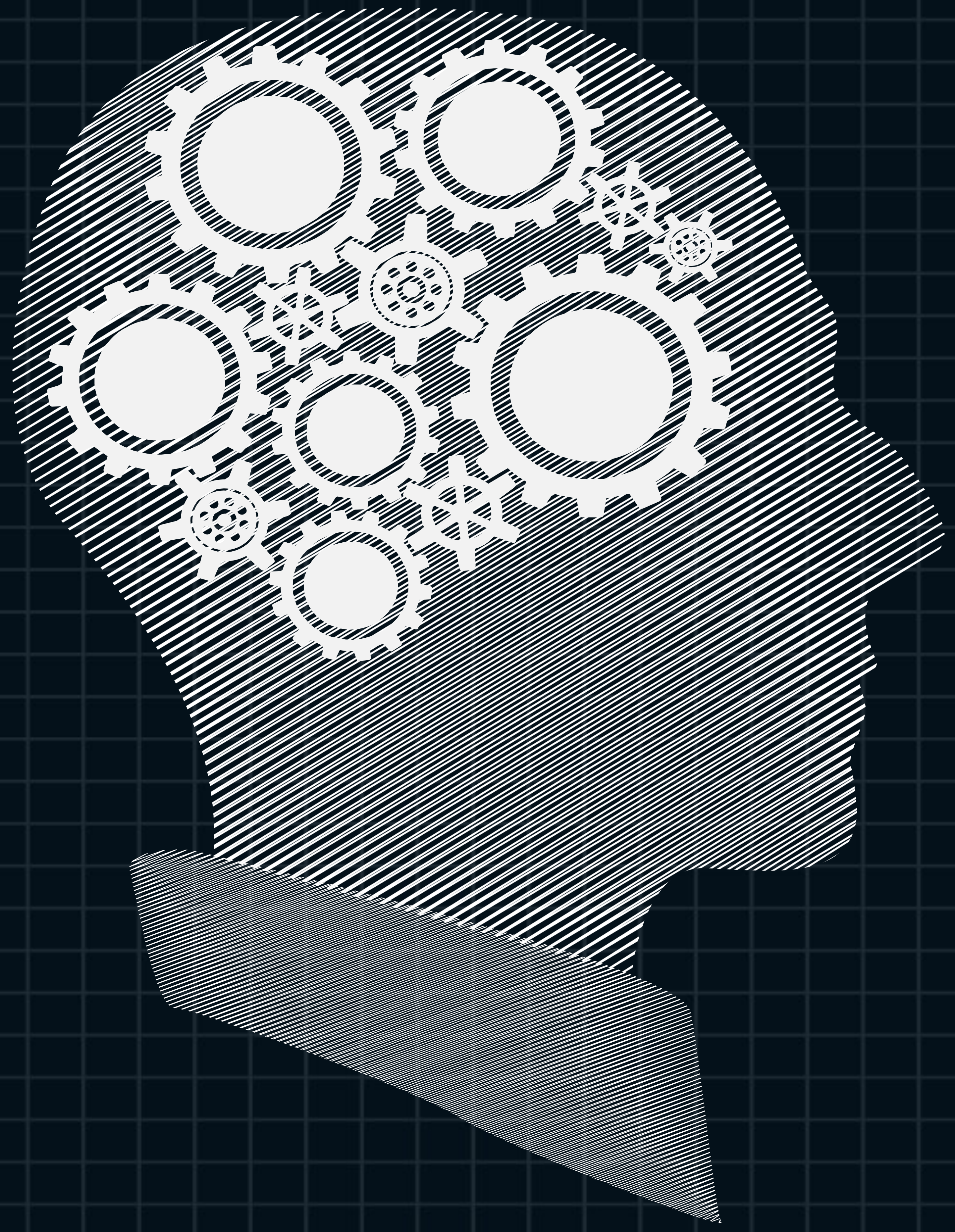
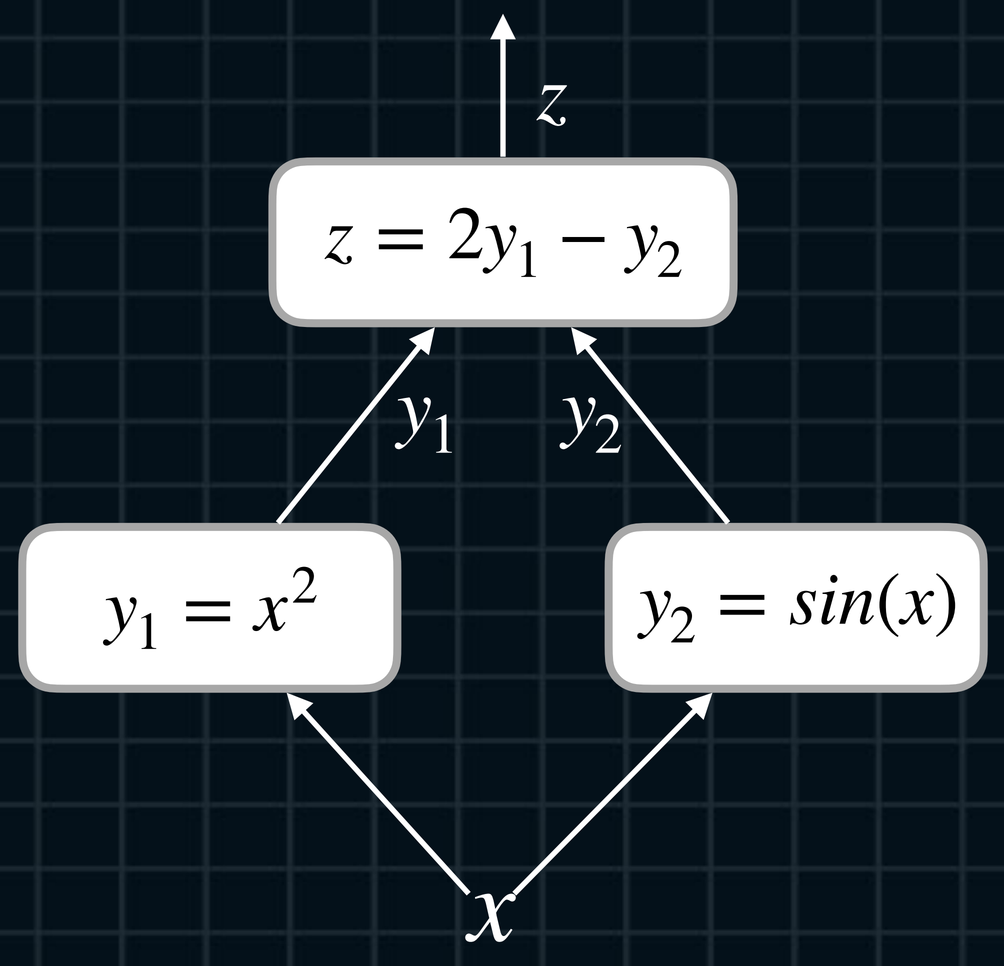


Backpropagation and Jacobian Matrices

Lecture.8
Total Derivatives

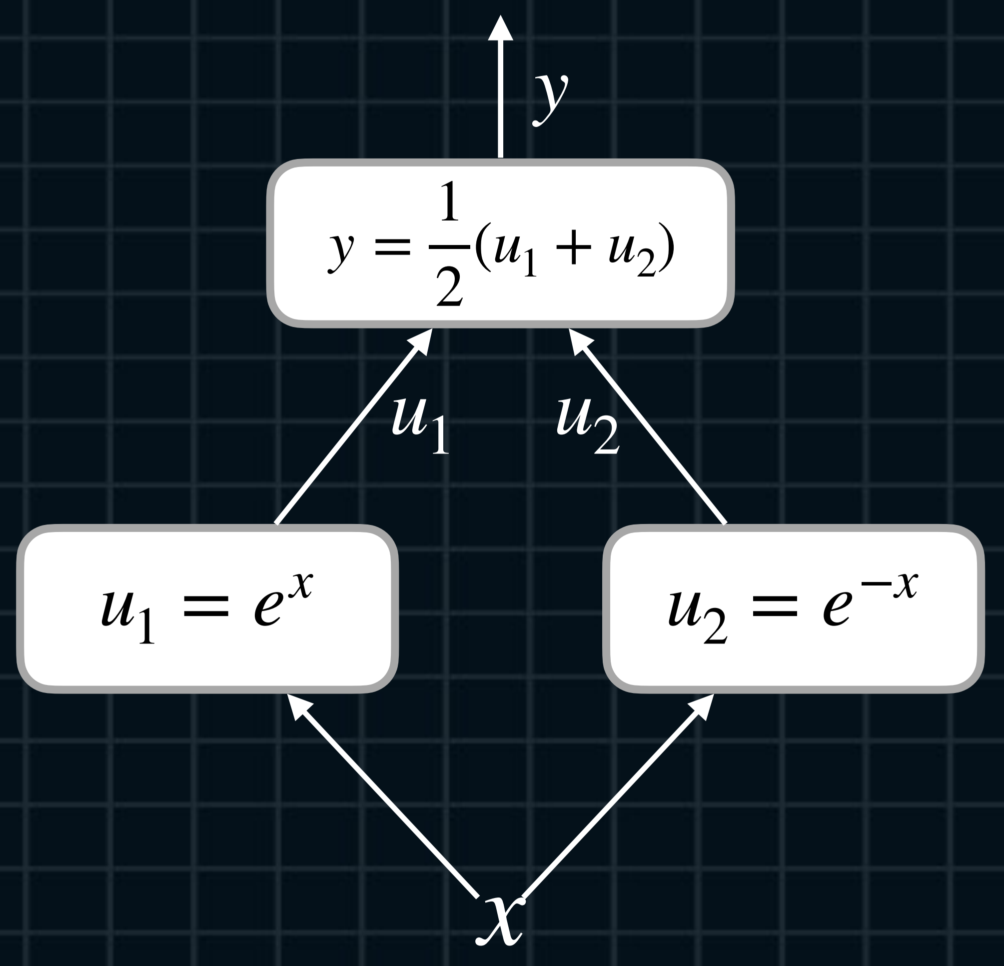
- Function with Multipath

Multipath Examples

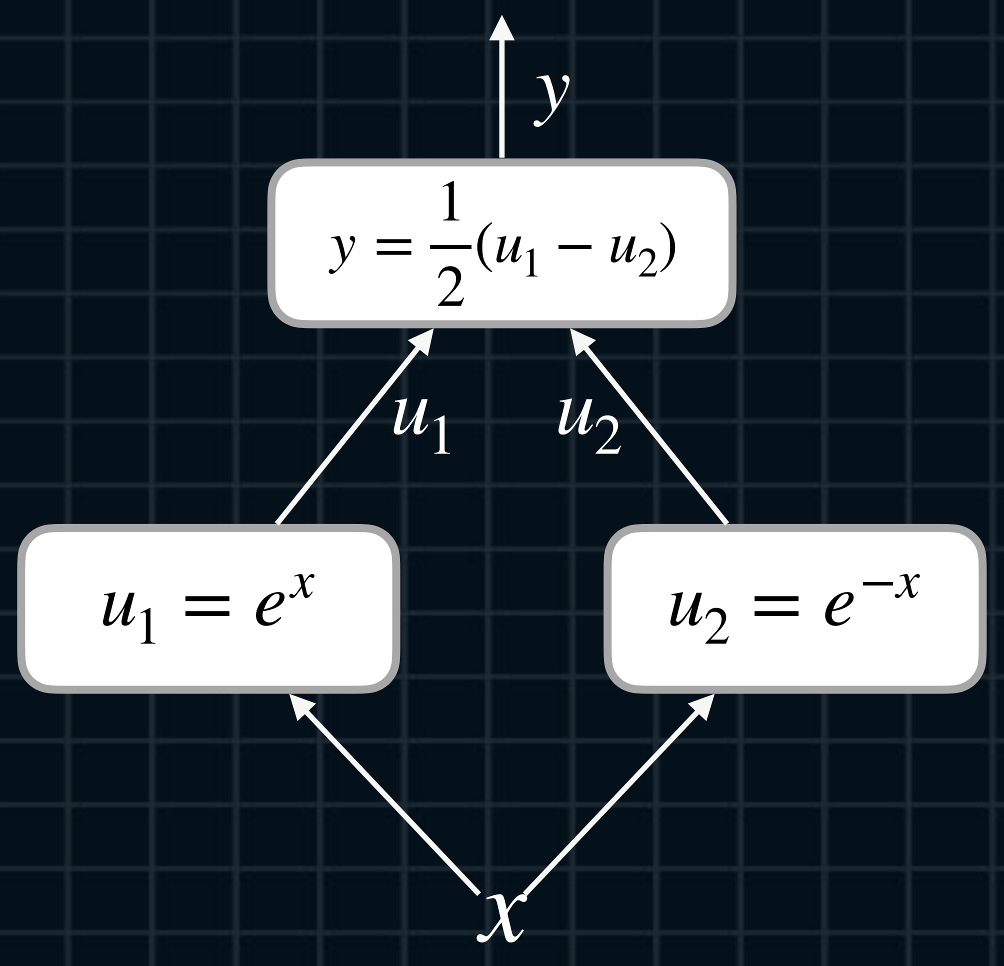


- Function with Multipath

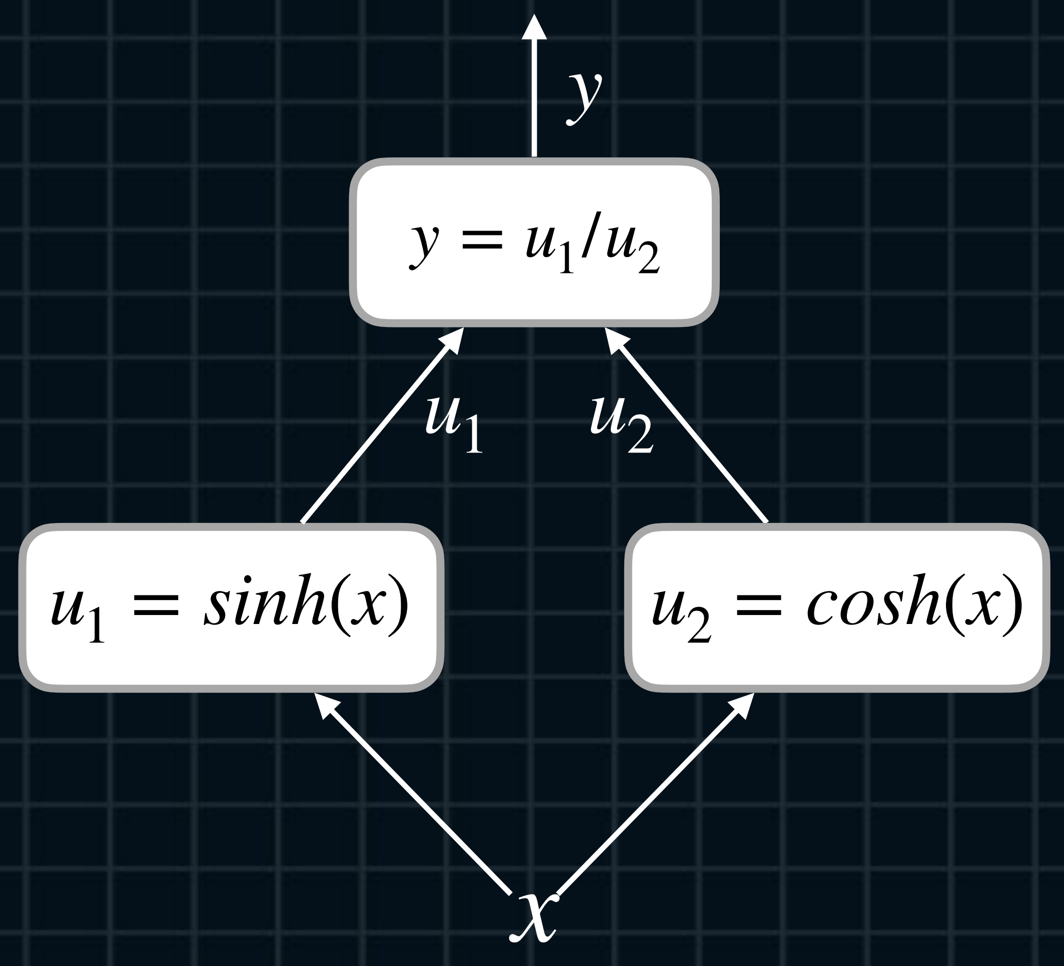
Multipath Examples



$y = cosh(x)$



$y = sinh(x)$

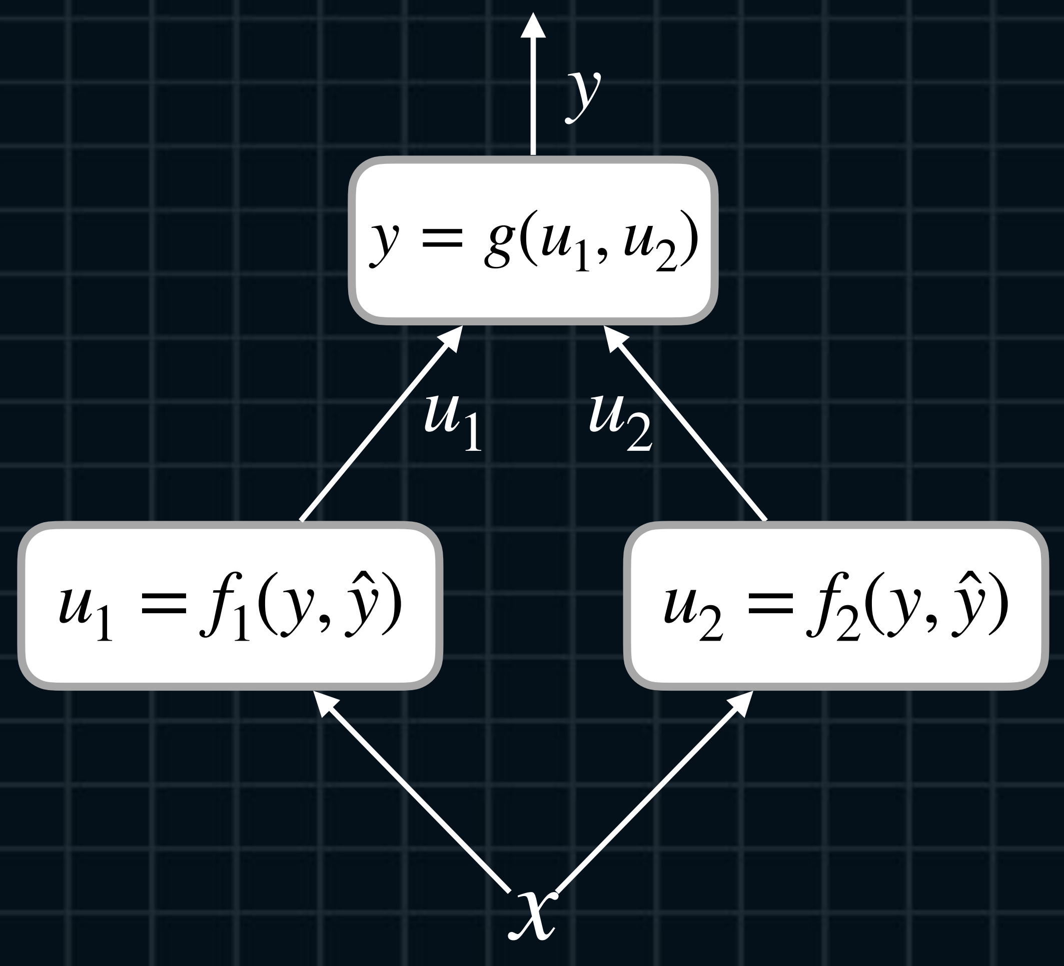


$y = tanh(x)$

Lecture.8 Total Derivatives

- Function with Multipath

Multipath Examples



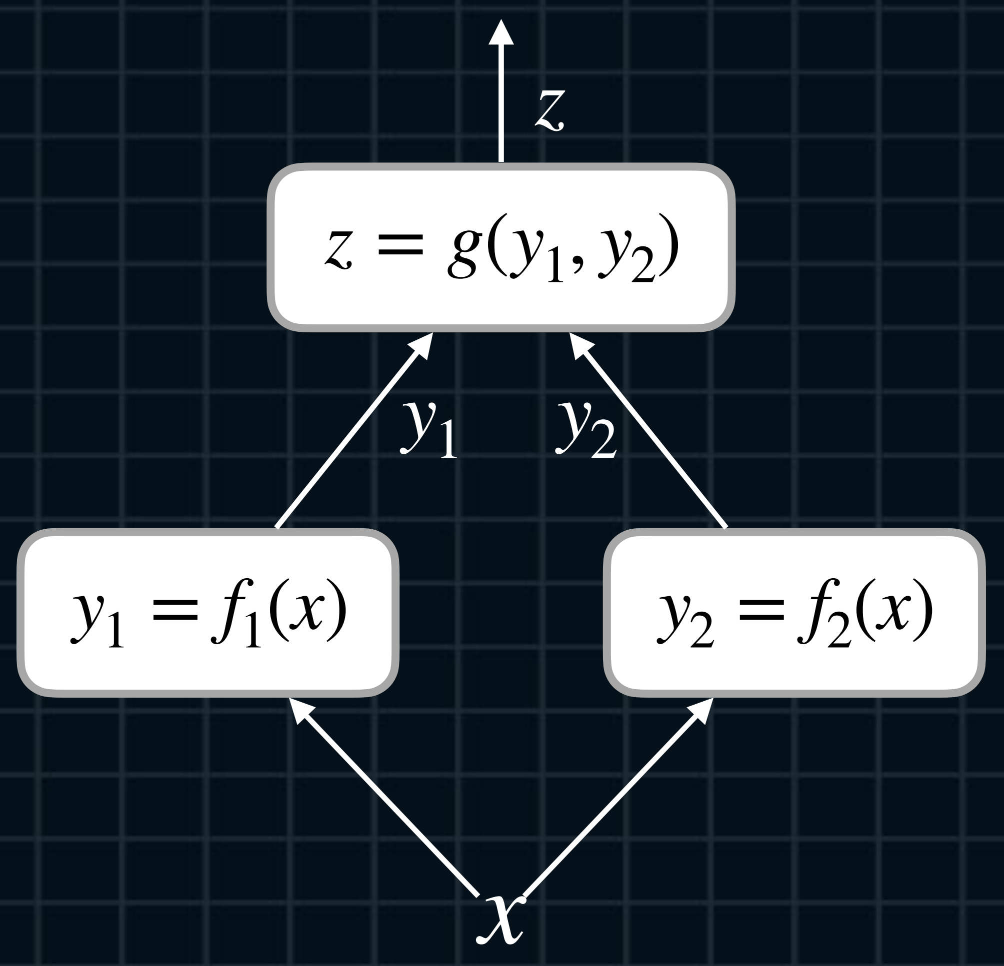
$$J_{BCE} = -\left[ylog(\hat{y}) + (1-y)log(1-\hat{y})\right]$$

$$y = g(u_1, u_2) = -(u_1 + u_2)$$

$$u_1 = f_1(y, \hat{y}) = ylog(\hat{y})$$

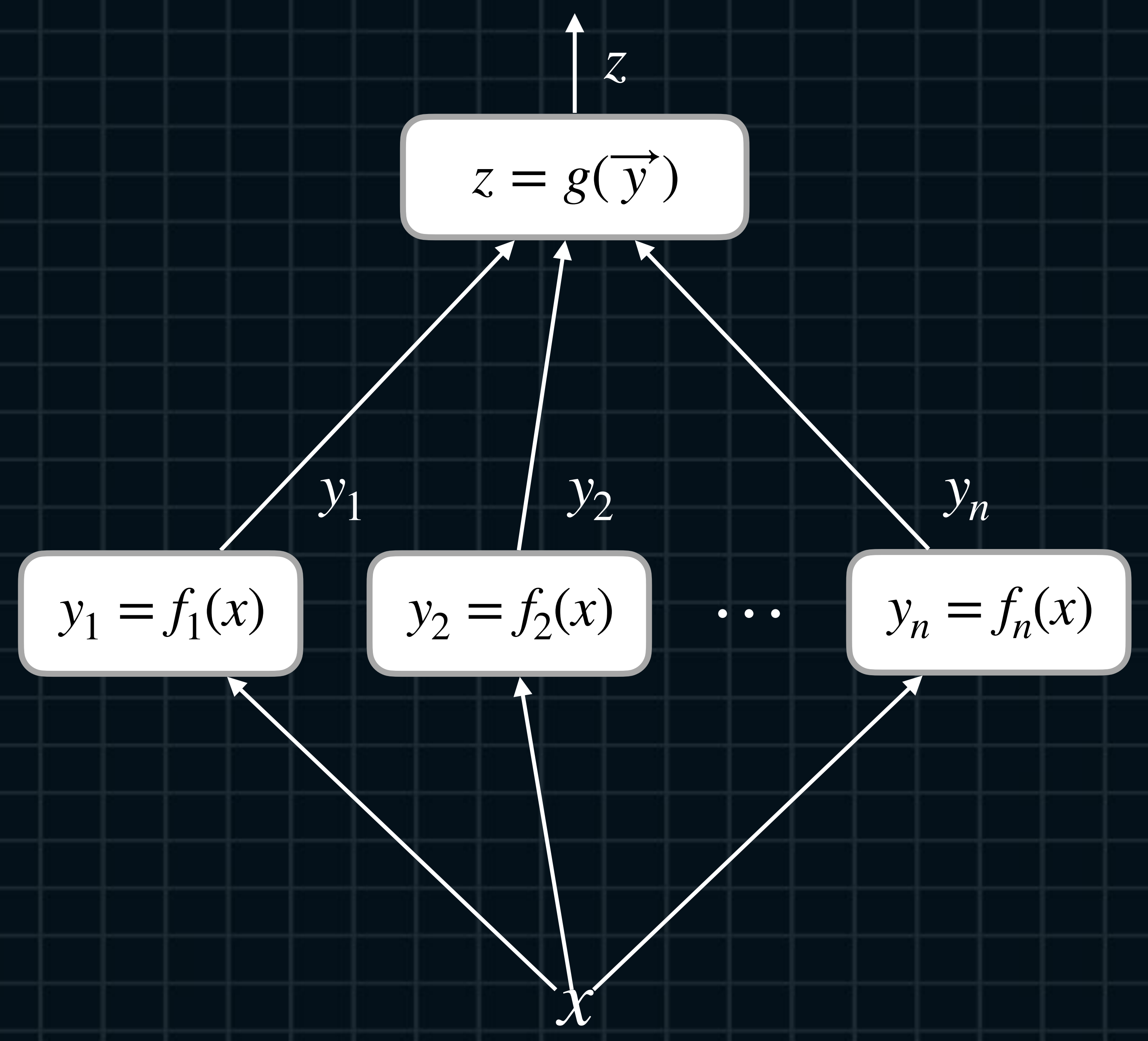
$$u_2 = f_2(y, \hat{y}) = (1-y)log(1-\hat{y})$$

Multipath Generalization



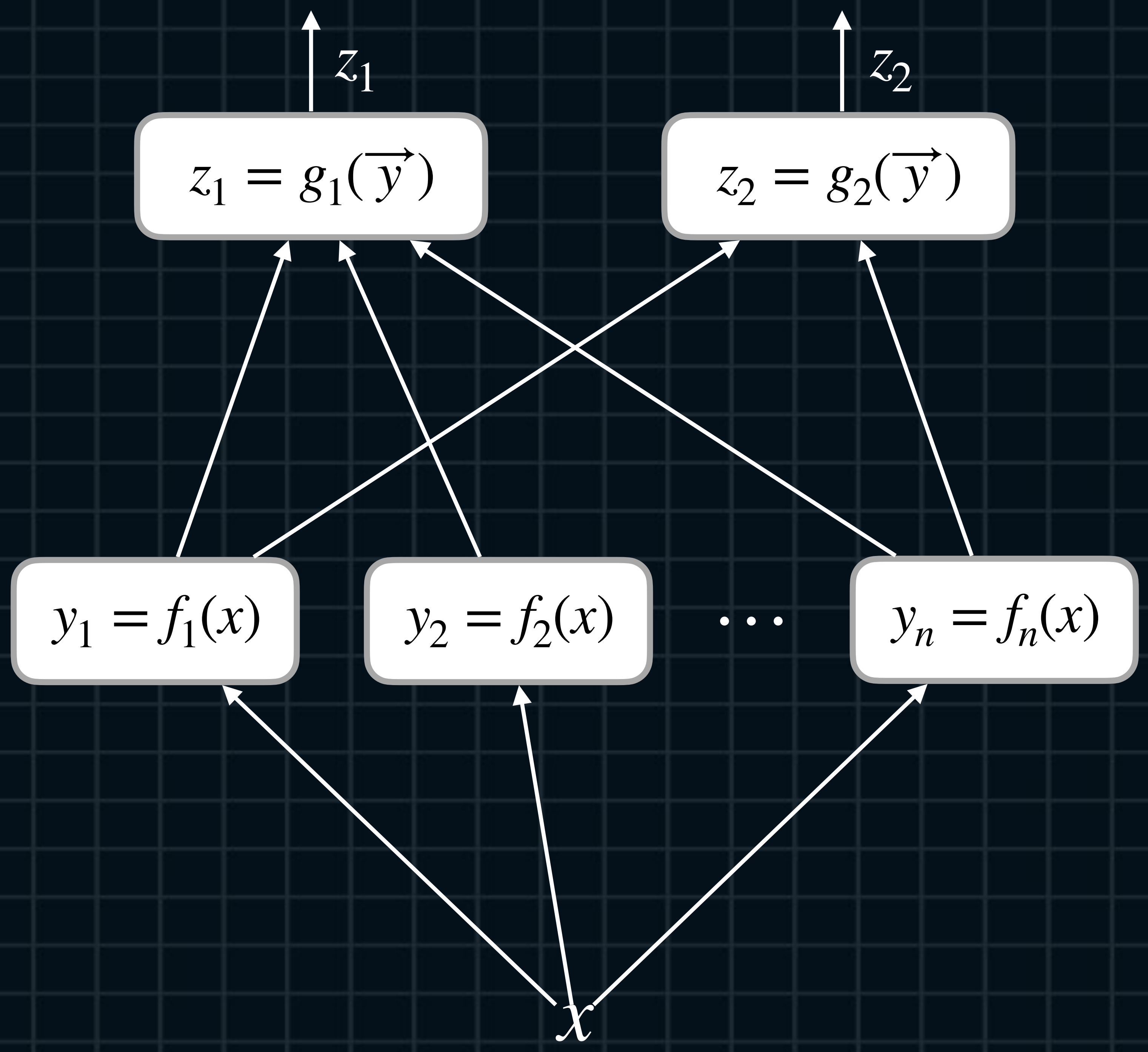
- Function with Multipath

Multipath Generalization



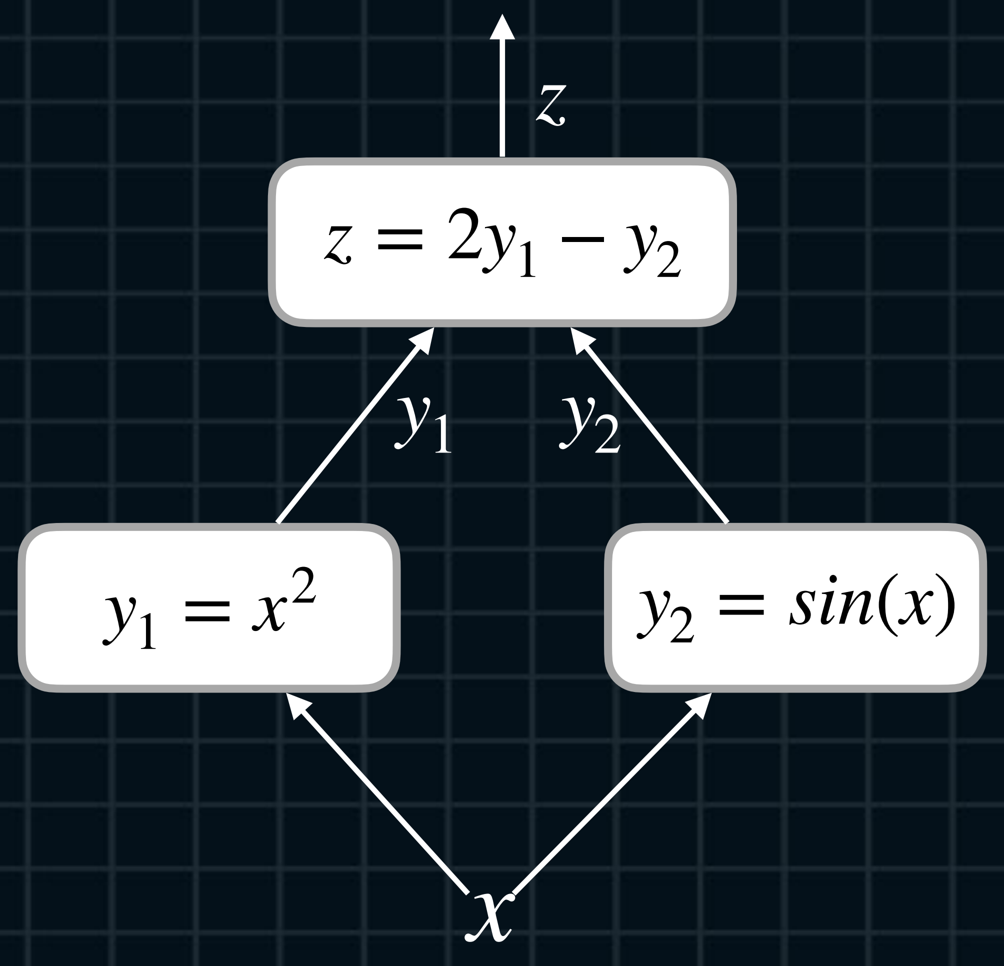
- Function with Multipath

Multipath Generalization

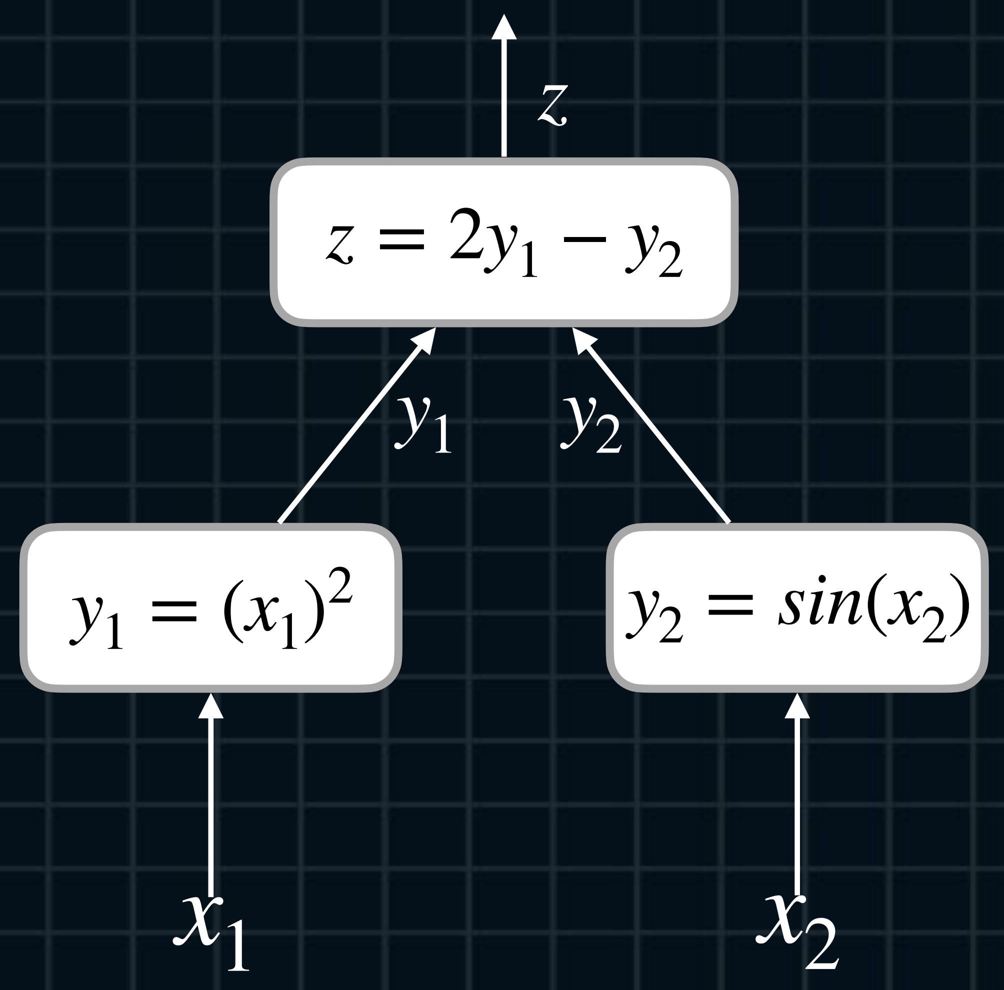


- Function with Multipath

Multipath vs Multivariate

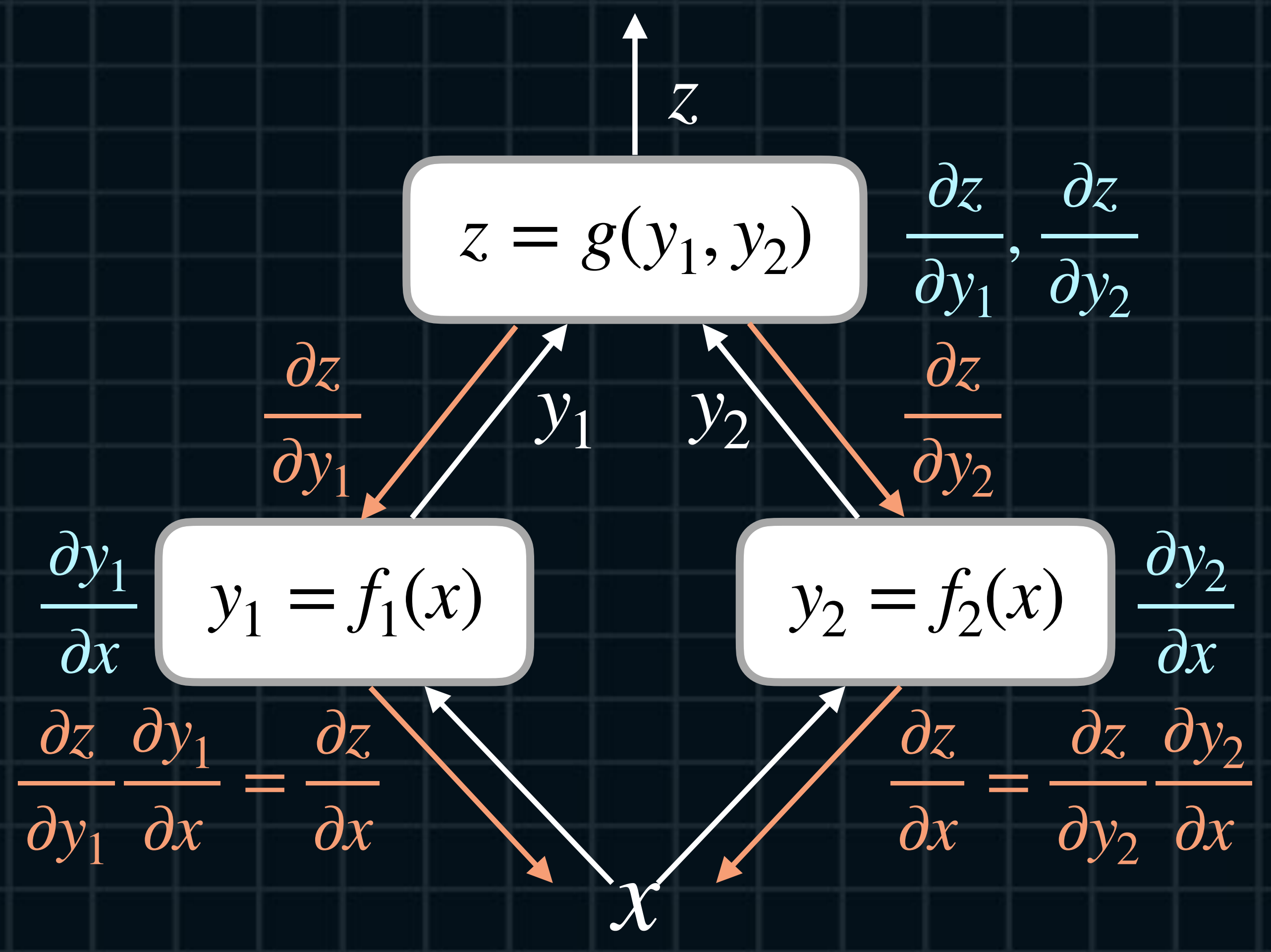


$$z = 2y_1 - y_2$$



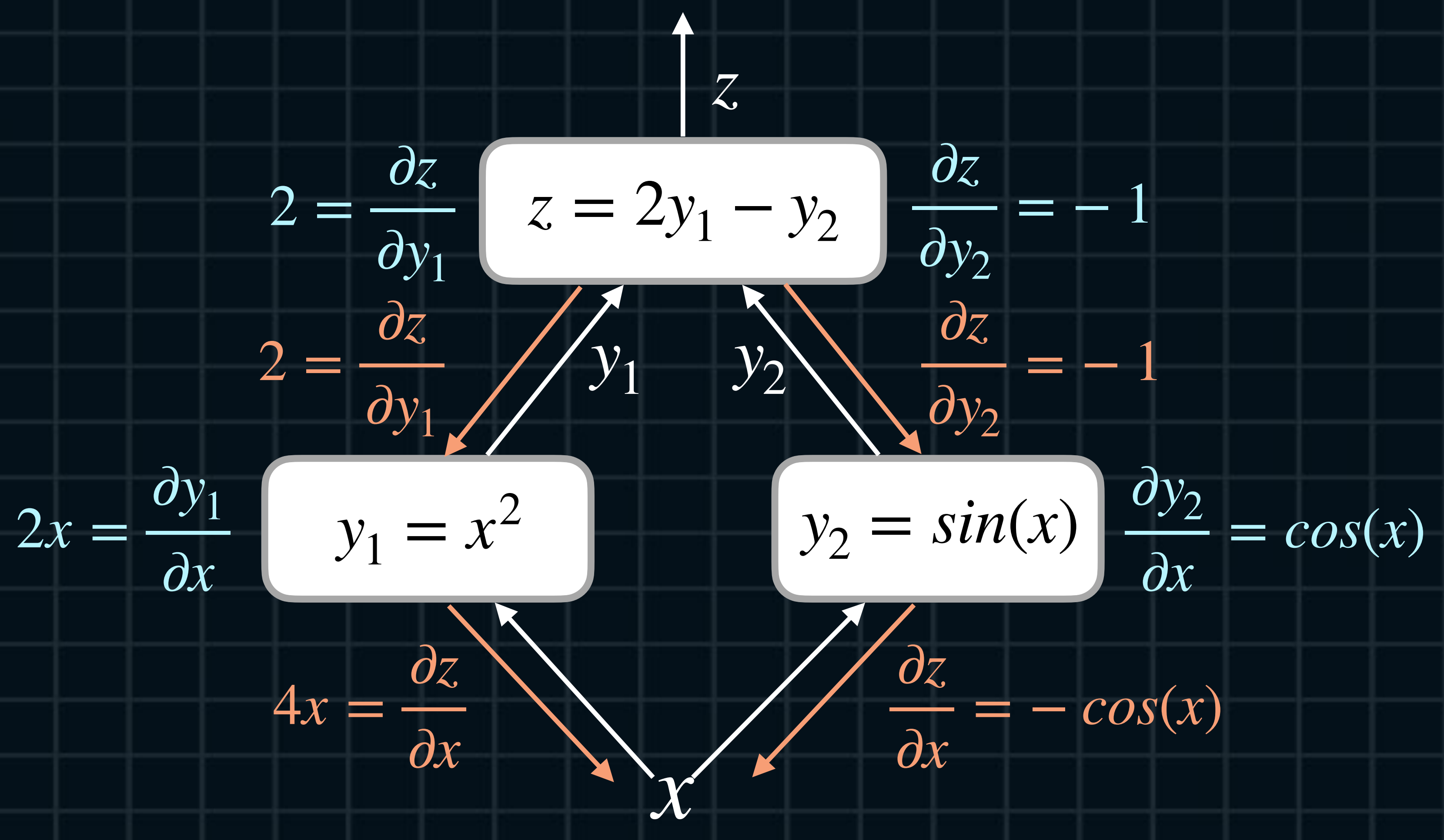
$$z = 2(x_1)^2 - sin(x_2)$$

Total Derivative



$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x}$$

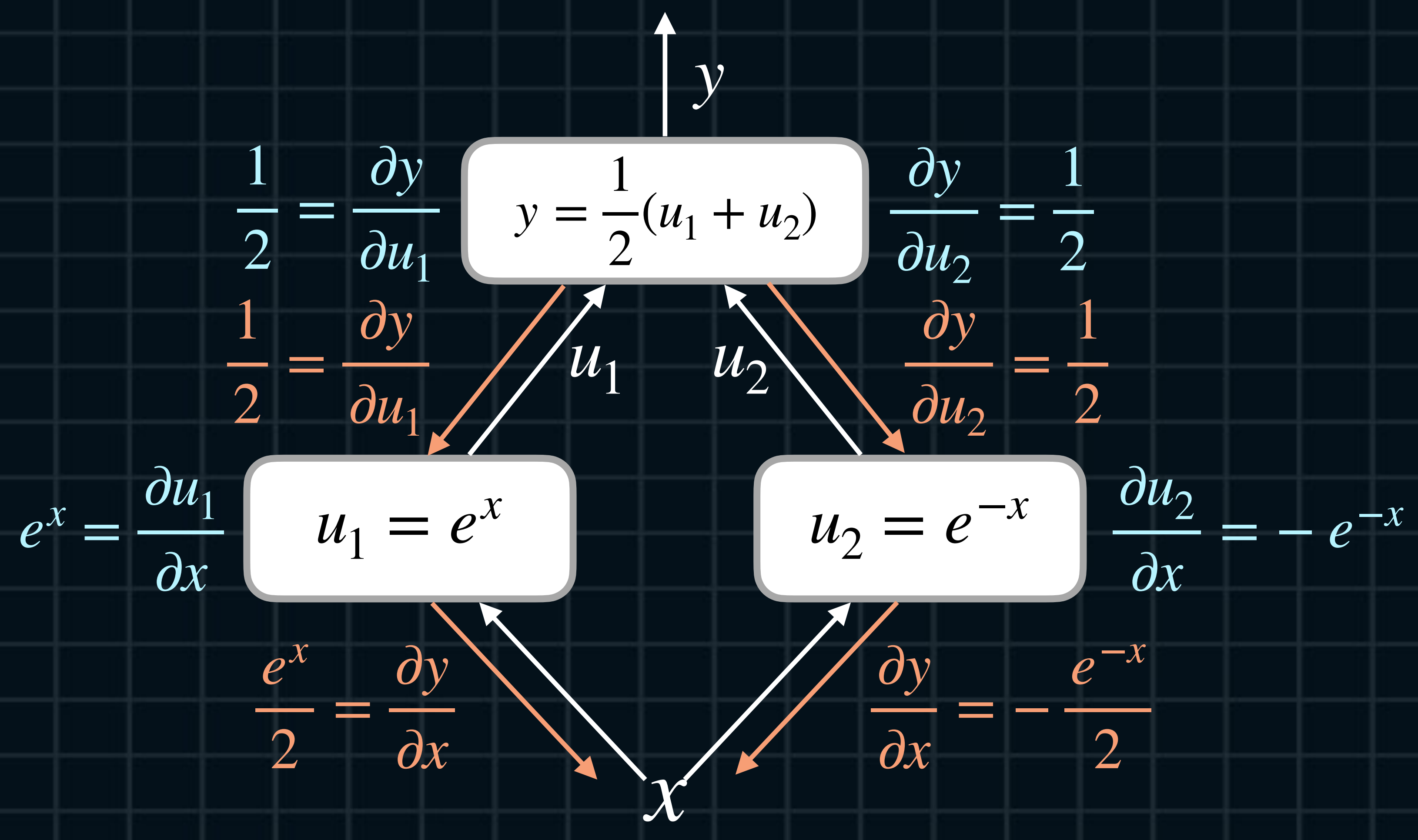
Total Derivative Examples



$$z = 2y_1 - y_2$$

$$\begin{aligned}\frac{\partial z}{\partial x} &= \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x} \\ &= 2 \cdot 2x + (-1) \cdot cos(x) \\ &= 4x - cos(x)\end{aligned}$$

Total Derivative Examples



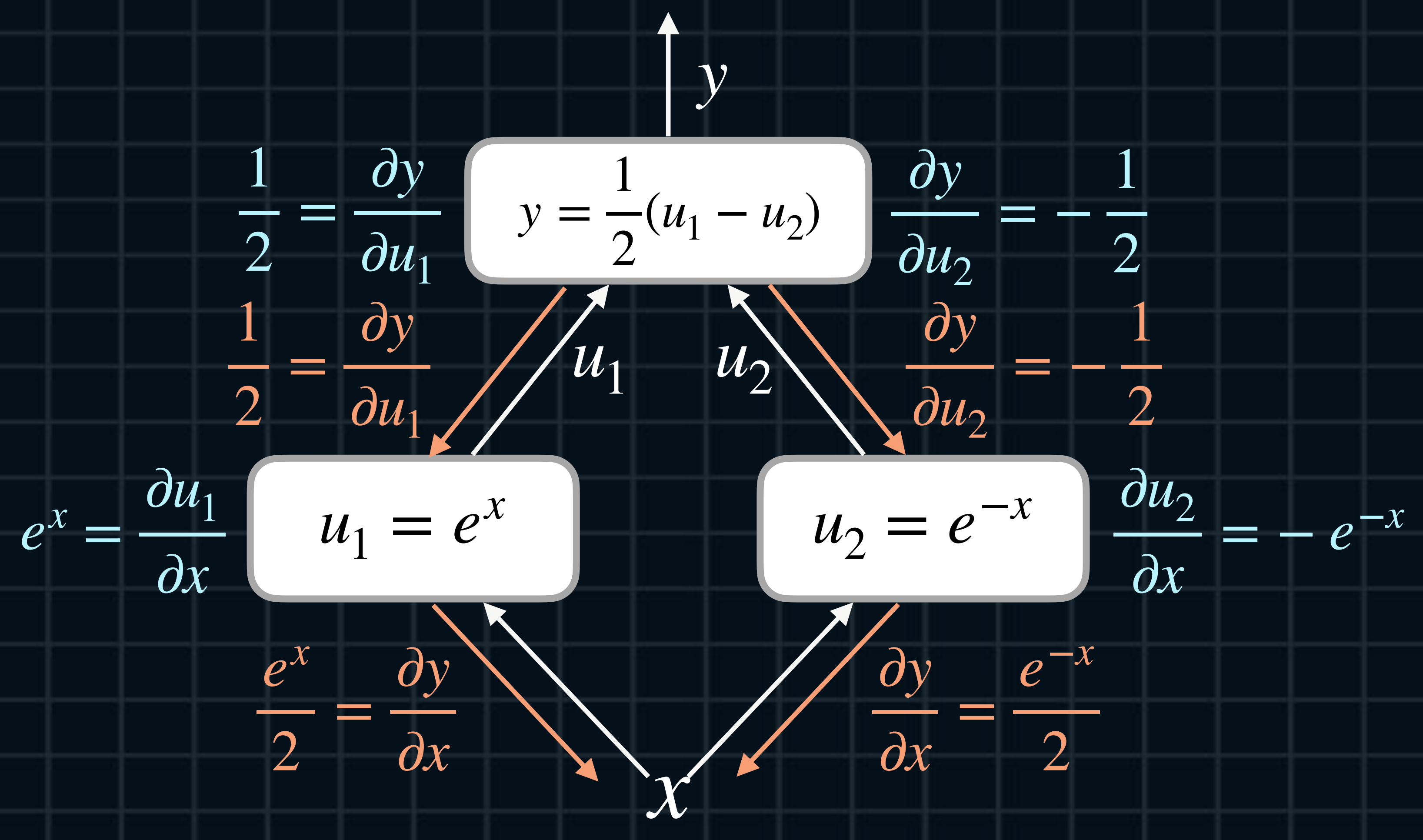
$y = cosh(x)$

$$y = cosh(x) = \frac{e^x + e^{-x}}{2}$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x}$$

$$= \frac{1}{2} \cdot e^x + \frac{1}{2} \cdot \left(-\frac{e^{-x}}{2}\right)$$

$$= \frac{e^x - e^{-x}}{2} = sinh(x)$$



$y = sinh(x)$

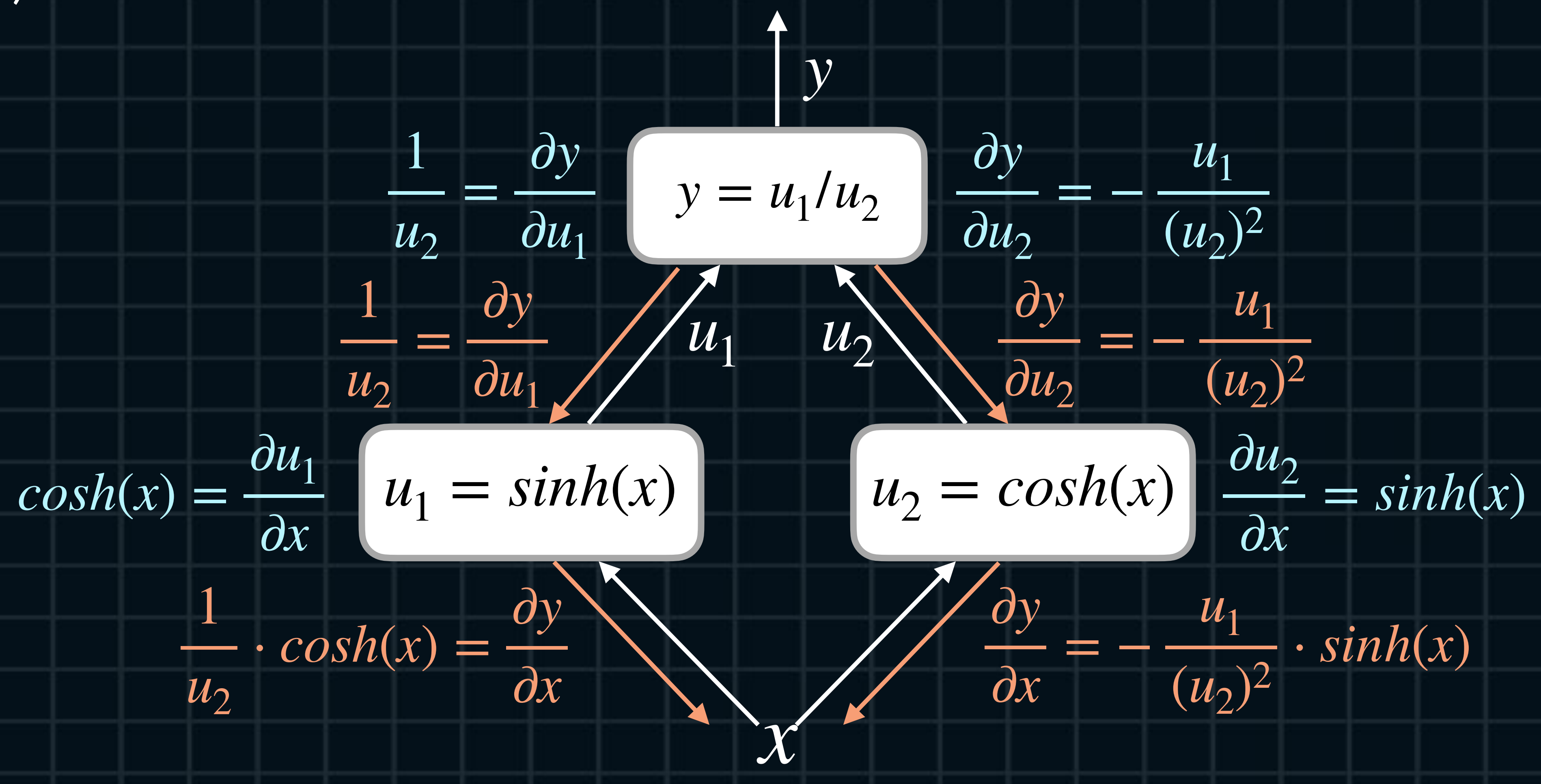
$$y = sinh(x) = \frac{e^x - e^{-x}}{2}$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u_1}\frac{\partial u_1}{\partial x} + \frac{\partial y}{\partial u_2}\frac{\partial u_2}{\partial x}$$

$$= \frac{1}{2} \cdot e^x + \left(-\frac{1}{2}\right) \cdot \left(-\frac{e^{-x}}{2}\right)$$

$$= \frac{e^x + e^{-x}}{2} = cosh(x)$$

Total Derivative Examples(Tanh)

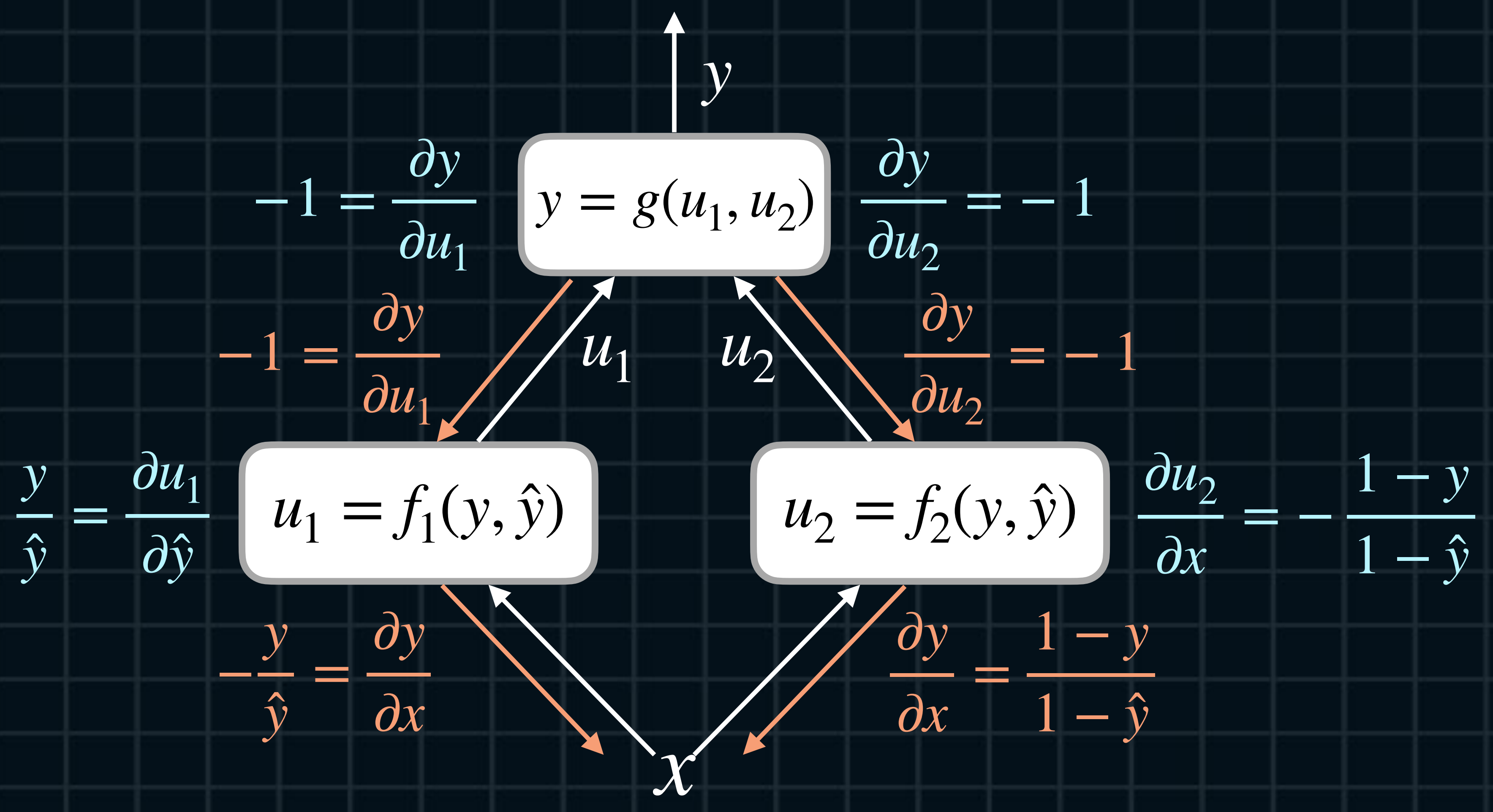


$$y = tanh(x) = \frac{sinh(x)}{cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\frac{\partial y}{\partial x} = \frac{1}{u_2} \cdot cosh(x) - \frac{u_1}{(u_2)^2} \cdot sinh(x)$$

$$= \big(1 + tanh(x)\big)\big(1 - tanh(x)\big)$$

Total Derivative Examples(BCEE)



$$y = g(u_1, u_2) = -(u_1 + u_2)$$

$$u_1 = f_1(y, \hat{y}) = ylog(\hat{y})$$

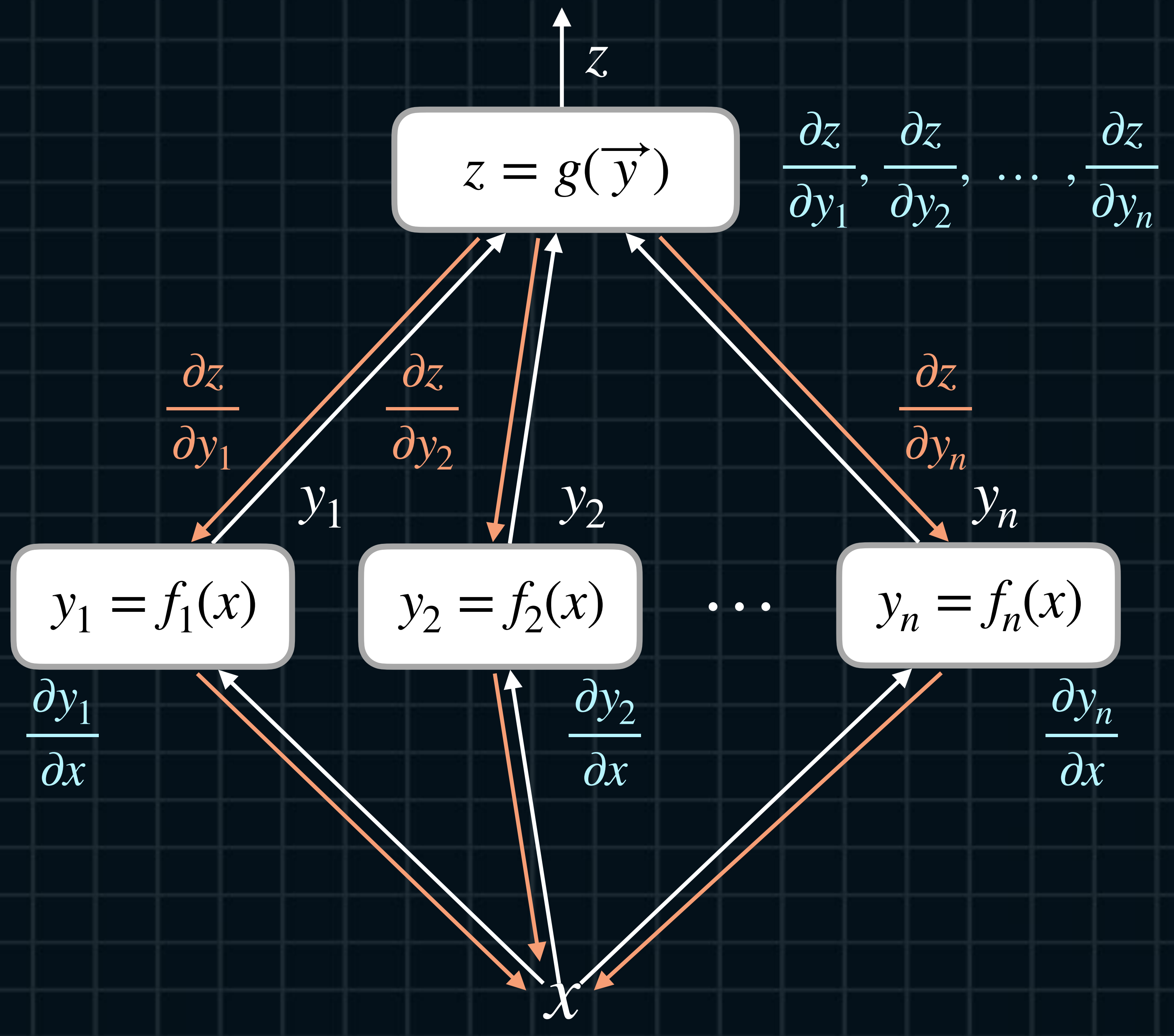
$$u_2 = f_2(y, \hat{y}) = (1-y)log(1-\hat{y})$$

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

$$\frac{\partial y}{\partial x} = \frac{1-y}{1-\hat{y}} - \frac{y}{\hat{y}}$$

- Total Derivatives

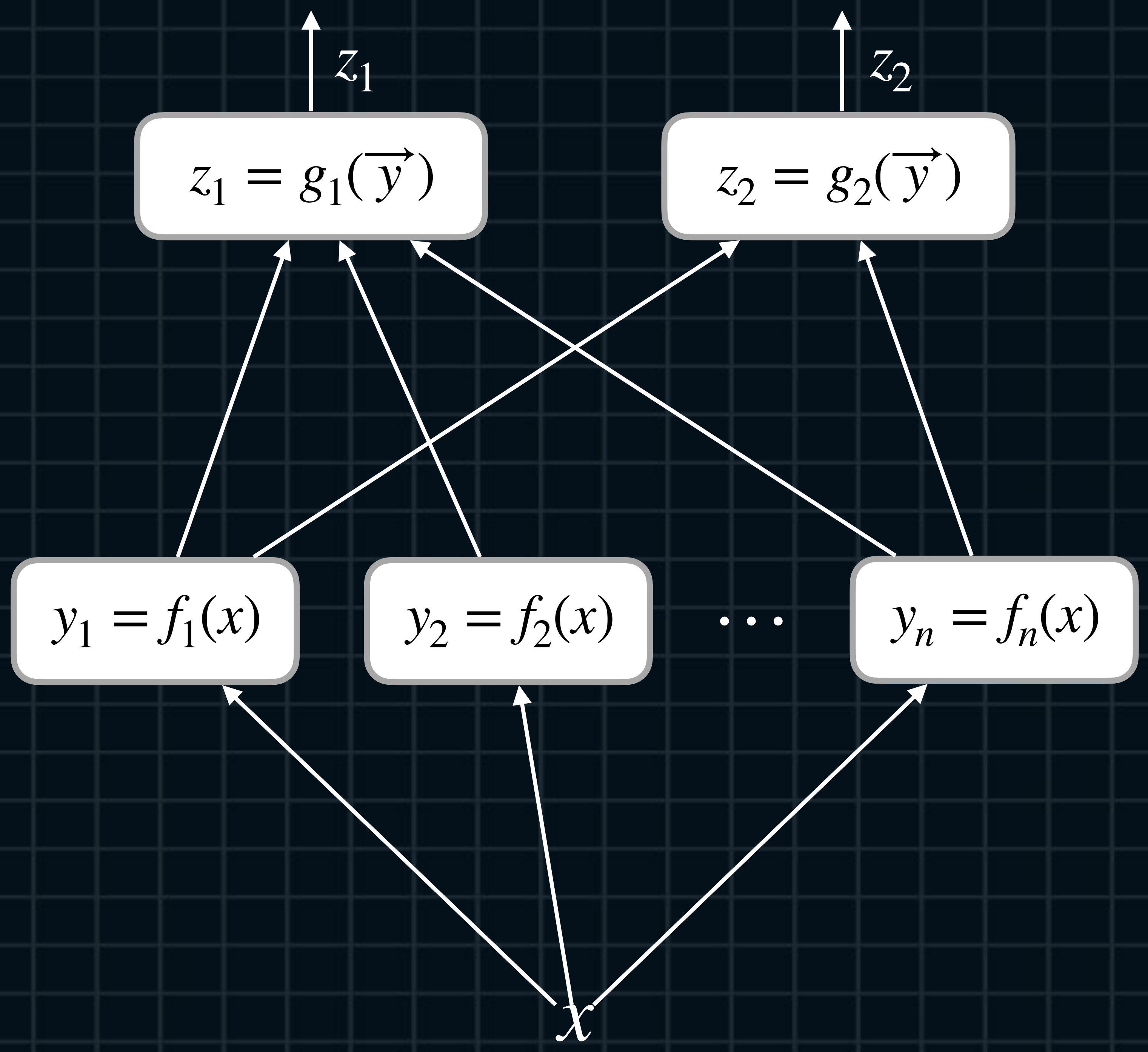
Total Derivative Generalization



$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x} + \dots + \frac{\partial z}{\partial y_n}\frac{\partial y_n}{\partial x}$$

$$= \sum_{i=1}^{n} \frac{\partial z}{\partial y_i}\frac{\partial y_i}{\partial x}$$

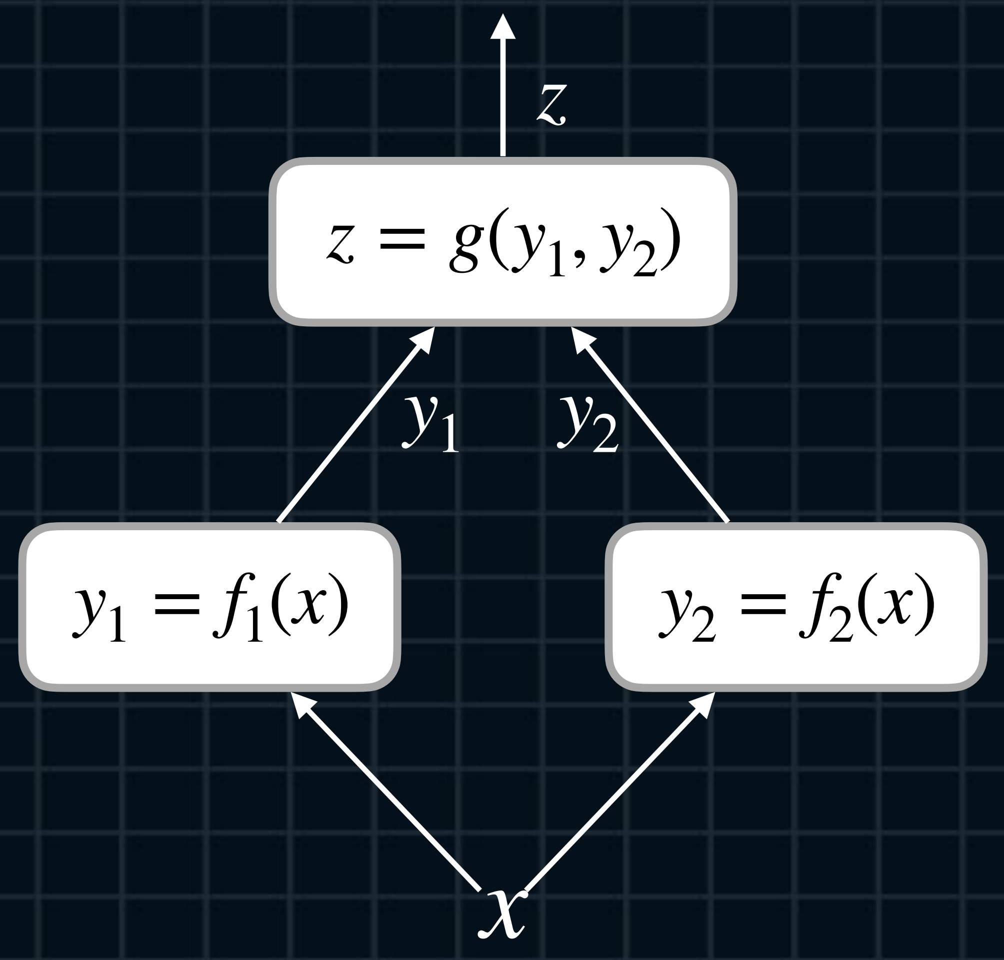
Total Derivative Generalization



$$\frac{\partial z_1}{\partial x} = \frac{\partial z_1}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z_1}{\partial y_2}\frac{\partial y_2}{\partial x} + \ldots + \frac{\partial z_1}{\partial y_n}\frac{\partial y_n}{\partial x} = \sum_{i=1}^{n} \frac{\partial z_1}{\partial y_i}\frac{\partial y_i}{\partial x}$$

$$\frac{\partial z_2}{\partial x} = \frac{\partial z_2}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z_2}{\partial y_2}\frac{\partial y_2}{\partial x} + \ldots + \frac{\partial z_2}{\partial y_n}\frac{\partial y_n}{\partial x} = \sum_{i=1}^{n} \frac{\partial z_2}{\partial y_i}\frac{\partial y_i}{\partial x}$$

- Vector Function and Total Derivative



Total Derivative

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x}$$

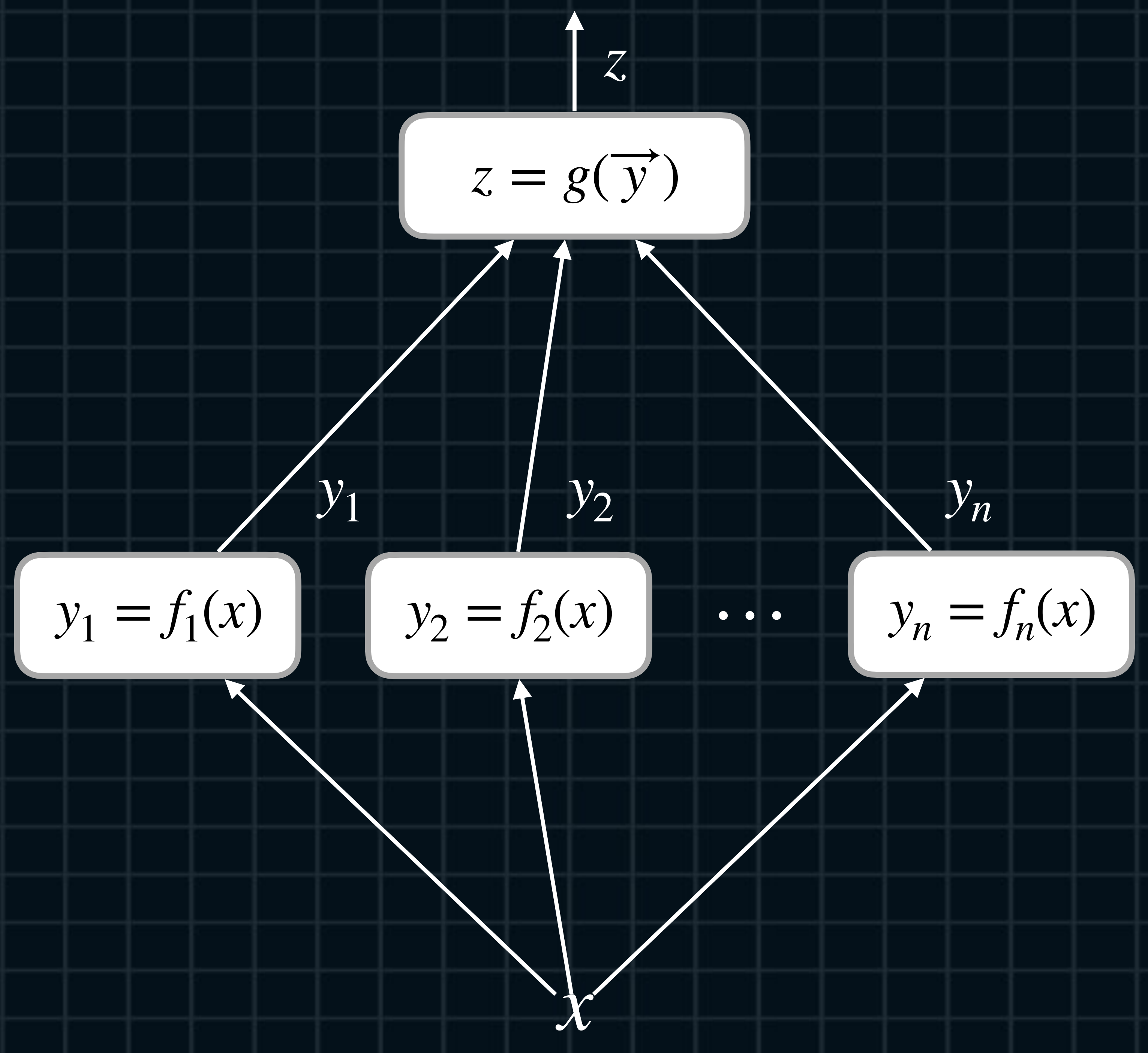
Vector Function

$$\frac{\partial z}{\partial \vec{y}} = \begin{pmatrix} \frac{\partial z}{\partial y_1} & \frac{\partial z}{\partial y_2} \end{pmatrix}, \quad \frac{\partial \vec{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \end{pmatrix}$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial \vec{y}}\frac{\partial \vec{y}}{\partial x} = \begin{pmatrix} \frac{\partial z}{\partial y_1} & \frac{\partial z}{\partial y_2} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \end{pmatrix} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x}$$

Lecture.8 Total Derivatives

- Vector Function and Total Derivative



Total Derivative

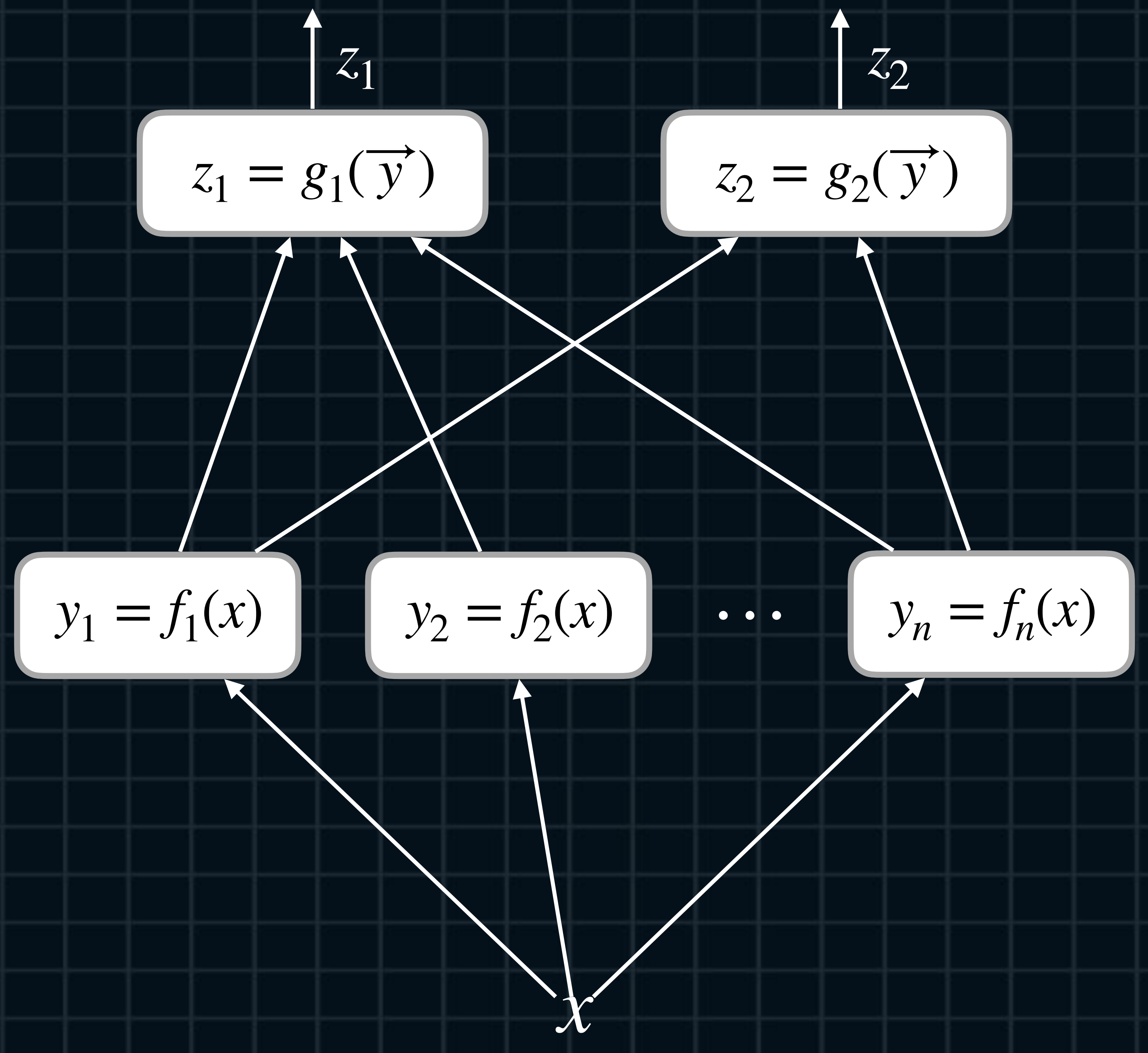
$$\frac{\partial z}{\partial x} = \sum_{i=1}^{n} \frac{\partial z}{\partial y_i}\frac{\partial y_i}{\partial x}$$

Vector Function

$$\frac{\partial z}{\partial \overrightarrow{y}} = \begin{pmatrix} \frac{\partial z}{\partial y_1} & \frac{\partial z}{\partial y_2} & \cdots & \frac{\partial z}{\partial y_n} \end{pmatrix}, \quad \frac{\partial \overrightarrow{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{pmatrix}$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial \overrightarrow{y}}\frac{\partial \overrightarrow{y}}{\partial x} = \begin{pmatrix} \frac{\partial z}{\partial y_1} & \frac{\partial z}{\partial y_2} & \cdots & \frac{\partial z}{\partial y_n} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial z}{\partial y_i}\frac{\partial y_i}{\partial x}$$

- Vector Function and Total Derivative



Total Derivative

$$\frac{\partial z_1}{\partial x} = \frac{\partial z_1}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z_1}{\partial y_2}\frac{\partial y_2}{\partial x} + \ldots + \frac{\partial z_1}{\partial y_n}\frac{\partial y_n}{\partial x} = \sum_{i=1}^{n} \frac{\partial z_1}{\partial y_i}\frac{\partial y_i}{\partial x}$$

$$\frac{\partial z_2}{\partial x} = \frac{\partial z_2}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z_2}{\partial y_2}\frac{\partial y_2}{\partial x} + \ldots + \frac{\partial z_2}{\partial y_n}\frac{\partial y_n}{\partial x} = \sum_{i=1}^{n} \frac{\partial z_2}{\partial y_i}\frac{\partial y_i}{\partial x}$$

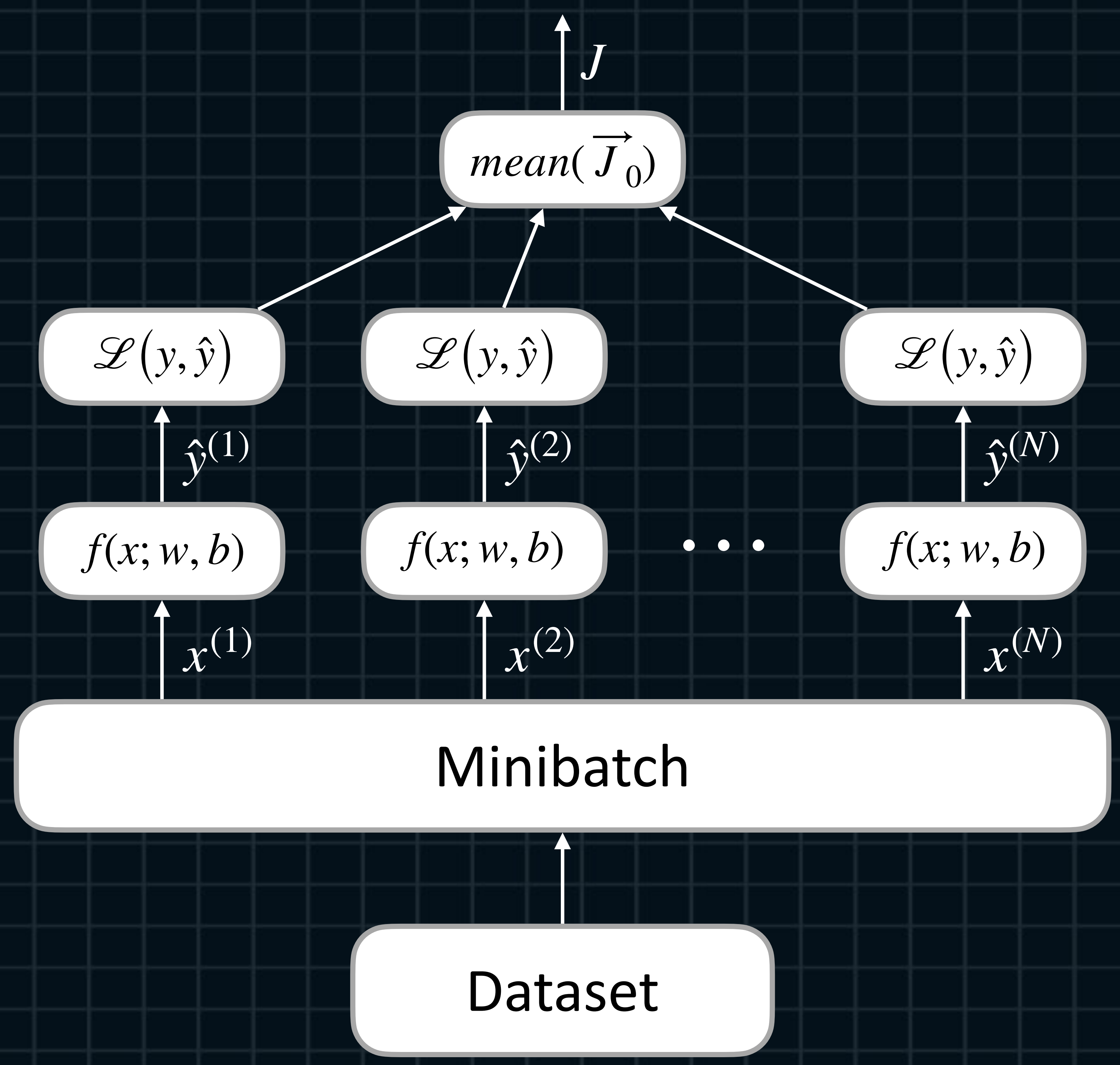
Vector Function

$$\frac{\partial \vec{z}}{\partial \overrightarrow{y}} = \begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_n} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_n} \end{pmatrix}, \quad \frac{\partial \overrightarrow{y}}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{pmatrix}$$

$$\frac{\partial \vec{z}}{\partial x} = \frac{\partial \vec{z}}{\partial \overrightarrow{y}}\frac{\partial \overrightarrow{y}}{\partial x} = \begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_n} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_n} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_n}{\partial x} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial z_1}{\partial y_i}\frac{\partial y_i}{\partial x} \\ \sum_{i=1}^{n} \frac{\partial z_2}{\partial y_i}\frac{\partial y_i}{\partial x} \end{pmatrix}$$

- Linear/Logistic Regression with Total Derivatives

Linear Regression



$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial J_0^{(1)}}\frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}}\frac{\partial \hat{y}^{(1)}}{\partial w} + \frac{\partial J}{\partial J_0^{(2)}}\frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}}\frac{\partial \hat{y}^{(2)}}{\partial w} + \dots \frac{\partial J}{\partial J_0^{(N)}}\frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}}\frac{\partial \hat{y}^{(N)}}{\partial w}$$

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

$$\frac{\partial J}{\partial J_0^{(i)}} = \frac{1}{N}, \quad \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = -2\left(y^{(i)} - \hat{y}^{(i)}\right), \quad \frac{\partial \hat{y}^{(i)}}{\partial w} = x^{(i)}, \quad \frac{\partial \hat{y}^{(i)}}{\partial b} = 1$$

$$\frac{\partial J}{\partial w} = \frac{1}{N}\cdot\left(-2\left(y^{(1)} - \hat{y}^{(1)}\right)\right)\cdot x^{(1)} + \frac{1}{N}\cdot\left(-2\left(y^{(2)} - \hat{y}^{(2)}\right)\right)\cdot x^{(2)} + \dots + \frac{1}{N}\cdot\left(-2\left(y^{(N)} - \hat{y}^{(N)}\right)\right)\cdot x^{(N)}$$

$$\frac{\partial J}{\partial b} = \frac{1}{N}\cdot\left(-2\left(y^{(1)} - \hat{y}^{(1)}\right)\right) + \frac{1}{N}\cdot\left(-2\left(y^{(2)} - \hat{y}^{(2)}\right)\right) + \dots + \frac{1}{N}\cdot\left(-2\left(y^{(N)} - \hat{y}^{(N)}\right)\right)$$

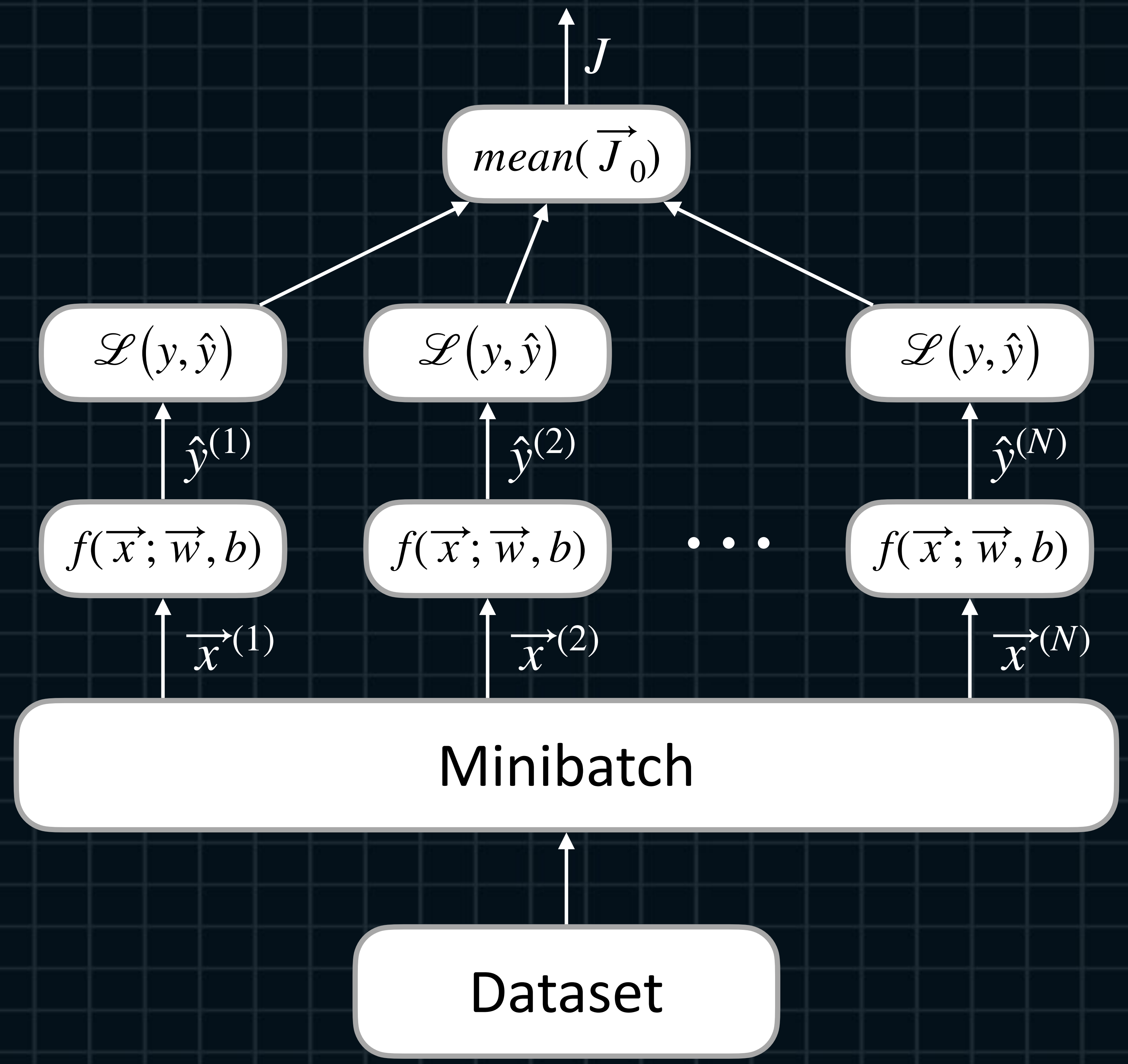
$$\frac{\partial J}{\partial w} = -\frac{2}{N}\sum_{i=1}^{N} x^{(i)}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$\frac{\partial J}{\partial b} = -\frac{2}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

Lecture.8 Total Derivatives

- Linear/Logistic Regression with Total Derivatives

Linear Regression



$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial J_0^{(1)}} \frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}} \frac{\partial \hat{y}^{(1)}}{\partial \vec{w}} + \frac{\partial J}{\partial J_0^{(2)}} \frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}} \frac{\partial \hat{y}^{(2)}}{\partial \vec{w}} + \dots \frac{\partial J}{\partial J_0^{(N)}} \frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}} \frac{\partial \hat{y}^{(N)}}{\partial \vec{w}}$$

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

$$\frac{\partial J}{\partial J_0^{(i)}} = \frac{1}{N}, \quad \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = -2\left(y^{(i)} - \hat{y}^{(i)}\right), \quad \frac{\partial \hat{y}^{(i)}}{\partial \vec{w}} = \left(\vec{x}^{(i)}\right)^T, \quad \frac{\partial \hat{y}^{(i)}}{\partial b} = 1$$

$$\frac{\partial J}{\partial w} = \frac{1}{N} \cdot \left(-2\left(y^{(1)} - \hat{y}^{(1)}\right)\right) \cdot \left(\vec{x}^{(1)}\right)^T + \frac{1}{N} \cdot \left(-2\left(y^{(2)} - \hat{y}^{(2)}\right)\right) \cdot \left(\vec{x}^{(2)}\right)^T + \dots + \frac{1}{N} \cdot \left(-2\left(y^{(N)} - \hat{y}^{(N)}\right)\right) \cdot \left(\vec{x}^{(N)}\right)^T$$

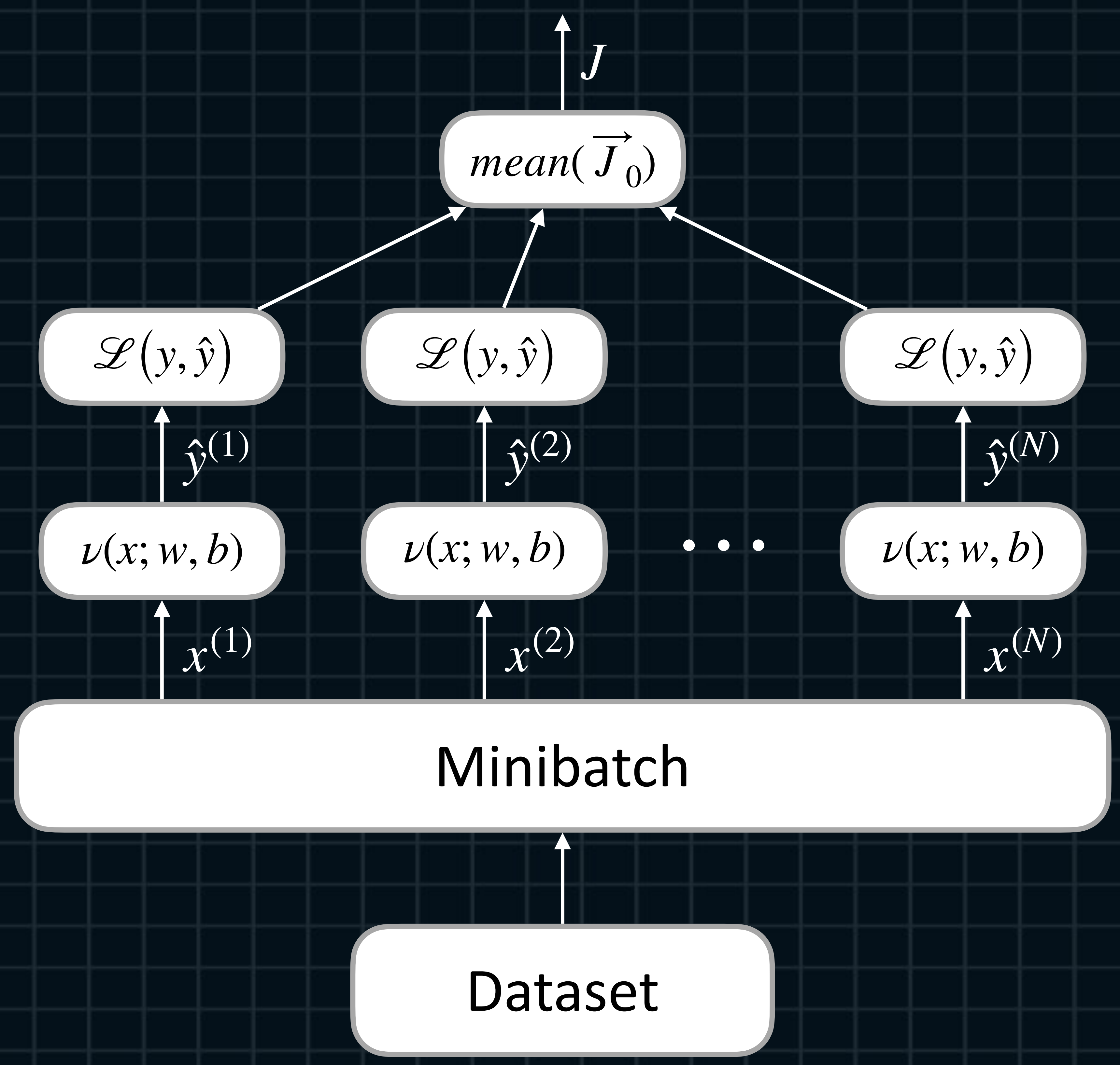
$$\frac{\partial J}{\partial b} = \frac{1}{N} \cdot \left(-2\left(y^{(1)} - \hat{y}^{(1)}\right)\right) + \frac{1}{N} \cdot \left(-2\left(y^{(2)} - \hat{y}^{(2)}\right)\right) + \dots + \frac{1}{N} \cdot \left(-2\left(y^{(N)} - \hat{y}^{(N)}\right)\right)$$

$$\frac{\partial J}{\partial w} = -\frac{2}{N} \sum_{i=1}^{N} \left(\vec{x}^{(i)}\right)^T \left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$\frac{\partial J}{\partial b} = -\frac{2}{N} \sum_{i=1}^{N} \left(y^{(i)} - \hat{y}^{(i)}\right)$$

- Linear/Logistic Regression with Total Derivatives

Logistic Regression



$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial J_0^{(1)}}\frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}}\frac{\partial \hat{y}^{(1)}}{\partial z^{(1)}}\frac{\partial z^{(1)}}{\partial w} + \frac{\partial J}{\partial J_0^{(2)}}\frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}}\frac{\partial \hat{y}^{(2)}}{\partial z^{(2)}}\frac{\partial z^{(2)}}{\partial w} + \dots \frac{\partial J}{\partial J_0^{(N)}}\frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}}\frac{\partial \hat{y}^{(N)}}{\partial z^{(N)}}\frac{\partial z^{(N)}}{\partial w}$$

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

$$\frac{\partial J}{\partial J_0^{(i)}} = \frac{1}{N}, \quad \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = \frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right)}, \quad \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} = \hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right), \quad \frac{\partial z^{(i)}}{\partial w} = x^{(i)}, \quad \frac{\partial z^{(i)}}{\partial b} = 1$$

$$\frac{\partial J}{\partial w} = \frac{1}{N}\cdot\left(-\left(y^{(1)} - \hat{y}^{(1)}\right)\right)\cdot x^{(1)} + \frac{1}{N}\cdot\left(-\left(y^{(2)} - \hat{y}^{(2)}\right)\right)\cdot x^{(2)} + \dots + \frac{1}{N}\cdot\left(-\left(y^{(N)} - \hat{y}^{(N)}\right)\right)\cdot x^{(N)}$$

$$\frac{\partial J}{\partial b} = \frac{1}{N}\cdot\left(-\left(y^{(1)} - \hat{y}^{(1)}\right)\right) + \frac{1}{N}\cdot\left(-\left(y^{(2)} - \hat{y}^{(2)}\right)\right) + \dots + \frac{1}{N}\cdot\left(-\left(y^{(N)} - \hat{y}^{(N)}\right)\right)$$

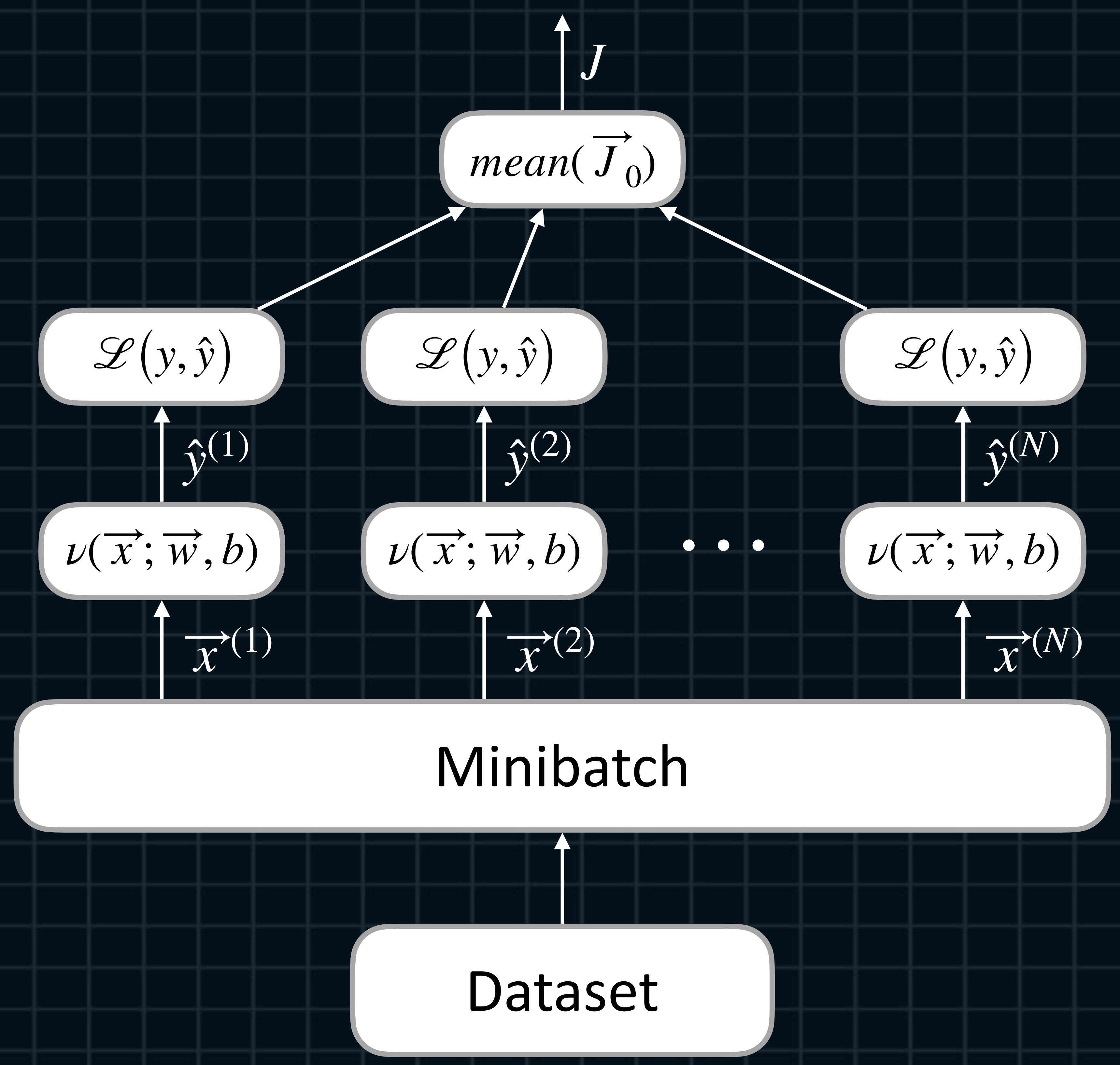
$$\frac{\partial J}{\partial w} = -\frac{1}{N}\sum_{i=1}^{N} x^{(i)}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$\frac{\partial J}{\partial b} = -\frac{1}{N}\sum_{i=1}^{N} \left(y^{(i)} - \hat{y}^{(i)}\right)$$

Lecture.8 Total Derivatives

- Linear/Logistic Regression with Total Derivatives

Logistic Regression



$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial J_0^{(1)}}\frac{\partial J_0^{(1)}}{\partial \hat{y}^{(1)}}\frac{\partial \hat{y}^{(1)}}{\partial z^{(1)}}\frac{\partial z^{(1)}}{\partial \vec{w}} + \frac{\partial J}{\partial J_0^{(2)}}\frac{\partial J_0^{(2)}}{\partial \hat{y}^{(2)}}\frac{\partial \hat{y}^{(2)}}{\partial z^{(2)}}\frac{\partial z^{(2)}}{\partial \vec{w}} + \dots \frac{\partial J}{\partial J_0^{(N)}}\frac{\partial J_0^{(N)}}{\partial \hat{y}^{(N)}}\frac{\partial \hat{y}^{(N)}}{\partial z^{(N)}}\frac{\partial z^{(N)}}{\partial \vec{w}}$$

$$= \sum_{i=1}^{N} \frac{\partial J}{\partial J_0^{(i)}}\frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}}\frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}}\frac{\partial z^{(i)}}{\partial \vec{w}}$$

$$\frac{\partial J}{\partial J_0^{(i)}} = \frac{1}{N}, \quad \frac{\partial J_0^{(i)}}{\partial \hat{y}^{(i)}} = \frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right)}, \quad \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} = \hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right), \quad \frac{\partial z^{(i)}}{\partial \vec{w}} = \left(\vec{x}^{(i)}\right)^T, \quad \frac{\partial z^{(i)}}{\partial b} = 1$$

$$\frac{\partial J}{\partial \vec{w}} = \frac{1}{N} \cdot \left(-\left(y^{(1)} - \hat{y}^{(1)}\right)\right) \cdot \left(\vec{x}^{(1)}\right)^T + \frac{1}{N} \cdot \left(-\left(y^{(2)} - \hat{y}^{(2)}\right)\right) \cdot \left(\vec{x}^{(2)}\right)^T + \dots + \frac{1}{N} \cdot \left(-\left(y^{(N)} - \hat{y}^{(N)}\right)\right) \cdot \left(\vec{x}^{(N)}\right)^T$$

$$\frac{\partial J}{\partial b} = \frac{1}{N} \cdot \left(-\left(y^{(1)} - \hat{y}^{(1)}\right)\right) + \frac{1}{N} \cdot \left(-\left(y^{(2)} - \hat{y}^{(2)}\right)\right) + \dots + \frac{1}{N} \cdot \left(-\left(y^{(N)} - \hat{y}^{(N)}\right)\right)$$

$$\frac{\partial J}{\partial w} = -\frac{1}{N}\sum_{i=1}^{N}\left(\vec{x}^{(i)}\right)^T\left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$\frac{\partial J}{\partial b} = -\frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

Backpropagation and Jacobian Matrices

Lecture.8

Total Derivatives