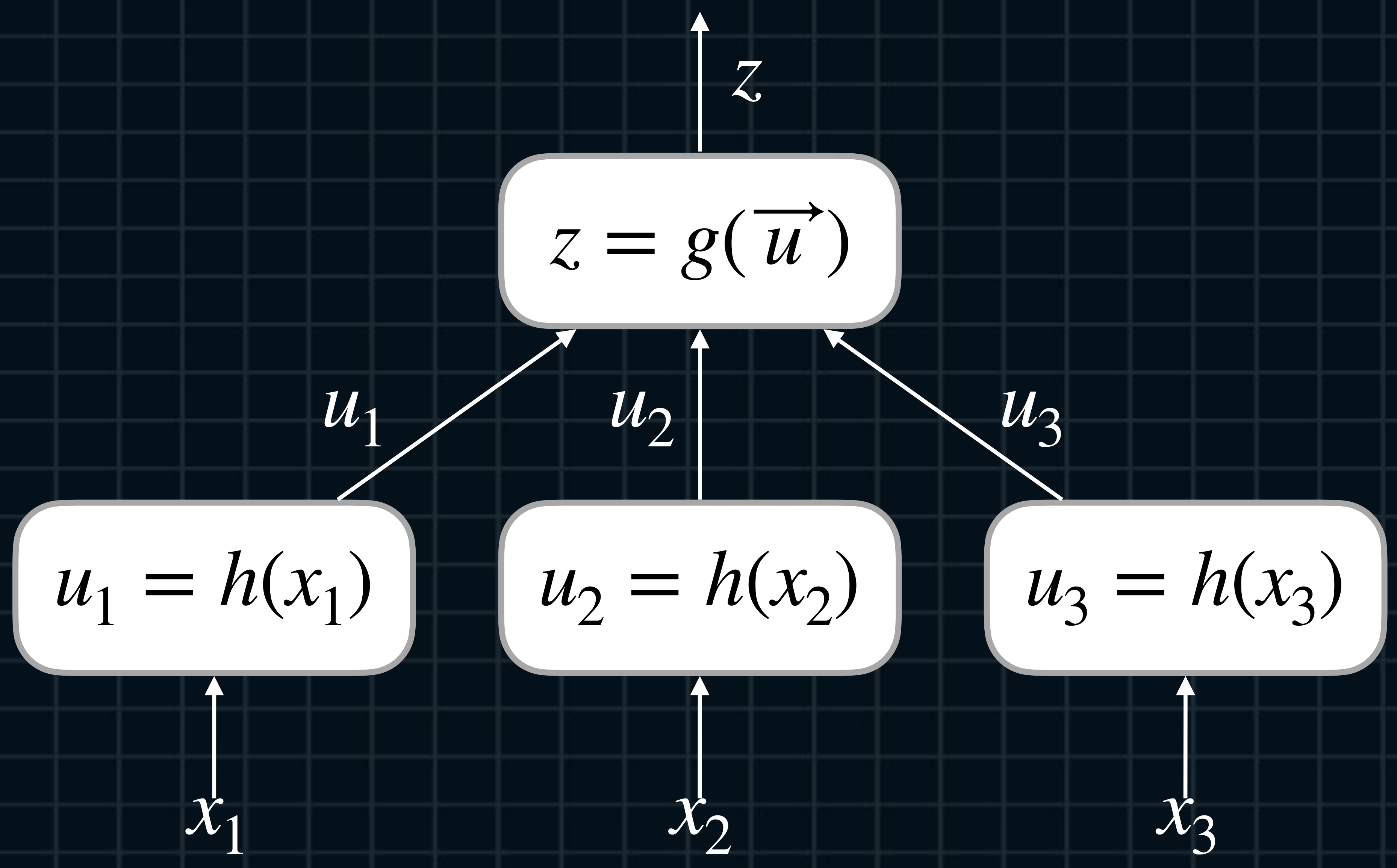
$$z = f(\overrightarrow{x}) = \frac{1}{3}\sum_{i=1}^{3} (x_i)^2$$

Multiple Input and Single Output

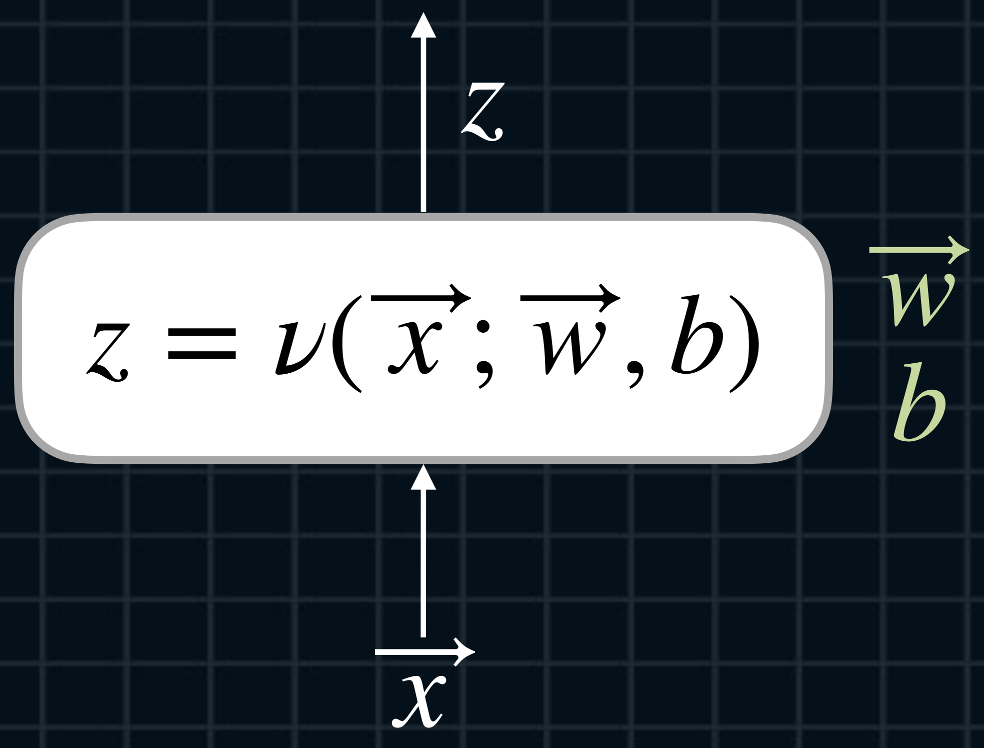
$$z = f(\overrightarrow{x}) = \frac{1}{3}\sum_{i=1}^{3} x_i$$

$$z = f(\overrightarrow{x}) = \frac{1}{3}\sum_{i=1}^{3} (x_i)^2$$





Multiple Input and Single Output



$$
\begin{aligned}
z = \nu(\vec{x}; \vec{w}, b) &= \vec{x}^T \cdot \vec{w} + b \\
&= x_1 w_1 + x_2 w_2 + \ldots + x_n w_n + b
\end{aligned}
$$

Multiple Input and Single Output

$$\vec{x} \longrightarrow \boxed{z = f(\vec{x})} \longrightarrow z$$

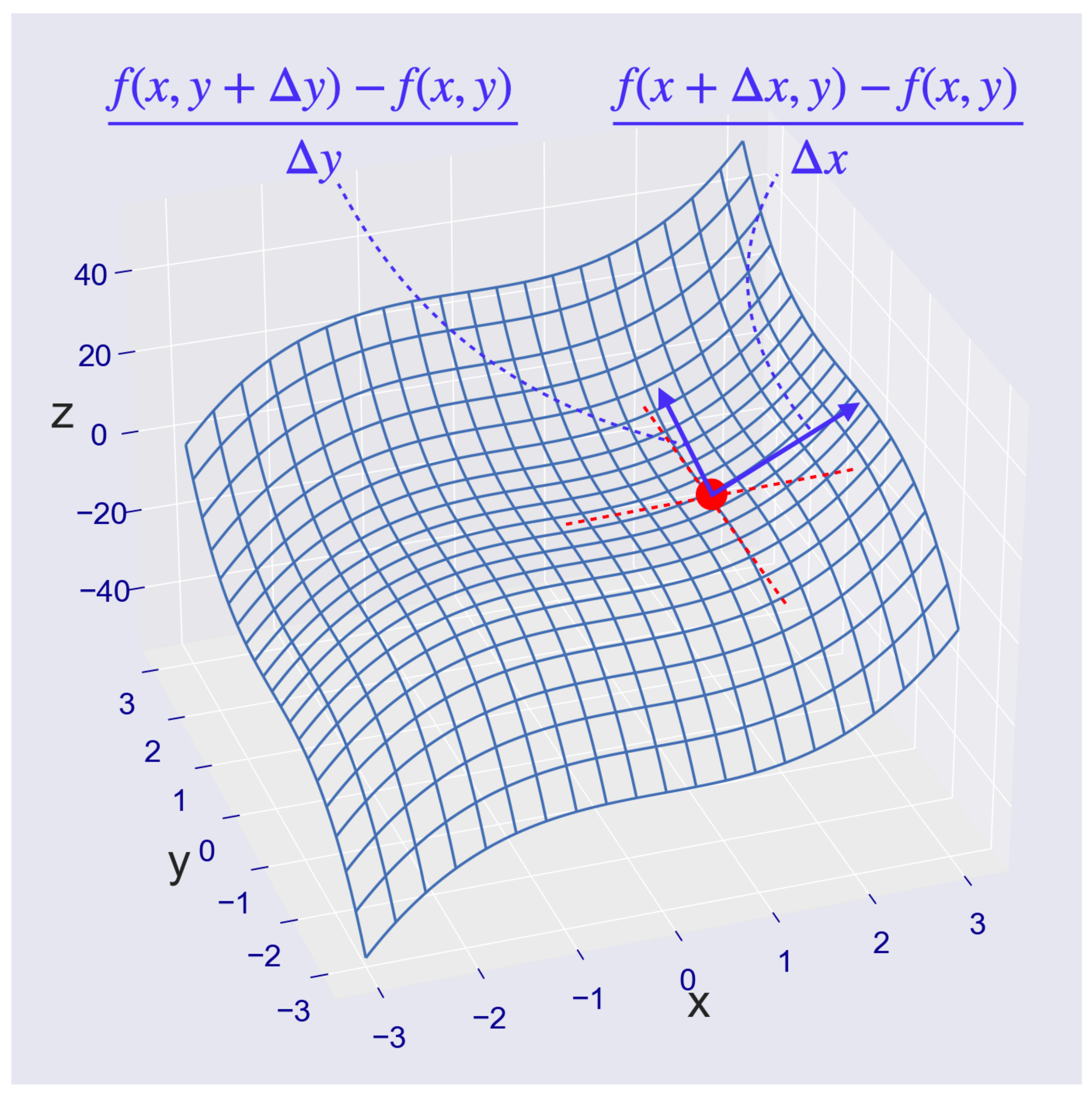
$$\mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\vec{x} \longrightarrow \boxed{z = \nu(\vec{x}; \vec{w}, b)} \longrightarrow z \qquad \vec{w},\ b$$

$$\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}$$

Partial Derivatives

$$z = f(x, y)$$



$$\frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

$$\frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

$$\frac{\partial f(x, y)}{\partial x} = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

$$\frac{\partial f(x, y)}{\partial y} = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

Parameter Update

$$x := x + \alpha \frac{\partial z}{\partial x}$$

increasing direction

$$x := x - \alpha \frac{\partial z}{\partial x}$$

decreasing direction

Parameter Update

$$x := x + \alpha\frac{\partial z}{\partial x},\ y := y + \alpha\frac{\partial z}{\partial y}$$ most increasing direction

$$x := x + \alpha\frac{\partial z}{\partial x},\ y := y - \alpha\frac{\partial z}{\partial y}$$

$$x := x - \alpha\frac{\partial z}{\partial x},\ y := y + \alpha\frac{\partial z}{\partial y}$$

$$x := x - \alpha\frac{\partial z}{\partial x},\ y := y - \alpha\frac{\partial z}{\partial y}$$ most decreasing direction

General Multivariate Functions and Partial Derivatives

$$z = f(\overrightarrow{x})$$

$$(\overrightarrow{x})^T = (x_1 \quad x_2 \quad \dots \quad x_n)$$

$$\frac{\partial f(\overrightarrow{x})}{\partial x_1}, \frac{\partial f(\overrightarrow{x})}{\partial x_2}, \dots, \frac{\partial f(\overrightarrow{x})}{\partial x_n}$$

$$x_i := x_i + \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_i}$$

$$x_i := x_i - \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_i}$$

Parameter Update

$$x_1 := x_1 + \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_1}, x_2 := x_2 + \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_2}, \ \dots \ , x_n := x_n + \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_n}$$

$$x_1 := x_1 - \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_1}, x_2 := x_2 - \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_2}, \ \dots \ , x_n := x_n - \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_n}$$

Gradients

$$z = f(\vec{x}),\ \vec{x} \in \mathbb{R}^n$$

$$\frac{\partial f(\vec{x})}{\partial x_1}, \frac{\partial f(\vec{x})}{\partial x_2}, \ldots, \frac{\partial f(\vec{x})}{\partial x_n}$$

$$\nabla_{\vec{x}} f(\vec{x}) = \begin{pmatrix} \frac{\partial f(\vec{x})}{\partial x_1} & \frac{\partial f(\vec{x})}{\partial x_2} & \ldots & \frac{\partial f(\vec{x})}{\partial x_n} \end{pmatrix}$$

$$\nabla_{\vec{x}} f(\vec{x}) = \frac{\partial f(\vec{x})}{\partial \vec{x}}$$

Gradients

$$\nabla_{\vec{x}} f(\vec{x}) = \begin{pmatrix} \frac{\partial f(\vec{x})}{\partial x_1} & \frac{\partial f(\vec{x})}{\partial x_2} & \cdots & \frac{\partial f(\vec{x})}{\partial x_n} \end{pmatrix}$$

$$z = f(x, y) = x + y$$

$$z = f(x, y) = x^2 + y^2$$

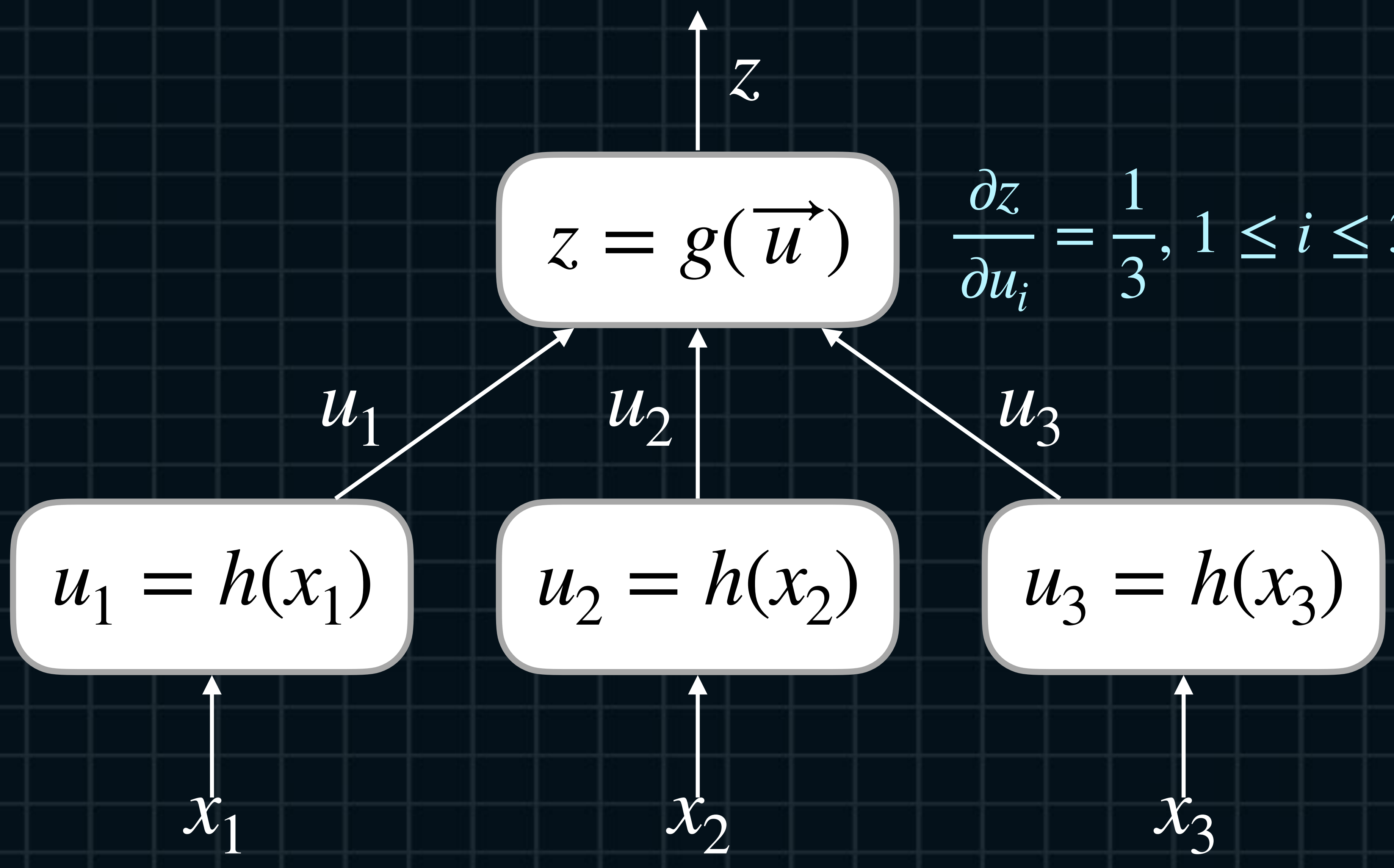
$$z = f(x, y) = xy$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad z = f(\vec{x}) = \frac{1}{3} \sum_{i=1}^{3} x_i$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad z = f(\vec{x}) = \frac{1}{3} \sum_{i=1}^{3} (x_i)^2$$

Lecture.3 Multivariate Functions and Jacobians - Partial Derivatives and Gradients

Gradients



$$\frac{\partial z}{\partial u_i} = \frac{1}{3}, 1 \le i \le 3$$

$$\frac{\partial u_i}{\partial z_i} = 2x_i, 1 \le i \le 3$$

$$\frac{\partial z}{\partial x_i} = \frac{\partial z}{\partial u_i}\frac{\partial u_i}{\partial x_i} = \frac{2x_i}{3}, 1 \le i \le 3$$

$$\frac{\partial z}{\partial \overrightarrow{x}} = \begin{pmatrix} \frac{2x_1}{3} & \frac{2x_2}{3} & \frac{2x_3}{3} \end{pmatrix}$$

Gradients and Jacobians

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

$$\frac{\partial f(\vec{x})}{\partial \vec{x}} = \nabla_{\vec{x}} f(\vec{x}) = \begin{pmatrix} \dfrac{\partial f(\vec{x})}{\partial x_1} & \dfrac{\partial f(\vec{x})}{\partial x_2} & \cdots & \dfrac{\partial f(\vec{x})}{\partial x_n} \end{pmatrix}$$

Gradients and Parameter Update

$$z = f(\overrightarrow{x})$$

$$\overrightarrow{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \frac{dz}{d\overrightarrow{x}} = \begin{pmatrix} \frac{\partial f(\overrightarrow{x})}{\partial x_1} & \frac{\partial f(\overrightarrow{x})}{\partial x_2} & \cdots & \frac{\partial f(\overrightarrow{x})}{\partial x_n} \end{pmatrix}$$

$$\overrightarrow{x} := \overrightarrow{x} + \alpha \left(\nabla_{\overrightarrow{x}} f \right)^T$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \alpha \begin{pmatrix} \frac{\partial f(\overrightarrow{x})}{\partial x_1} \\ \frac{\partial f(\overrightarrow{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\overrightarrow{x})}{\partial x_n} \end{pmatrix}$$

$$x_i := x_i + \alpha \frac{\partial f(\overrightarrow{x})}{\partial x_i}$$

Gradients and Parameter Update

$$-\nabla_{\overrightarrow{x}} f = \begin{pmatrix} -\frac{\partial f(\overrightarrow{x})}{\partial x_1} & -\frac{\partial f(\overrightarrow{x})}{\partial x_2} & \cdots & -\frac{\partial f(\overrightarrow{x})}{\partial x_n} \end{pmatrix}$$

$$\overrightarrow{x} := \overrightarrow{x} - \alpha\left(\nabla_{\overrightarrow{x}} f\right)^T$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \alpha \begin{pmatrix} \frac{\partial f(\overrightarrow{x})}{\partial x_1} \\ \frac{\partial f(\overrightarrow{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\overrightarrow{x})}{\partial x_n} \end{pmatrix}$$

$$x_i := x_i - \alpha\frac{\partial f(\overrightarrow{x})}{\partial x_i}$$

Gradients and Parameter Update

$$\nabla_{\vec{x}} f = \left(\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \ldots \quad \frac{\partial f}{\partial x_n} \right)$$ the most increasing direction

$$-\nabla_{\vec{x}} f = \left(-\frac{\partial f}{\partial x_1} \quad -\frac{\partial f}{\partial x_2} \quad \ldots \quad -\frac{\partial f}{\partial x_n} \right)$$ the most decreasing direction

$$\vec{x} := \vec{x} + \alpha \left(\nabla_{\vec{x}} f \right)^T$$

$$\vec{x} := \vec{x} - \alpha \left(\nabla_{\vec{x}} f \right)^T$$

Dot Product and Jacobians

$$\vec{u}^T = (u_1 \quad u_2 \quad u_3), \quad \vec{v}^T = (v_1 \quad v_2 \quad v_3)$$

$$z = f(\vec{u}, \vec{v}) = u_1 v_1 + u_2 v_2 + u_3 v_3$$

$$\frac{\partial z}{\partial \vec{u}} \qquad \frac{\partial z}{\partial \vec{v}}$$

$$\frac{\partial z}{\partial u_i} = v_i, \ 1 \le i \le 3$$

$$\frac{\partial z}{\partial \vec{u}} = (v_1 \quad v_2 \quad v_3) \qquad \frac{\partial z}{\partial \vec{v}} = (u_1 \quad u_2 \quad u_3)$$

Dot Product and Jacobians

$$\vec{u}, \vec{v} \in \mathbb{R}^n$$

$$z = f(\vec{u}, \vec{v}) = \sum_{i=1}^{n} u_i v_i$$

$$\frac{\partial z}{\partial u_i} = v_i, \frac{\partial z}{\partial v_j} = u_j$$

$$\frac{\partial z}{\partial \vec{u}} = \begin{pmatrix} \frac{\partial z}{\partial u_1} & \frac{\partial z}{\partial u_2} & \dots & \frac{\partial z}{\partial u_n} \end{pmatrix} = \vec{v}^T$$

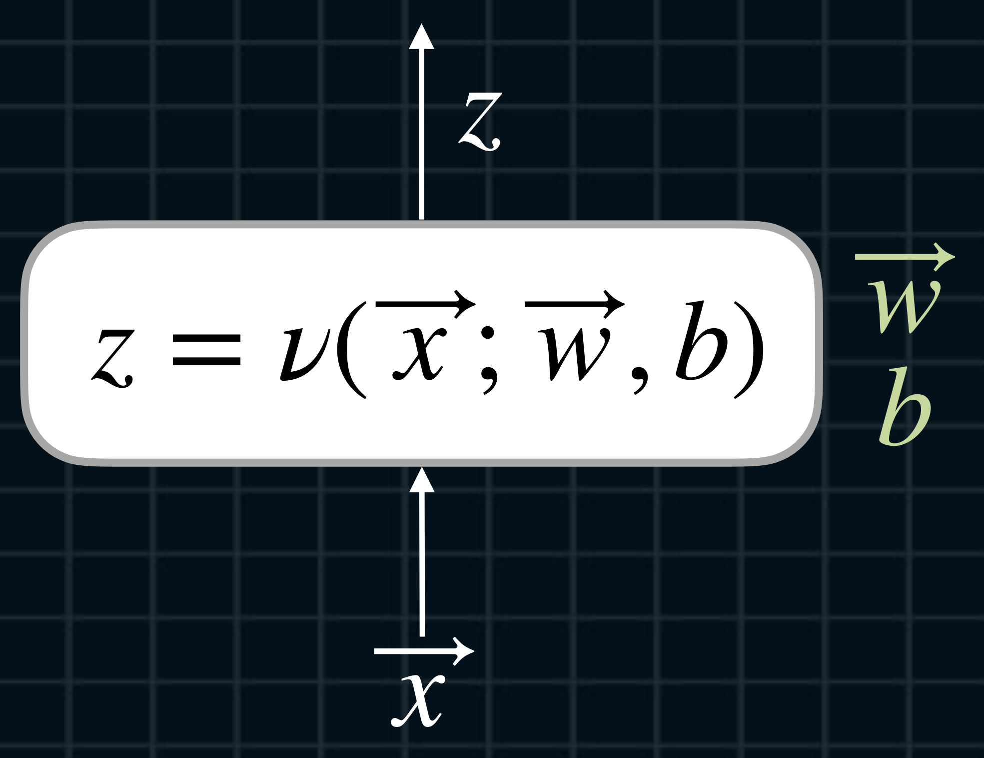
$$\frac{\partial z}{\partial \vec{v}} = \begin{pmatrix} \frac{\partial z}{\partial v_1} & \frac{\partial z}{\partial v_2} & \dots & \frac{\partial z}{\partial v_n} \end{pmatrix} = \vec{u}^T$$

Affine Functions and Jacobians

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

$$\frac{\partial z}{\partial \overrightarrow{x}} \qquad \frac{\partial z}{\partial \overrightarrow{w}} \qquad \frac{\partial z}{\partial b}$$

$$\frac{\partial z}{\partial \overrightarrow{x}} = \overrightarrow{w}^T, \quad \frac{\partial z}{\partial \overrightarrow{w}} = \overrightarrow{x}^T, \quad \frac{\partial z}{\partial b} = 1$$



$$\frac{\partial z}{\partial \overrightarrow{x}} = \overrightarrow{w}^T$$

$$\frac{\partial z}{\partial \overrightarrow{w}} = \overrightarrow{x}^T$$

$$\frac{\partial z}{\partial b} = 1$$

Affine Functions and Jacobians

z

$$z = \nu(\vec{x}; \vec{w}, b)$$

$\vec{w}$
b

$\vec{x}$

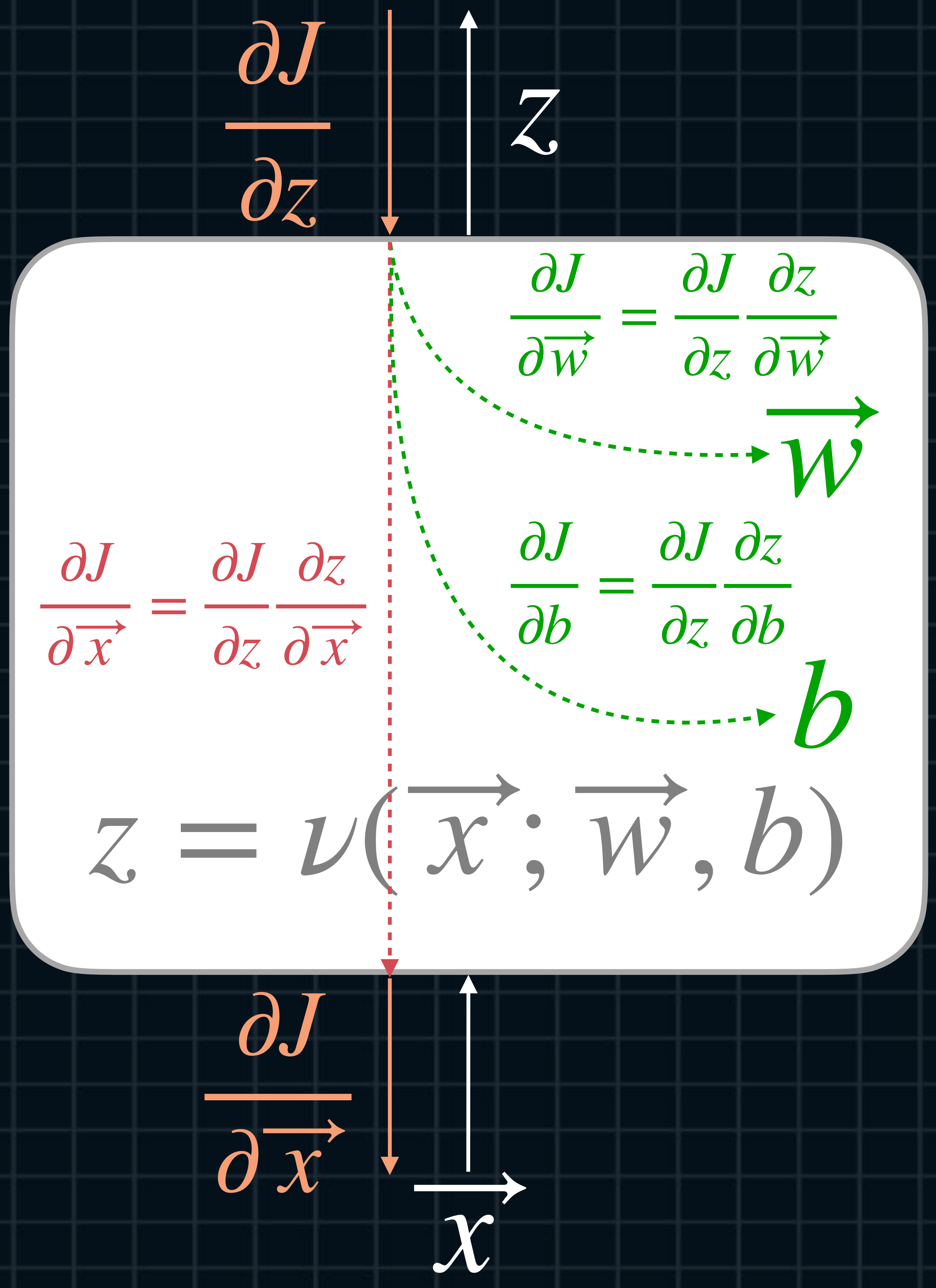
$$\frac{\partial J}{\partial z}\frac{\partial z}{\partial \vec{x}} = \frac{\partial J}{\partial \vec{x}} = \frac{\partial J}{\partial z} \cdot \vec{w}^T$$

$$\frac{\partial J}{\partial z}\frac{\partial z}{\partial \vec{w}} = \frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial z} \cdot \vec{x}^T$$

$$\frac{\partial J}{\partial z}\frac{\partial z}{\partial b} = \frac{\partial J}{\partial b} = \frac{\partial J}{\partial z}$$

$$\vec{w} := \vec{w} - \alpha\left(\frac{\partial J}{\partial \vec{w}}\right)^T, \; b := b - \alpha\frac{\partial J}{\partial b}$$

Affine Functions and Jacobians



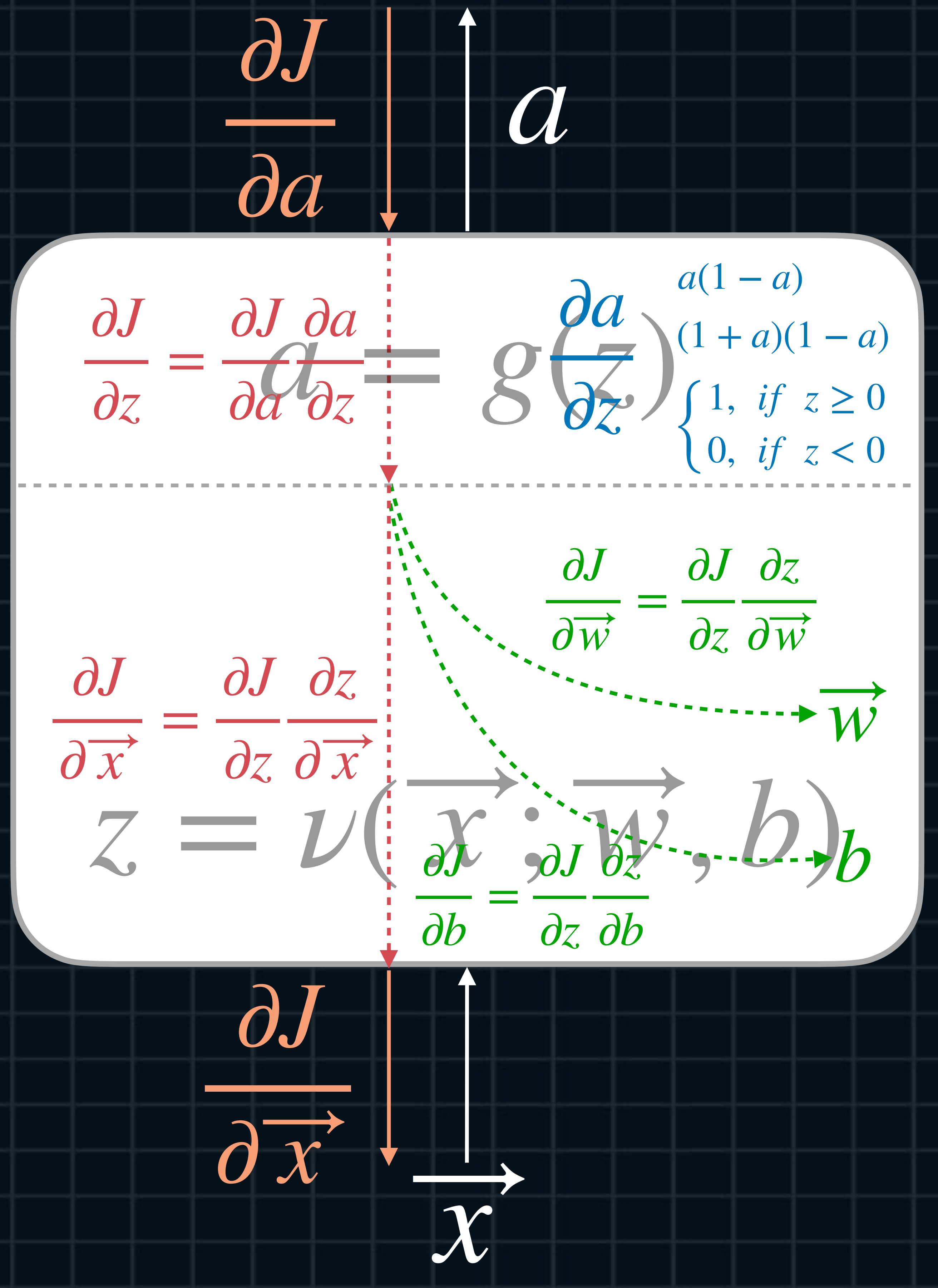
Activation Functions and Jacobians

$$\frac{\partial a}{\partial z} = a(1-a),\ if\ \ g = \sigma \qquad \frac{\partial J}{\partial z} = \frac{\partial J}{\partial a} \cdot a(1-a),\ if\ \ g = \sigma$$

$$\frac{\partial a}{\partial z} = (1+a)(1-a),\ if\ \ g = tanh \qquad \frac{\partial J}{\partial z} = \frac{\partial J}{\partial a} \cdot (1+a)(1-a),\ if\ \ g = tanh$$

$$\frac{\partial a}{\partial z} = \begin{cases} 1, & if\ \ z \geq 0 \\ 0, & if\ \ z < 0 \end{cases},\ \ if\ \ g = ReLU \qquad \frac{\partial J}{\partial z} = \frac{\partial J}{\partial a} \cdot \begin{cases} 1, & if\ \ z \geq 0 \\ 0, & if\ \ z < 0 \end{cases},\ \ if\ \ g = ReLU$$

Affine Functions and Jacobians



$$\frac{\partial J}{\partial \vec{x}} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial \vec{x}} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z} \cdot \vec{w}^T$$

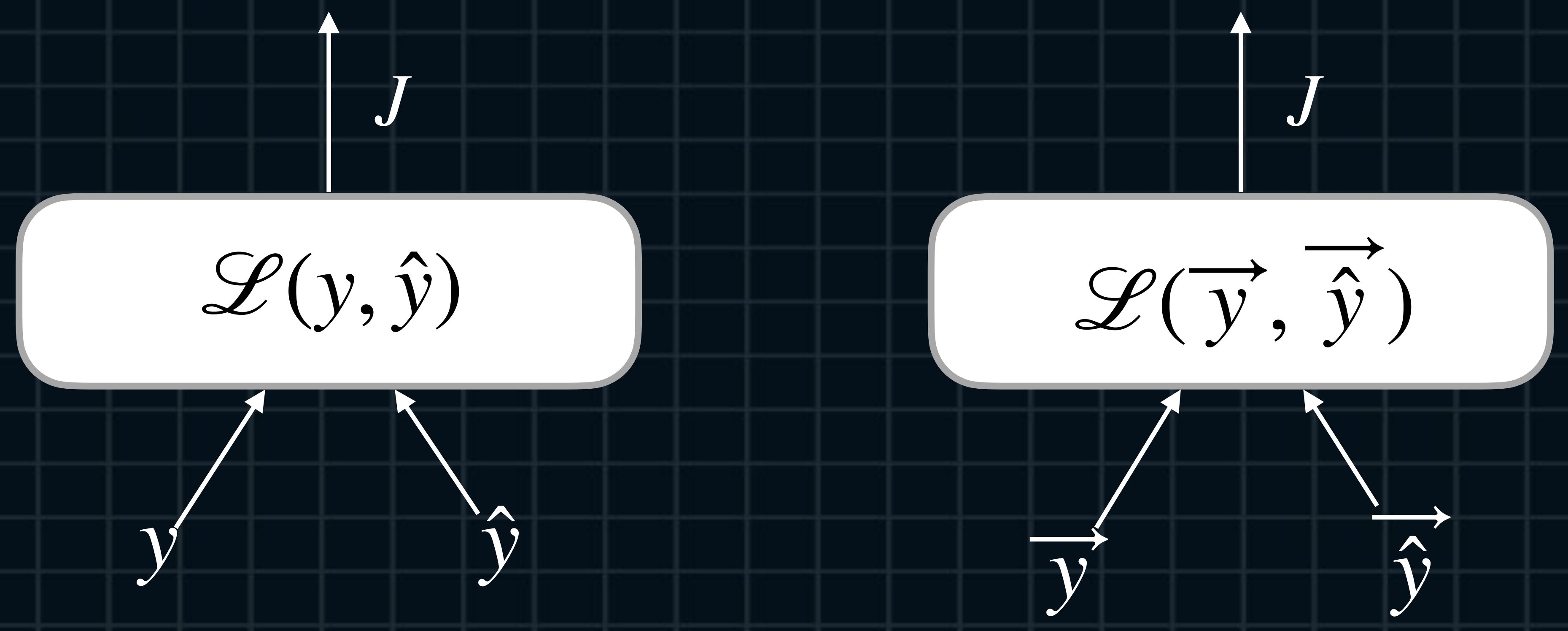
$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial \vec{w}} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z} \cdot \vec{x}^T$$

$$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial b} = \frac{\partial J}{\partial a}\frac{\partial a}{\partial z}$$

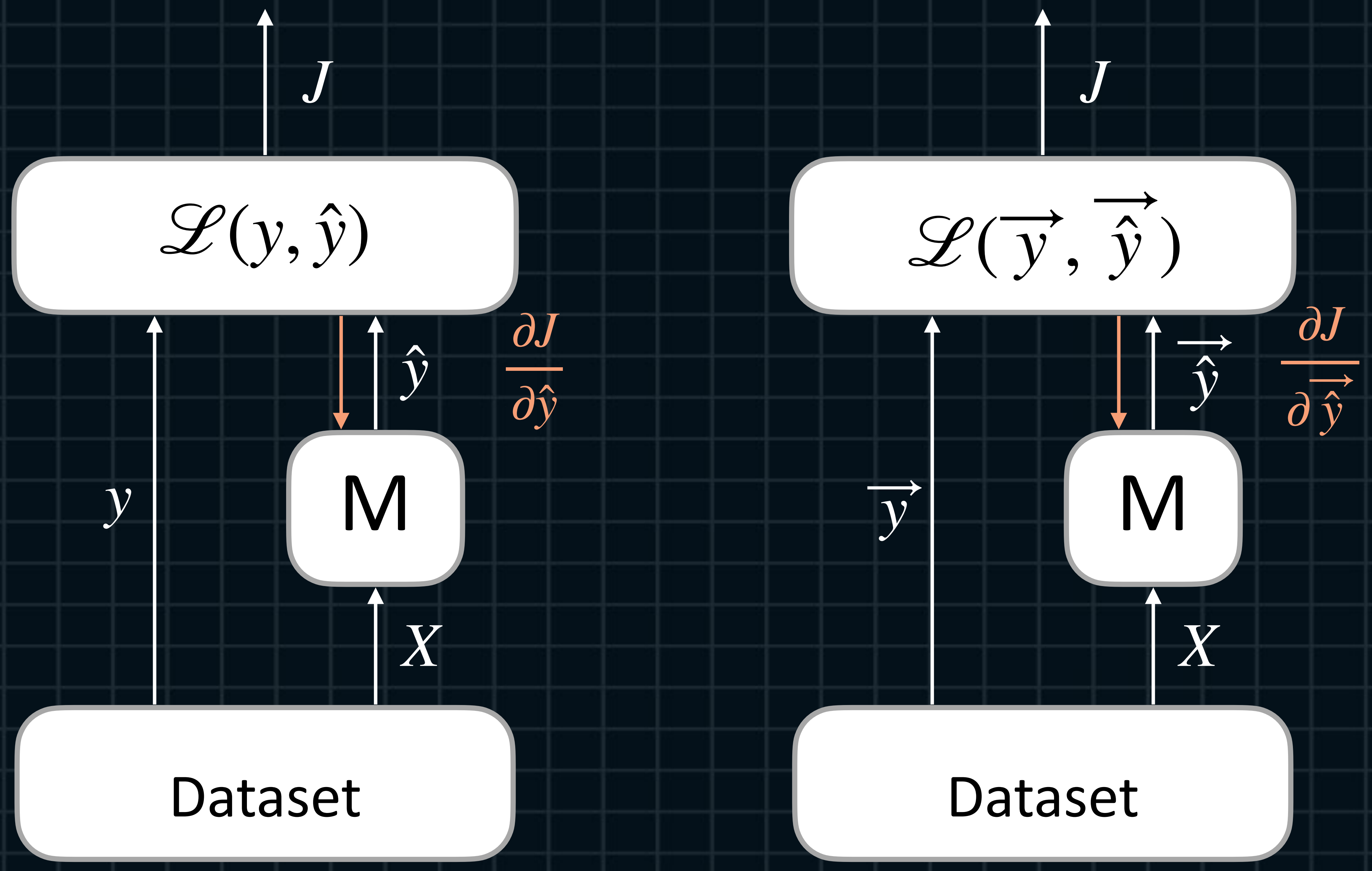
$$\vec{w} := \vec{w} - \alpha\left(\frac{\partial J}{\partial \vec{w}}\right)^T = \vec{w} - \alpha\frac{\partial J}{\partial a}\frac{\partial a}{\partial z} \cdot \vec{x}$$

$$b := b - \alpha\frac{\partial J}{\partial b} = b - \alpha\frac{\partial J}{\partial a}\frac{\partial a}{\partial z}$$

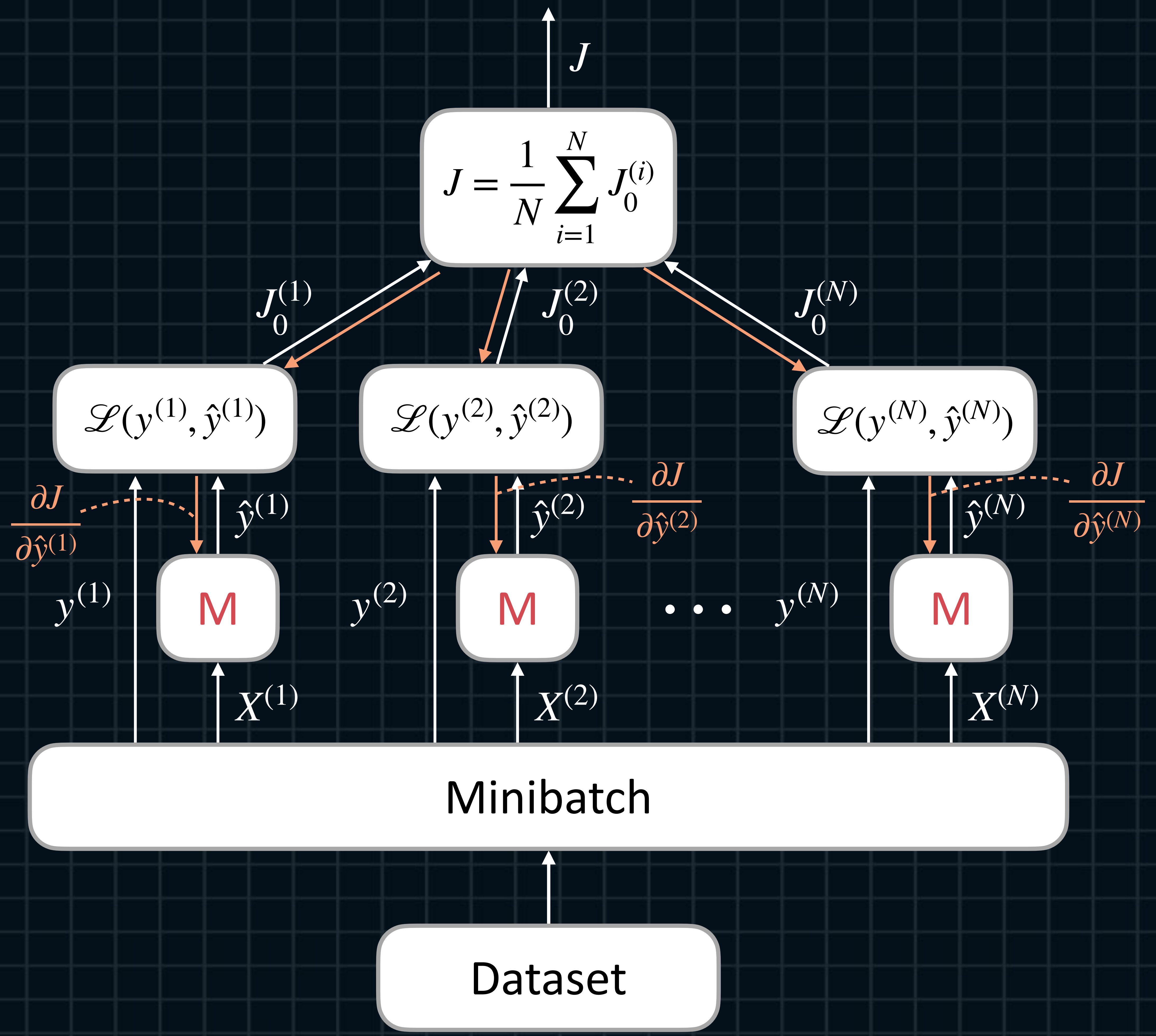
Loss Functions and Multivariate Functions



Loss Functions and Backpropagation

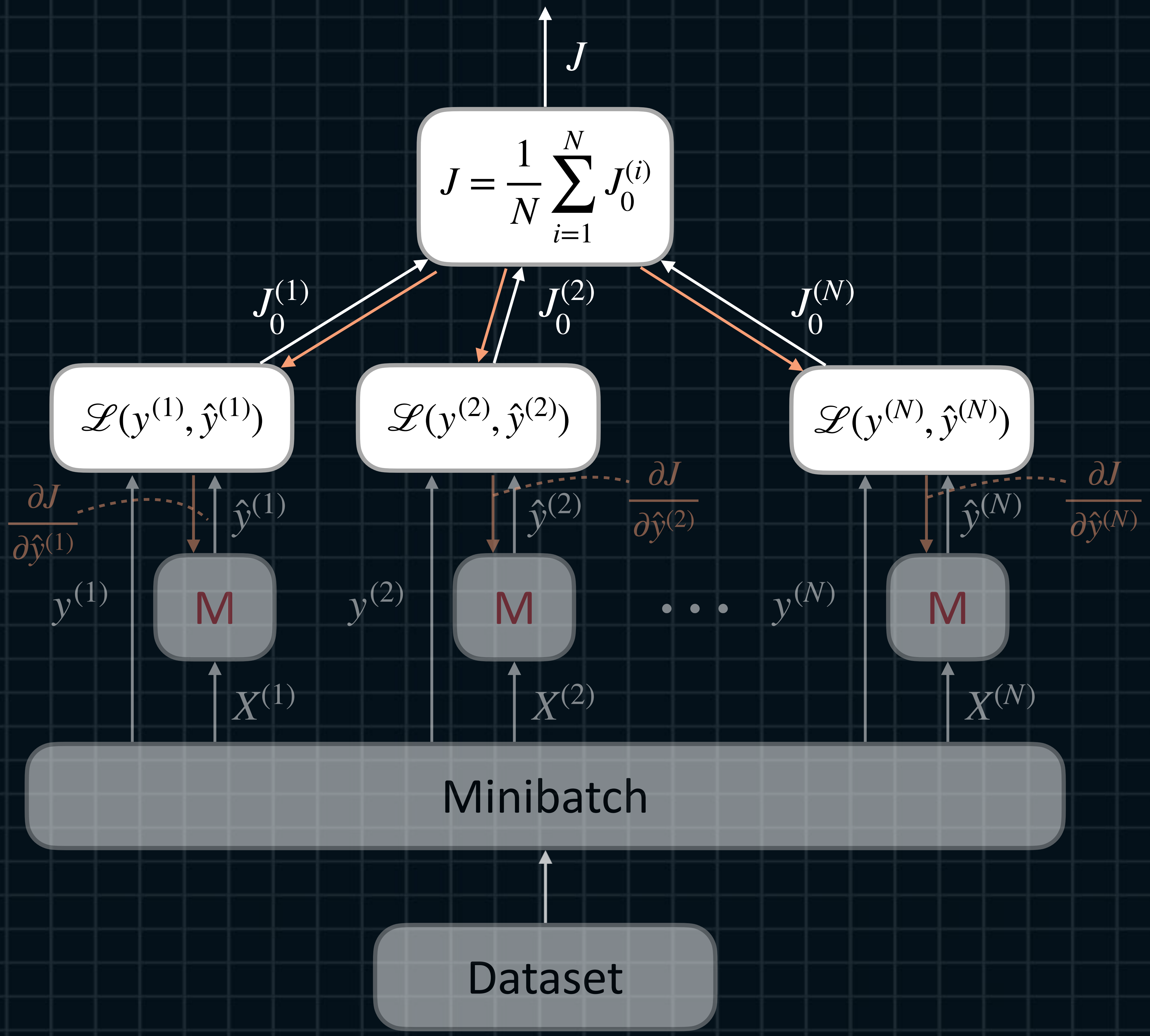


Loss Functions and Backpropagation



- Loss Functions and Jacobians

Gradients for Mini-batch



$$J_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(1)}\right)^2$$

$$J_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}log(\hat{y}^{(i)}) + (1 - y^{(i)})log(1 - \hat{y}^{(i)})\right]$$

$$J_{CCEE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_k^{(i)}log(\hat{y}_k^{(i)})\right]$$

Gradients for Mini-batch

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\left(\vec{J_0}\right)^T = \begin{pmatrix} \frac{\partial J}{\partial J_0^{(1)}} & \frac{\partial J}{\partial J_0^{(2)}} & \cdots & \frac{\partial J}{\partial J_0^{(N)}} \end{pmatrix}$$

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

$$\frac{\partial J}{\partial \vec{J_0}} = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

$$\frac{\partial J}{\partial J_0^{(i)}} = \frac{1}{N}$$

$$\frac{\partial J}{\partial \hat{y}^{(i)}} = \frac{\partial J}{\partial J_0^{(i)}} \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = \frac{1}{N} \cdot \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}}$$

$$\vec{J_0} \in \mathbb{R}^N \quad \frac{\partial J}{\partial \vec{J_0}} \in \mathbb{R}^{1\times N}$$

MSE and Jacobians

$$J_0^{(i)} = (y^{(i)} - \hat{y}^{(i)})^2$$

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

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

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

BCEE and Jacobians

$$J = -\left[ylog(\hat{y}) + (1-y)log(1-\hat{y})\right]$$

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

$$= -\left[\frac{\partial}{\partial \hat{y}}\left[ylog(\hat{y})\right] + \frac{\partial}{\partial \hat{y}}\left[(1-y)log(1-\hat{y})\right]\right]$$

$$= -\left(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}\right) = -\frac{y - y\hat{y} - \hat{y} + \hat{y}y}{\hat{y}(1-\hat{y})}$$

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

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

BCEE and Jacobians

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

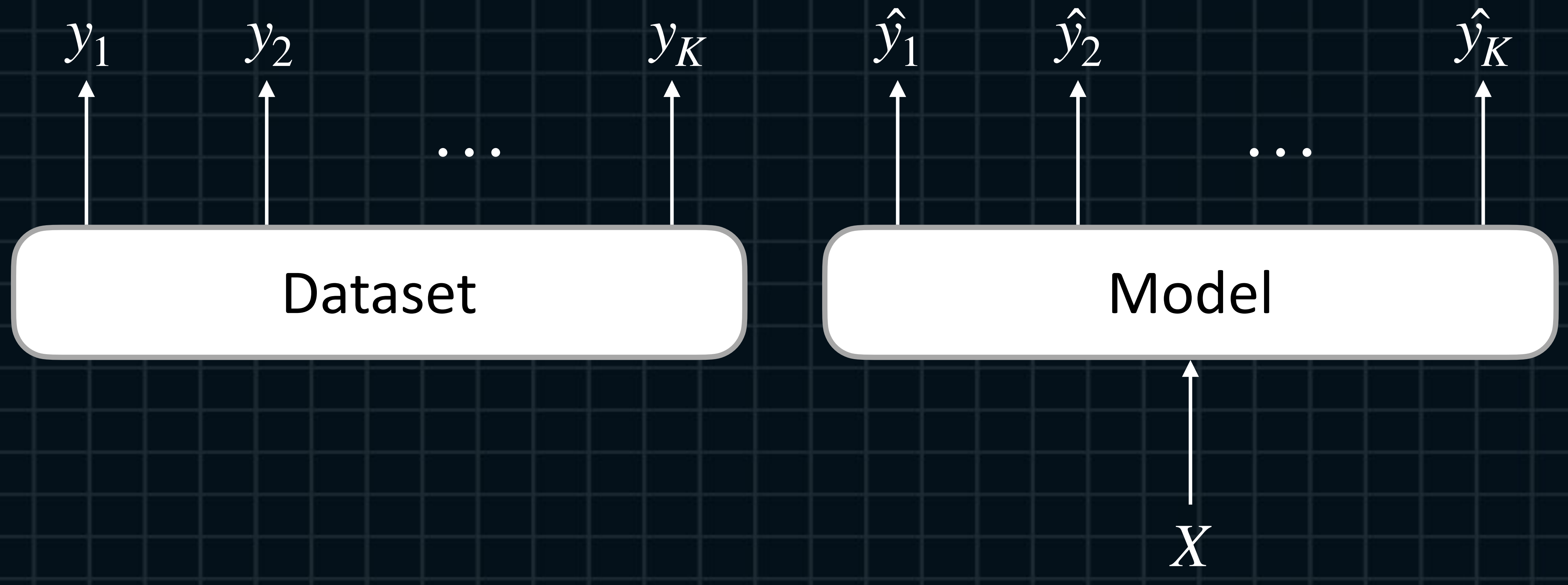
$$= -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}log(\hat{y}^{(i)}) + (1-y^{(i)})log(1-\hat{y}^{(i)})\right]$$

$$\frac{\partial J}{\partial \hat{y}^{(i)}} = \frac{\partial J}{\partial J_0^{(i)}}\frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = \frac{1}{N}\frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = \frac{1}{N}\frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}(1-\hat{y}^{(i)})}$$

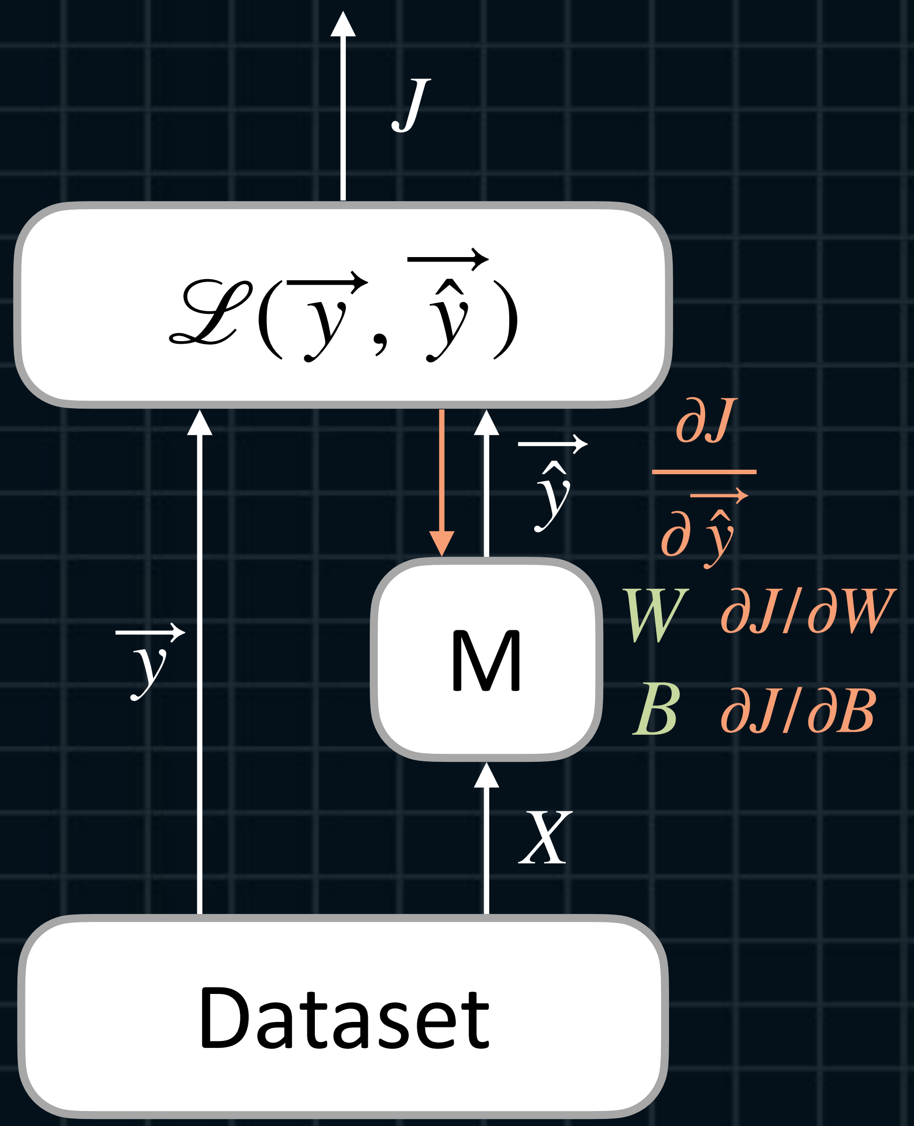
$$\frac{\partial J}{\partial \vec{\hat{y}}} = \begin{pmatrix} \frac{\partial J}{\partial \hat{y}^{(1)}} & \frac{\partial J}{\partial \hat{y}^{(2)}} & \cdots & \frac{\partial J}{\partial \hat{y}^{(N)}} \end{pmatrix}$$

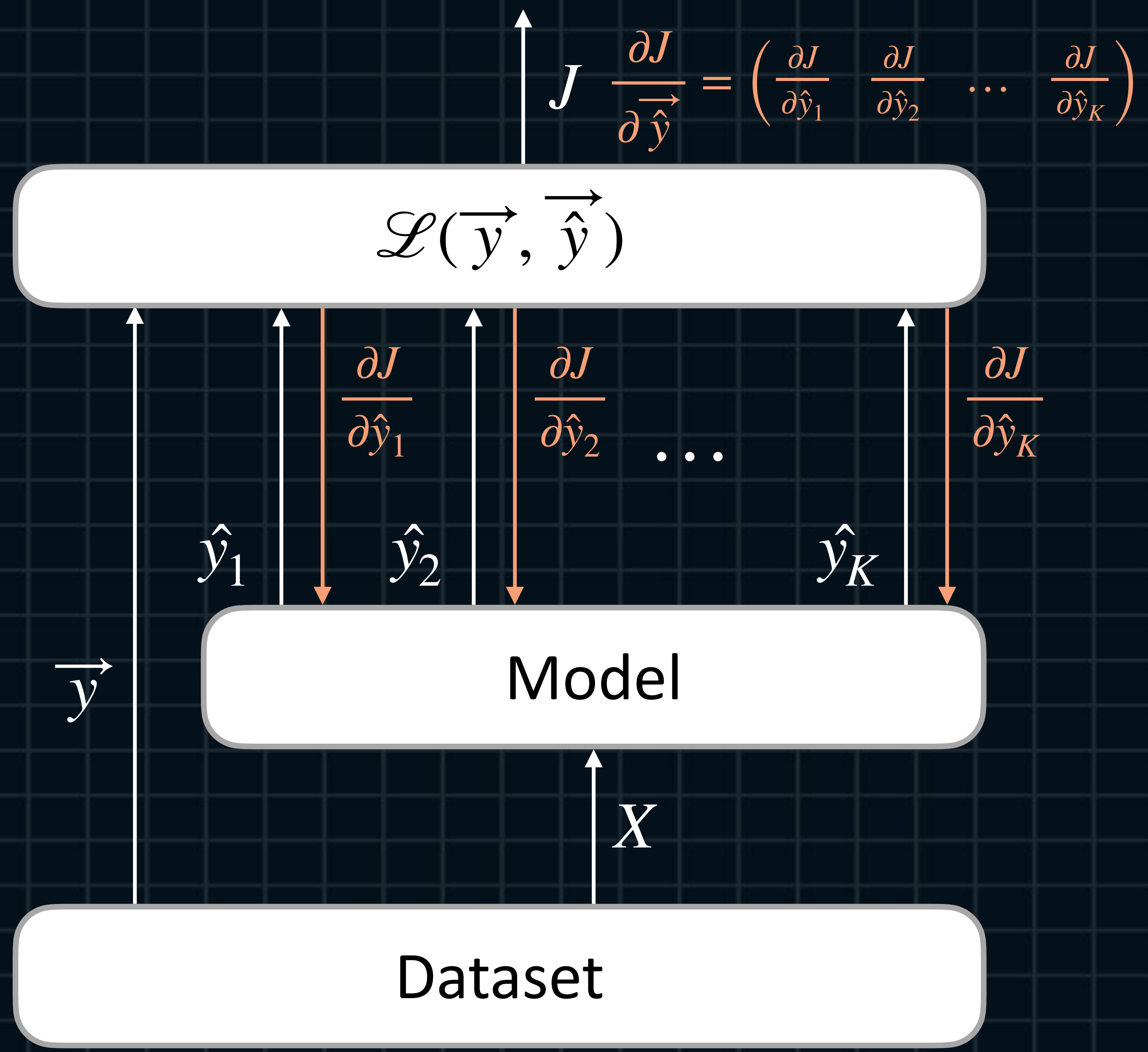
$$= \frac{1}{N}\begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}(1-\hat{y}^{(1)})} & \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}(1-\hat{y}^{(2)})} & \cdots & \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}(1-\hat{y}^{(N)})} \end{pmatrix}$$

CCEE and Jacobians

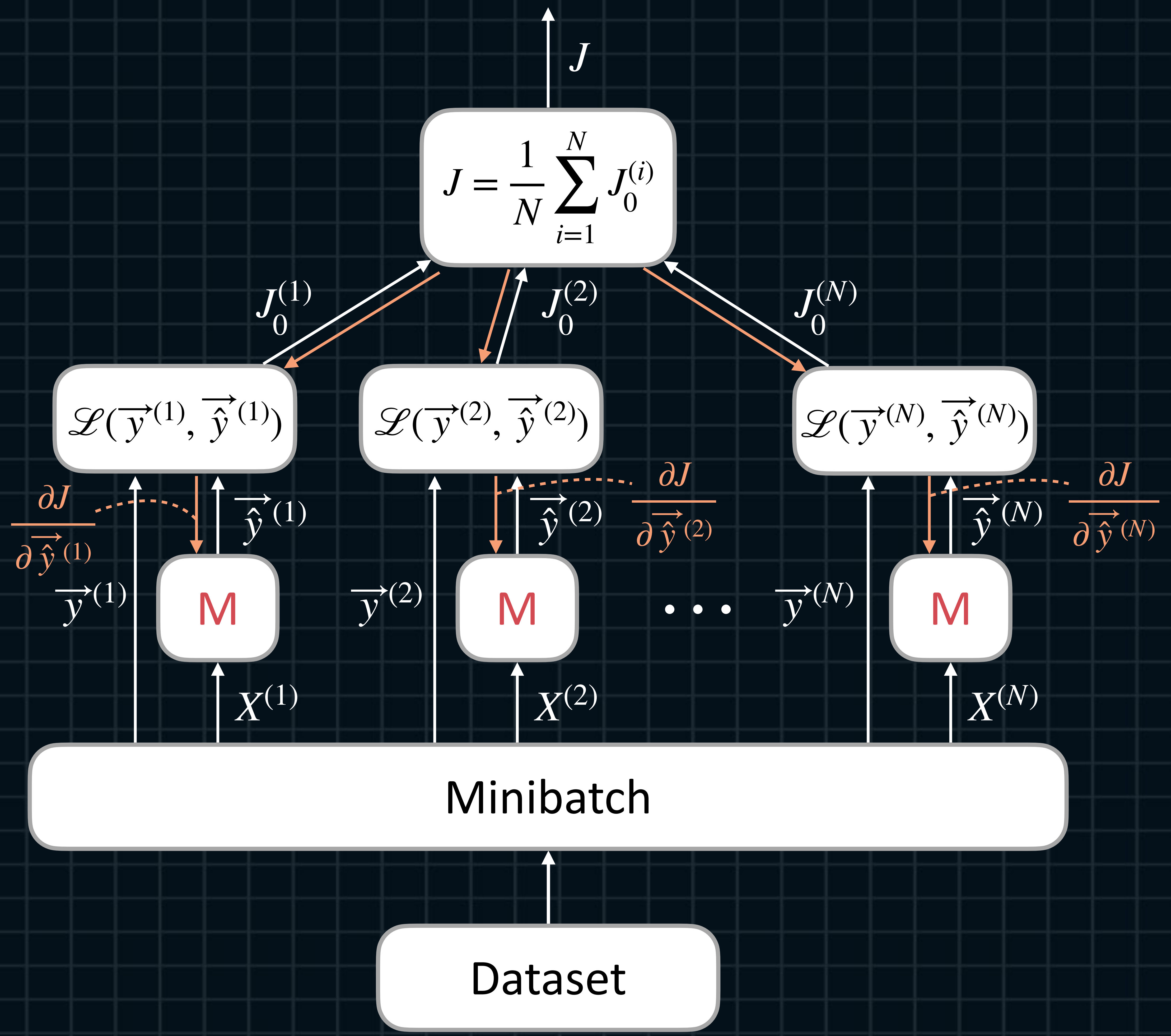


CCEE and Jacobians





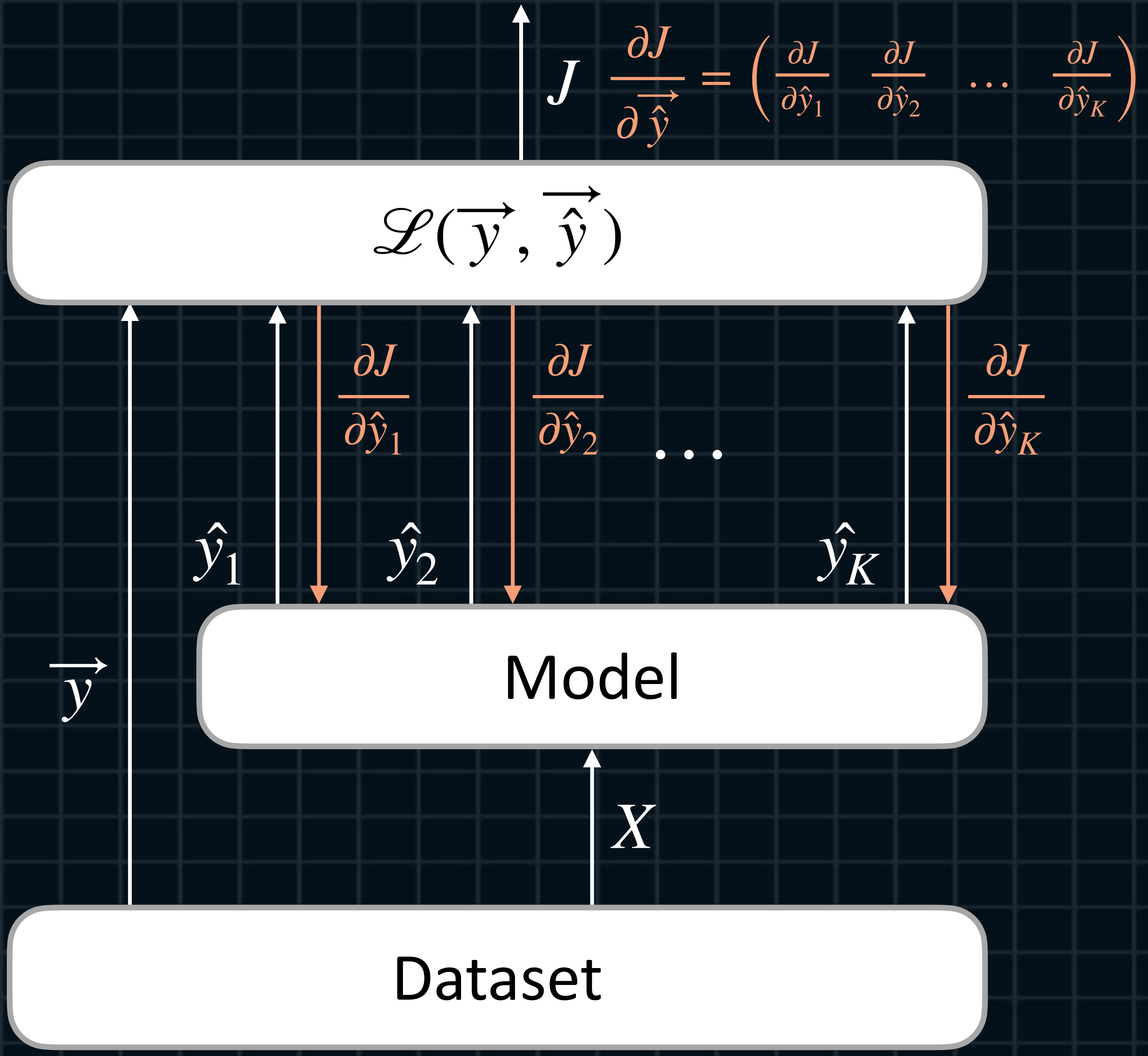
CCEE and Jacobians



Lecture.3 Multivariate Functions and Jacobians - Loss Functions and Jacobians

CCEE and Jacobians

$$J = -\sum_{k=1}^{K} y_k log(\hat{y}_k)$$



$$\frac{\partial J}{\partial \hat{y}_j} = \frac{\partial}{\partial \hat{y}_j}\left[-\sum_{i=1}^{K} y_i log(\hat{y}_i) \right]$$

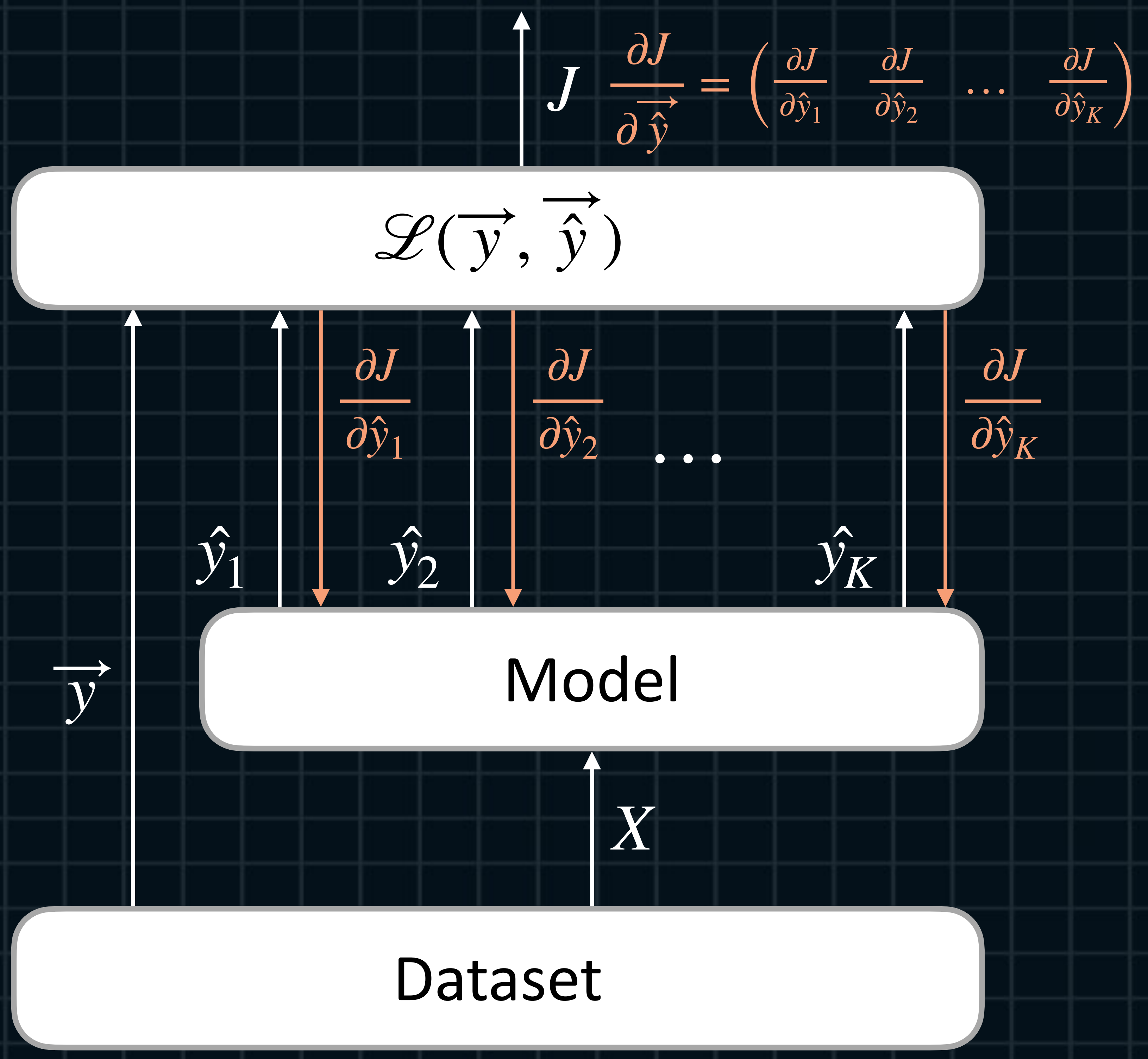
$$= -\frac{\partial}{\partial \hat{y}_j}\left[y_1 log(\hat{y}_1) + y_2 log(\hat{y}_2) + \dots + y_j log(\hat{y}_j) + \dots + y_K log(\hat{y}_K) \right]$$

$$= -\frac{y_j}{\hat{y}_j}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \left(\frac{\partial J}{\partial \hat{y}_1} \quad \frac{\partial J}{\partial \hat{y}_2} \quad \dots \quad \frac{\partial J}{\partial \hat{y}_K} \right)$$

$$= -\left(\frac{y_1}{\hat{y}_1} \quad \frac{y_2}{\hat{y}_2} \quad \dots \quad \frac{y_K}{\hat{y}_K} \right)$$

CCEE and Jacobians



$$y_i \in (0,1),\ \sum_{i=1}^{K} y_i = 1$$

$$(\vec{y})^T = (y_1 \quad y_2 \quad \cdots \quad y_K) = (1 \quad 0 \quad \dots \quad 0)$$

$$\frac{\partial J}{\partial \hat{y}_i} = \begin{cases} -1/\hat{y}_i, & \text{if } i = 1 \\ 0, & \text{if } i \neq 1 \end{cases}$$

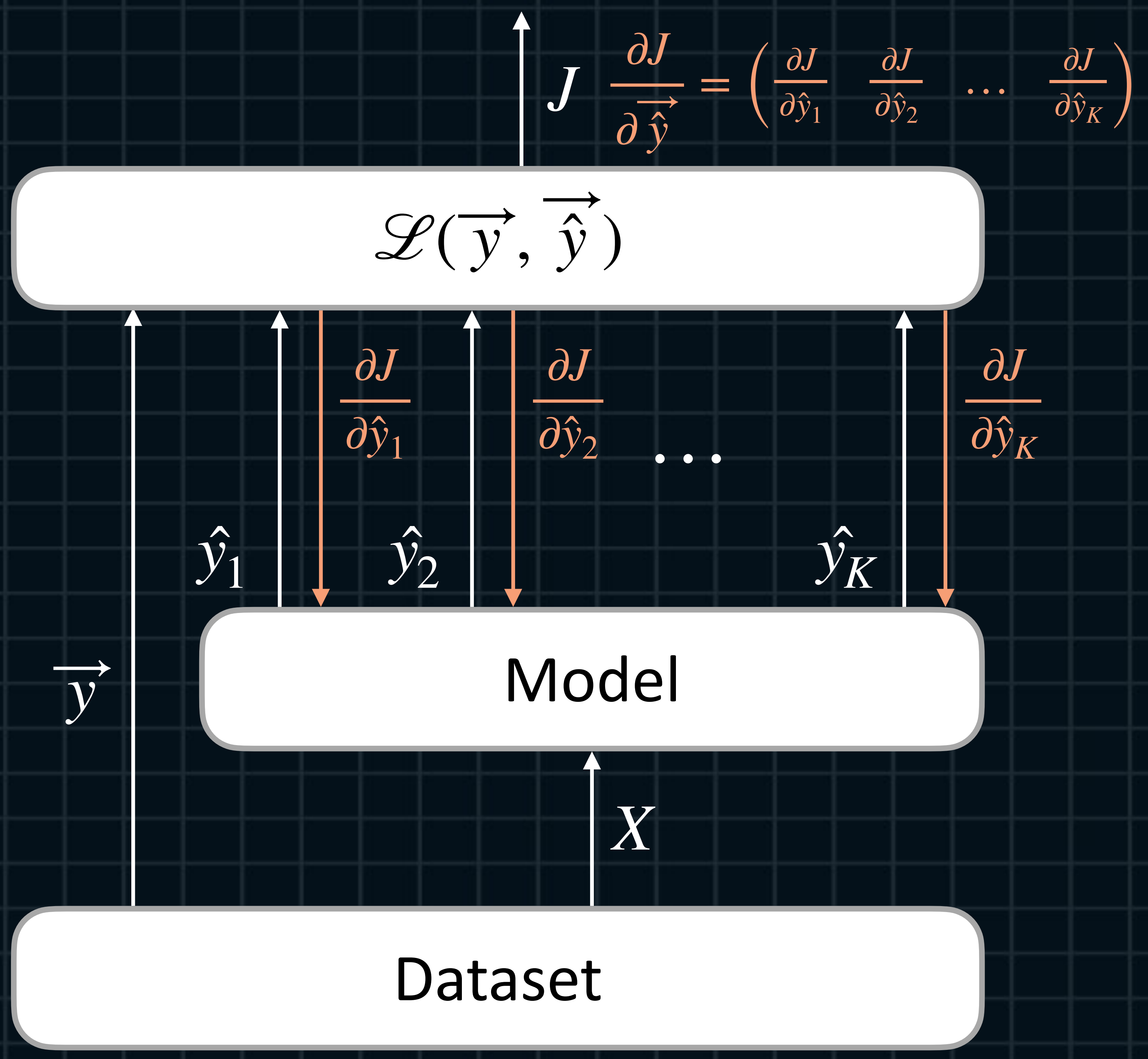
$$\frac{\partial J}{\partial \vec{\hat{y}}} = \begin{pmatrix} -\frac{1}{\hat{y}_1} & 0 & \dots & 0 \end{pmatrix}$$

$$(\vec{y})^T = (y_1 \quad y_2 \quad \cdots \quad y_K) = (0 \quad 0 \quad \dots \quad 1)$$

$$\frac{\partial J}{\partial \hat{y}_i} = \begin{cases} -1/\hat{y}_i, & \text{if } i = K \\ 0, & \text{if } i \neq K \end{cases}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \begin{pmatrix} 0 & 0 & \dots & -\frac{1}{\hat{y}_K} \end{pmatrix}$$

CCEE and Jacobians



$$label = \alpha \;\Rightarrow\; y_i = \begin{cases} 1, \; i = \alpha \\ 0, \; i \neq \alpha \end{cases}$$

$$J = -\sum_{i=1}^{K} y_i log(\hat{y}_i) = -\,log(\hat{y}_\alpha)$$

$$\frac{\partial J}{\partial \hat{y}_i} = \begin{cases} -1/\hat{y}_\alpha, & i = \alpha \\ 0, & i \neq \alpha \end{cases}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \left(\frac{\partial J}{\partial \hat{y}_1} \quad \frac{\partial J}{\partial \hat{y}_2} \quad \dots \quad \frac{\partial J}{\partial \hat{y}_\alpha} \quad \dots \quad \frac{\partial J}{\partial \hat{y}_K}\right)$$

$$= \left(0 \quad 0 \quad \dots \quad -1/\hat{y}_\alpha \quad \dots \quad 0\right)$$

CCEE and Jacobians

$$J_0^{(j)} = -\sum_{i=1}^{K} y_i^{(j)} log(\hat{y}_i^{(j)})$$

$$1 \le j \le N, 1 \le i \le K$$

$$\frac{\partial J_0^{(j)}}{\partial \hat{y}_i^{(j)}} = -\frac{y_i^{(j)}}{\hat{y}_i^{(j)}}$$

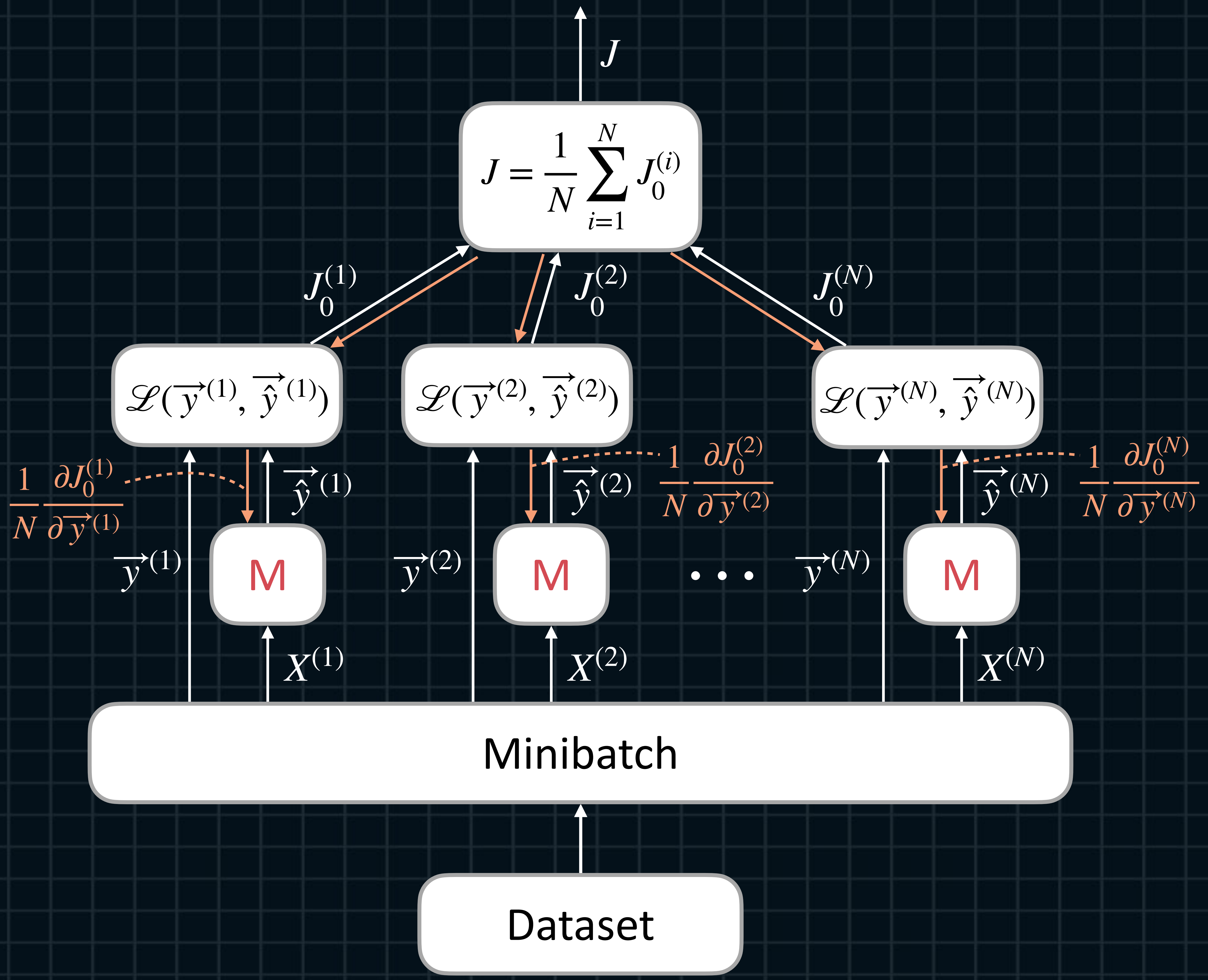
$$\frac{\partial J_0^{(j)}}{\partial \vec{\hat{y}}^{(j)}} = \begin{pmatrix} \frac{\partial J}{\partial \hat{y}_1^{(j)}} & \frac{\partial J}{\partial \hat{y}_2^{(j)}} & \cdots & \frac{\partial J}{\partial \hat{y}_K^{(j)}} \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{y_1^{(j)}}{\hat{y}_1^{(j)}} & -\frac{y_2^{(j)}}{\hat{y}_2^{(j)}} & \cdots & -\frac{y_K^{(j)}}{\hat{y}_K^{(j)}} \end{pmatrix}$$

$$J = \frac{1}{N}\sum_{j=1}^{N} J_0^{(j)} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K} y_i^{(j)} log(\hat{y}_i^{(j)})$$

$$\frac{\partial J}{\partial \vec{\hat{y}}^{(j)}} = \frac{\partial J}{\partial J_0^{(j)}}\frac{\partial J_0^{(j)}}{\partial \vec{\hat{y}}^{(j)}} = \frac{1}{N}\begin{pmatrix} -\frac{y_1^{(j)}}{\hat{y}_1^{(j)}} & -\frac{y_2^{(j)}}{\hat{y}_2^{(j)}} & \cdots & -\frac{y_K^{(j)}}{\hat{y}_K^{(j)}} \end{pmatrix}$$

CCEE and Jacobians



$$J = \frac{1}{N}\sum_{j=1}^{N} J_0^{(j)} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K} y_i^{(j)} log(\hat{y}_i^{(j)})$$

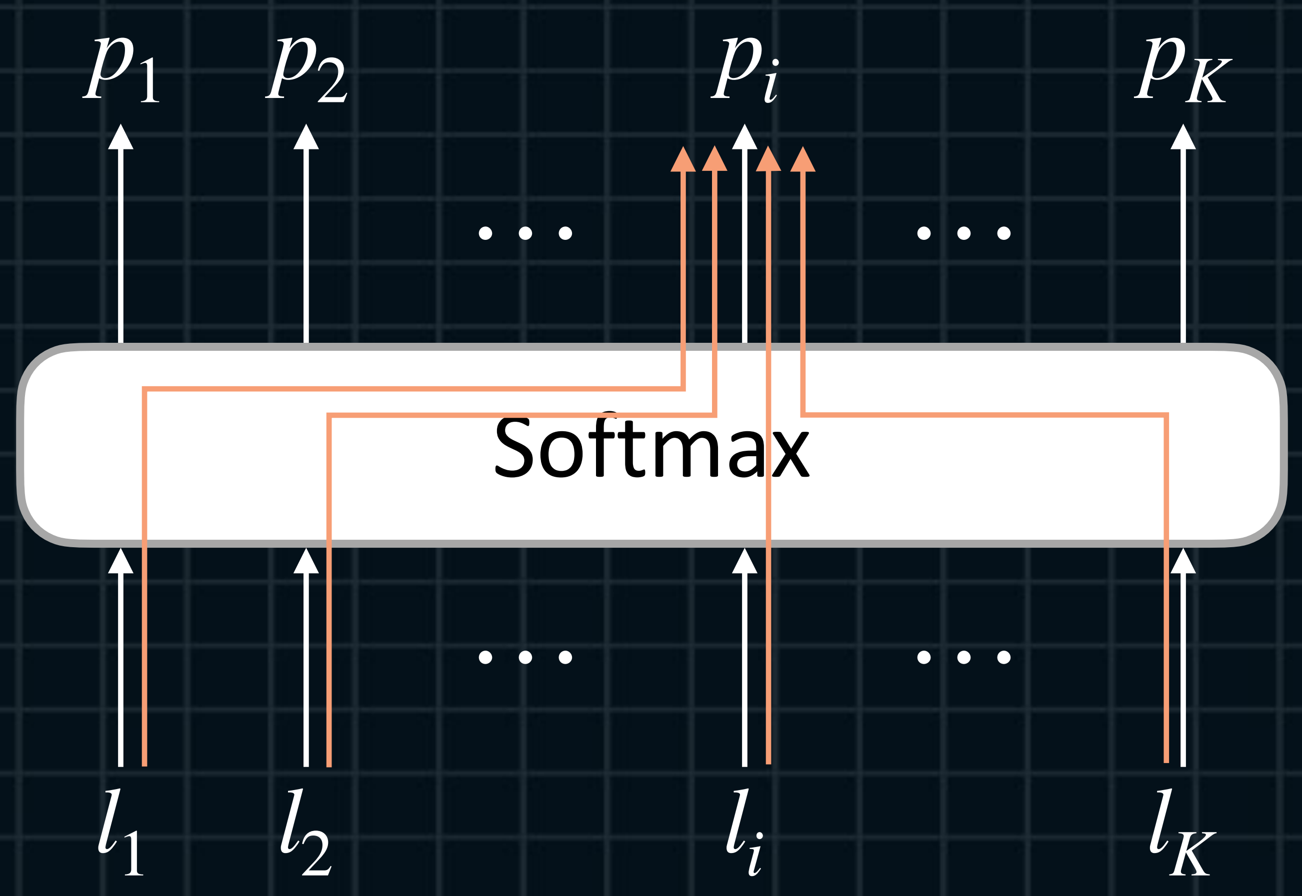
$$\frac{\partial J}{\partial \hat{\vec{y}}^{(j)}} = \frac{\partial J}{\partial J_0^{(j)}}\frac{\partial J_0^{(j)}}{\partial \hat{\vec{y}}^{(j)}} = \frac{1}{N}\left(-\frac{y_1^{(j)}}{\hat{y}_1^{(j)}} \quad -\frac{y_2^{(j)}}{\hat{y}_2^{(j)}} \quad \cdots \quad -\frac{y_K^{(j)}}{\hat{y}_K^{(j)}}\right)$$

- Softmax and Jacobians

Calculation of Softmax

$$p_i = \frac{e^{l_i}}{\sum_{k=1}^{K} e^{l_k}} = \frac{e^{l_i}}{e^{l_1} + e^{l_2} + \ldots + e^{l_K}}$$

$$p_i = \frac{e^{l_i}}{S}$$



$$\frac{\partial p_i}{\partial l_j} = \begin{cases} \frac{\partial p_i}{\partial l_i}, & i = j \\ \frac{\partial p_i}{\partial l_j}, & i \neq j \end{cases}$$

Diff. of the Denominator

$$
\begin{aligned}
\frac{\partial S}{\partial l_j} &= \frac{\partial}{\partial l_j}\left[\sum_{k=1}^{K} e^k\right] \\
&= \frac{\partial}{\partial l_j}\left[e^{l_1} + e^{l_2} + \ldots + e^{l_j} + \ldots + e^{l_K}\right] \\
&= \frac{\partial}{\partial l_j}\left[e^{l_1}\right] + \frac{\partial}{\partial l_j}\left[e^{l_2}\right] + \ldots + \frac{\partial}{\partial l_j}\left[e^{l_j}\right] + \ldots + \frac{\partial}{\partial l_j}\left[e^{l_K}\right] \\
&= e^{l_j}
\end{aligned}
$$

Jacobians of Softmax

$if \;\; i = j$

$$\frac{\partial p_i}{\partial l_i} = \frac{\partial}{\partial l_i}\left[\frac{e^{l_i}}{S}\right] = \frac{\frac{\partial e^{l_i}}{\partial l_i} \cdot S - e^{l_i} \cdot \frac{\partial S}{\partial l_i}}{S^2}$$

$$= \frac{e^{l_i} \cdot S - e^{2l_i}}{S^2} = \frac{e^{l_i}\left(S - e^{l_i}\right)}{S^2}$$

$$= \frac{e^{l_i}}{S} \cdot \frac{S - e^{l_i}}{S} = p_i(1 - p_i)$$

$$\frac{\partial p_i}{\partial l_i} = p_i(1 - p_i)$$

$if \;\; i \neq j$

$$\frac{\partial p_i}{\partial l_j} = \frac{\partial}{\partial l_j}\left[\frac{e^{l_i}}{S}\right] = \frac{\frac{\partial e^{l_i}}{\partial l_j} \cdot S - e^{l_i} \cdot \frac{\partial S}{\partial l_j}}{S^2}$$

$$= \frac{-e^{l_i}e^{l_j}}{S^2} = -\frac{e^{l_i}}{S} \cdot \frac{e^{l_j}}{S}$$

$$= -p_i p_j$$

$$\frac{\partial p_i}{\partial l_j} = -p_i p_j$$

Jacobians of Softmax

$$\frac{\partial p_i}{\partial l_j} = \begin{cases} p_i(1-p_i), & i = j \\ -p_i p_j, & i \neq j \end{cases}$$

$$\frac{\partial p_i}{\partial l_j} = \begin{cases} -p_i p_i + p_i, & i = j \\ -p_i p_j, & i \neq j \end{cases}$$

- Softmax and Jacobians

Jacobians of Softmax

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

$$\begin{aligned}\frac{\partial p_i}{\partial \vec{l}} &= \begin{pmatrix} \frac{\partial p_i}{\partial l_1} & \frac{\partial p_i}{\partial l_2} & \dots & \frac{\partial p_i}{\partial l_i} & \dots & \frac{\partial p_i}{\partial l_K} \end{pmatrix} \\ &= \begin{pmatrix} -p_i p_1 & -p_i p_2 & \dots & p_i(1-p_i) & \dots & -p_i p_K \end{pmatrix}\end{aligned}$$

Jacobians of Softmax

$$\frac{\partial p_1}{\partial \vec{l}} = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_i & \dots & -p_1p_K \end{pmatrix}$$

$$\frac{\partial p_2}{\partial \vec{l}} = \begin{pmatrix} -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_i & \dots & -p_2p_K \end{pmatrix}$$

$$\vdots$$

$$\frac{\partial p_i}{\partial \vec{l}} = \begin{pmatrix} -p_ip_1 & -p_ip_2 & \dots & p_i(1-p_i) & \dots & -p_ip_K \end{pmatrix}$$

$$\vdots$$

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

Backpropagation and Jacobian Matrices

Lecture.3
Multivariate Functions and Jacobians