Fully-Connected Layer

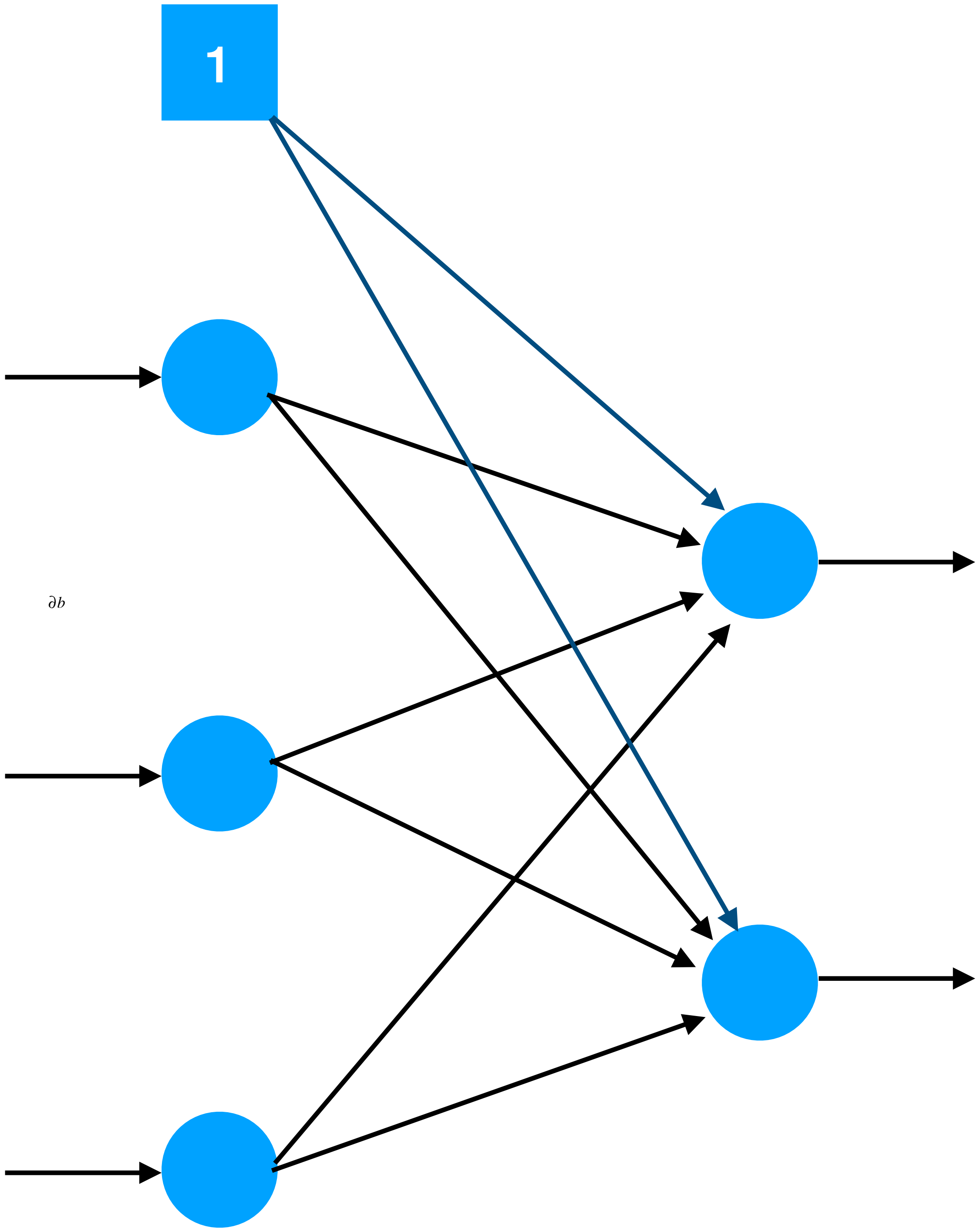
Dr. Thanh-Sach LE
LTSACH@hcmut.edu.vn

GVLab:
Graphics and Vision Laboratory

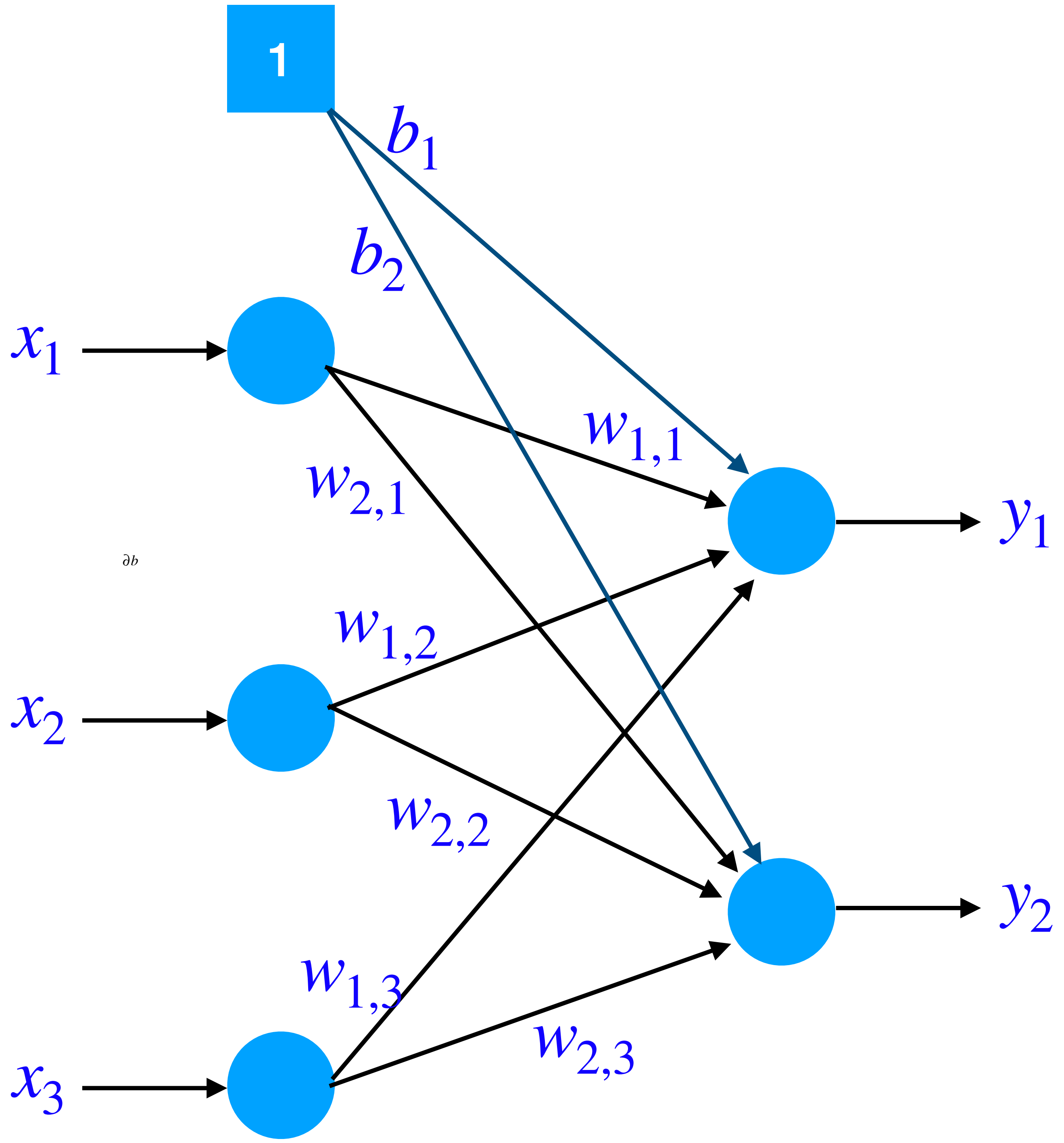
Faculty of Computer Science and Engineering,
HCMUT

Contents

3 Computation model

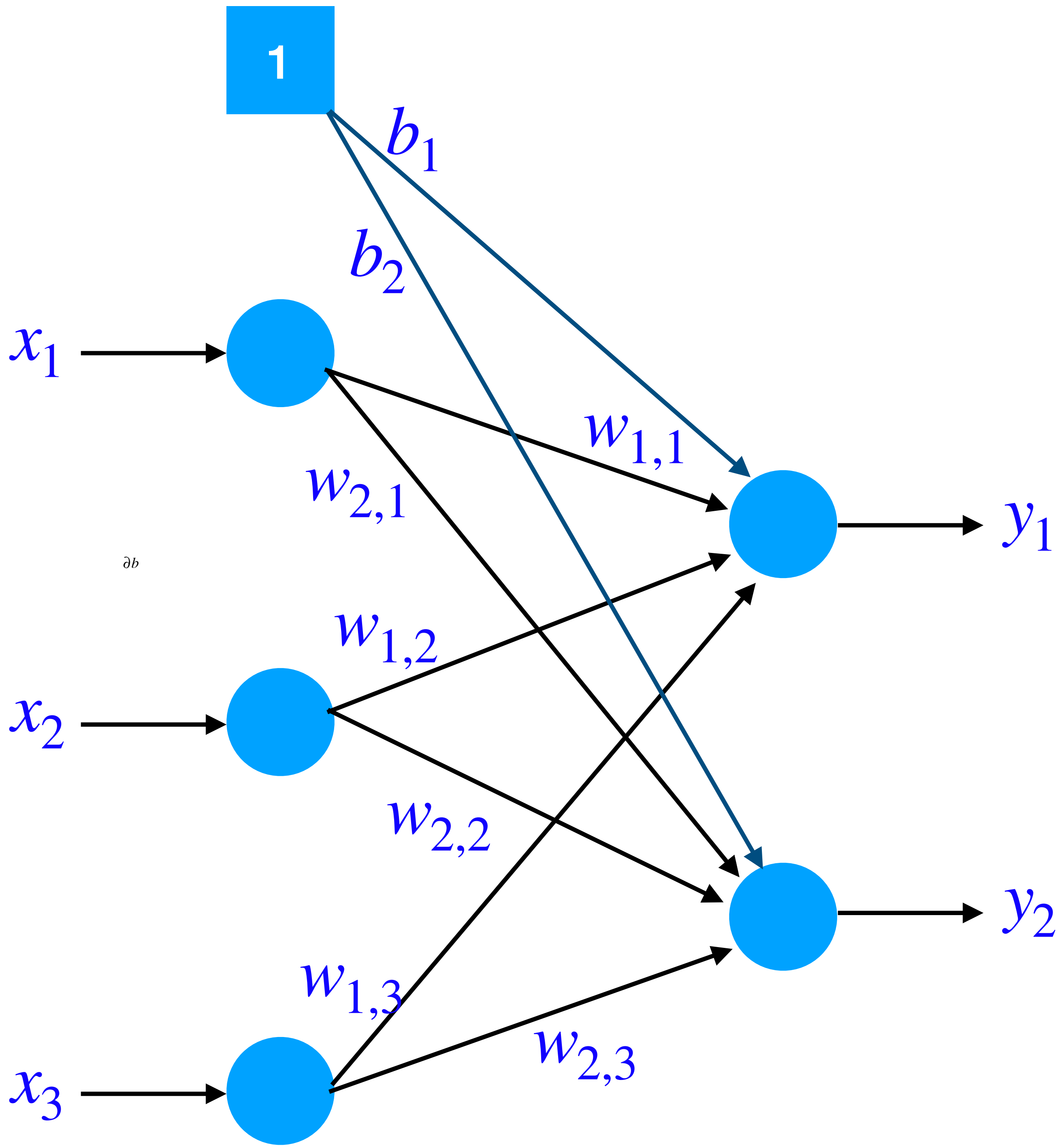


Computation model



5

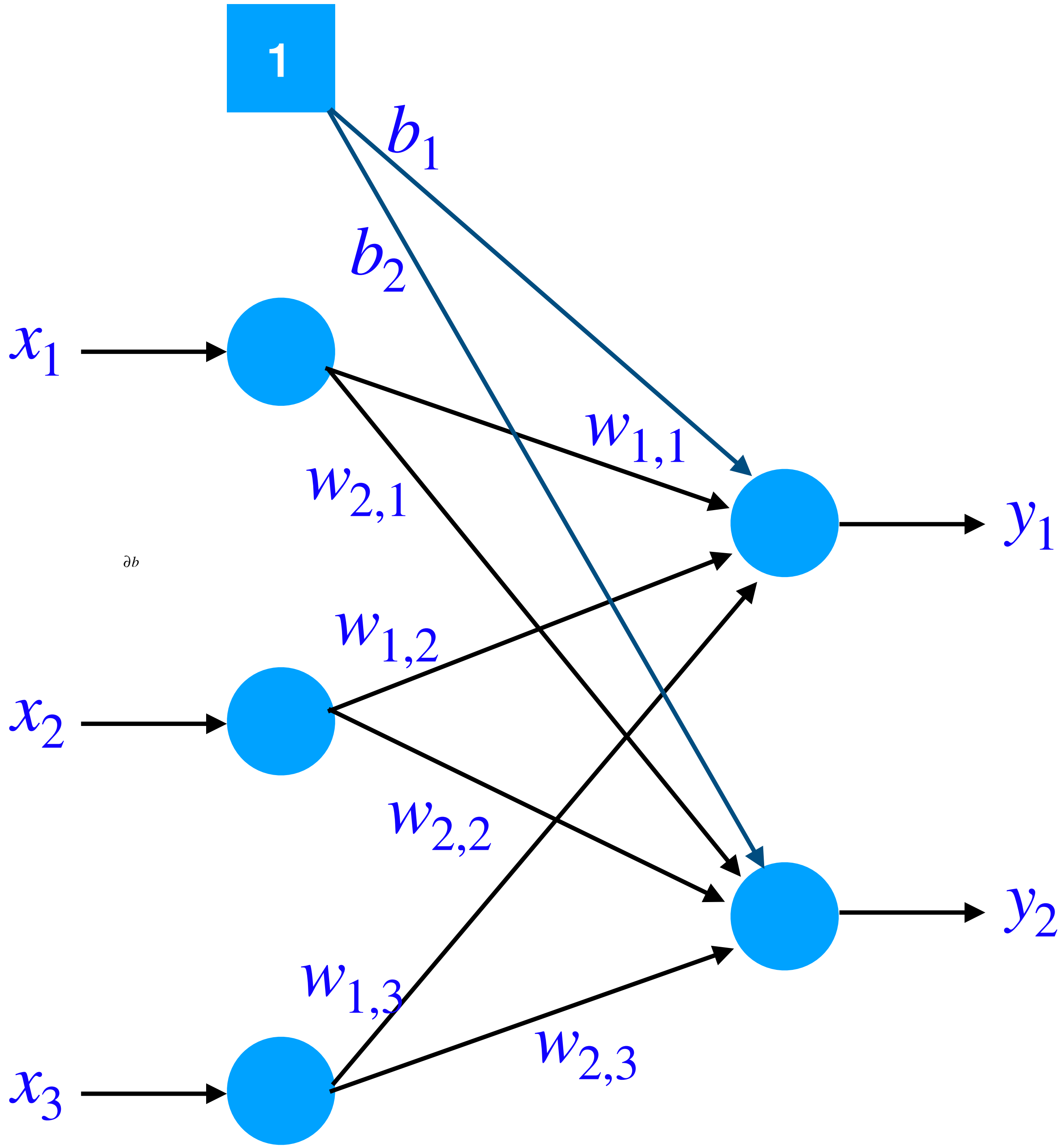
Computation model



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Computation model



Notation:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Computation model

$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

$$\mathbf{y} = \mathbf{W} * \mathbf{x} + \mathbf{b}$$

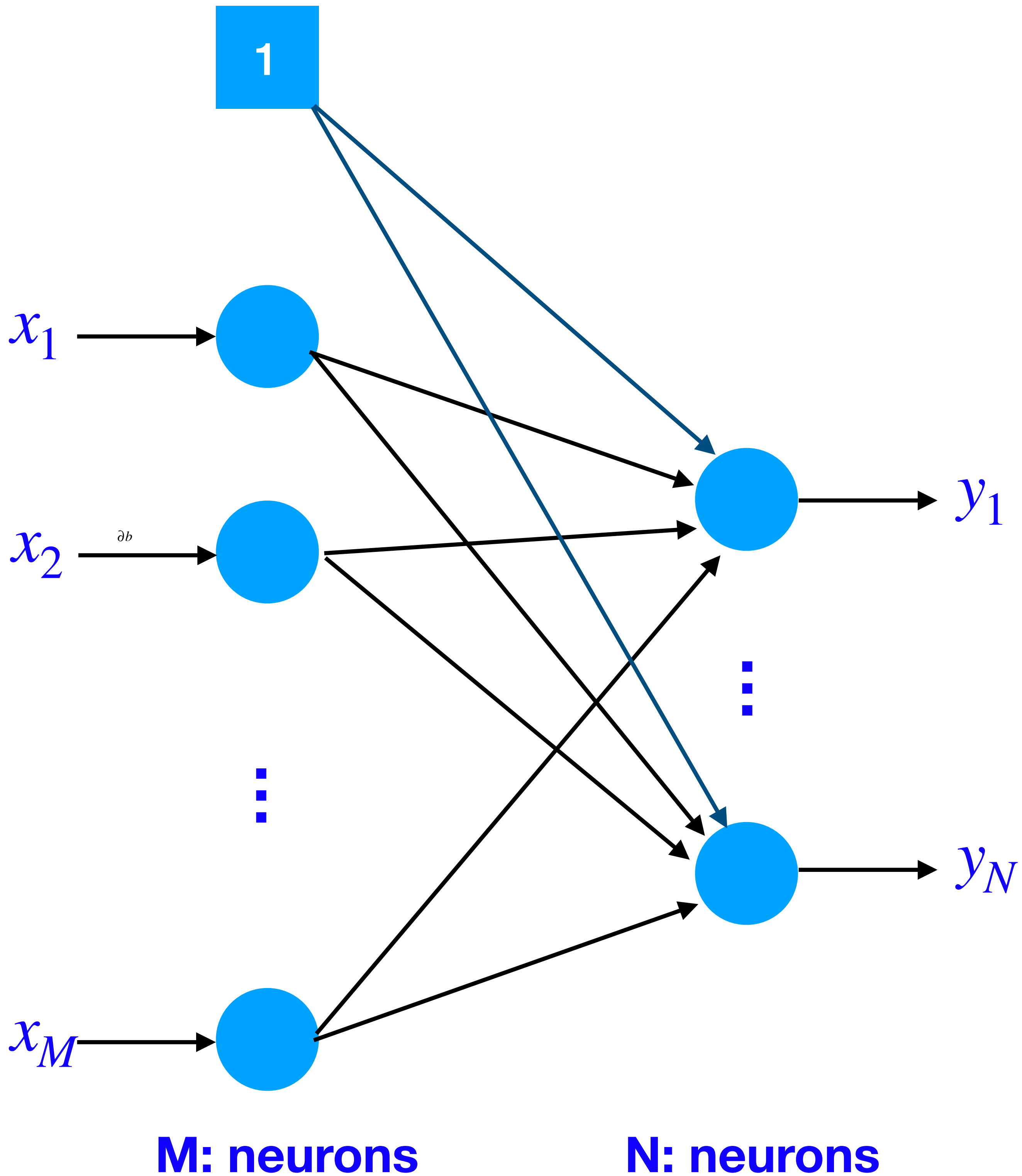
*: Matrix-vector multiplication

∂b

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix} \begin{bmatrix} . \\ . \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Learnable Parameters



$\mathbf{W} =$

$w_{1,1}$	$w_{1,2}$	$w_{1,3}$	...	$w_{1,M}$
$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	...	$w_{2,M}$
$w_{N,1}$	$w_{2,2}$	$w_{2,3}$	...	$w_{N,M}$

$\mathbf{b} =$

b_1
b_2
N

#parameters $= M \times N + N$

Back-propagation through FC-Layer

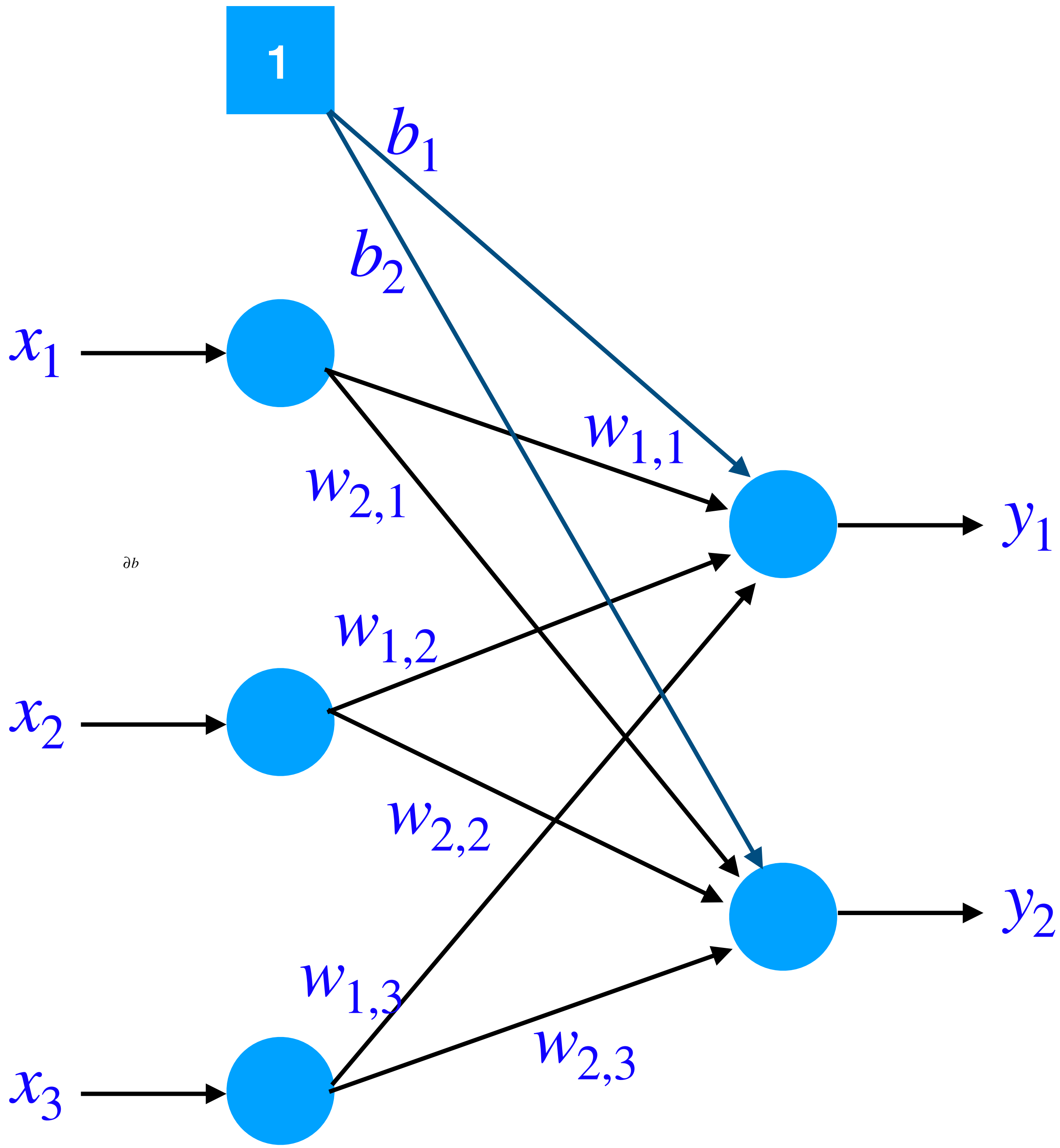
Dr. Thanh-Sach LE
LTSACH@hcmut.edu.vn

GVLab:
Graphics and Vision Laboratory

Faculty of Computer Science and Engineering,
HCMUT

Back-propagation through FC-Layer

Derivation of ΔW



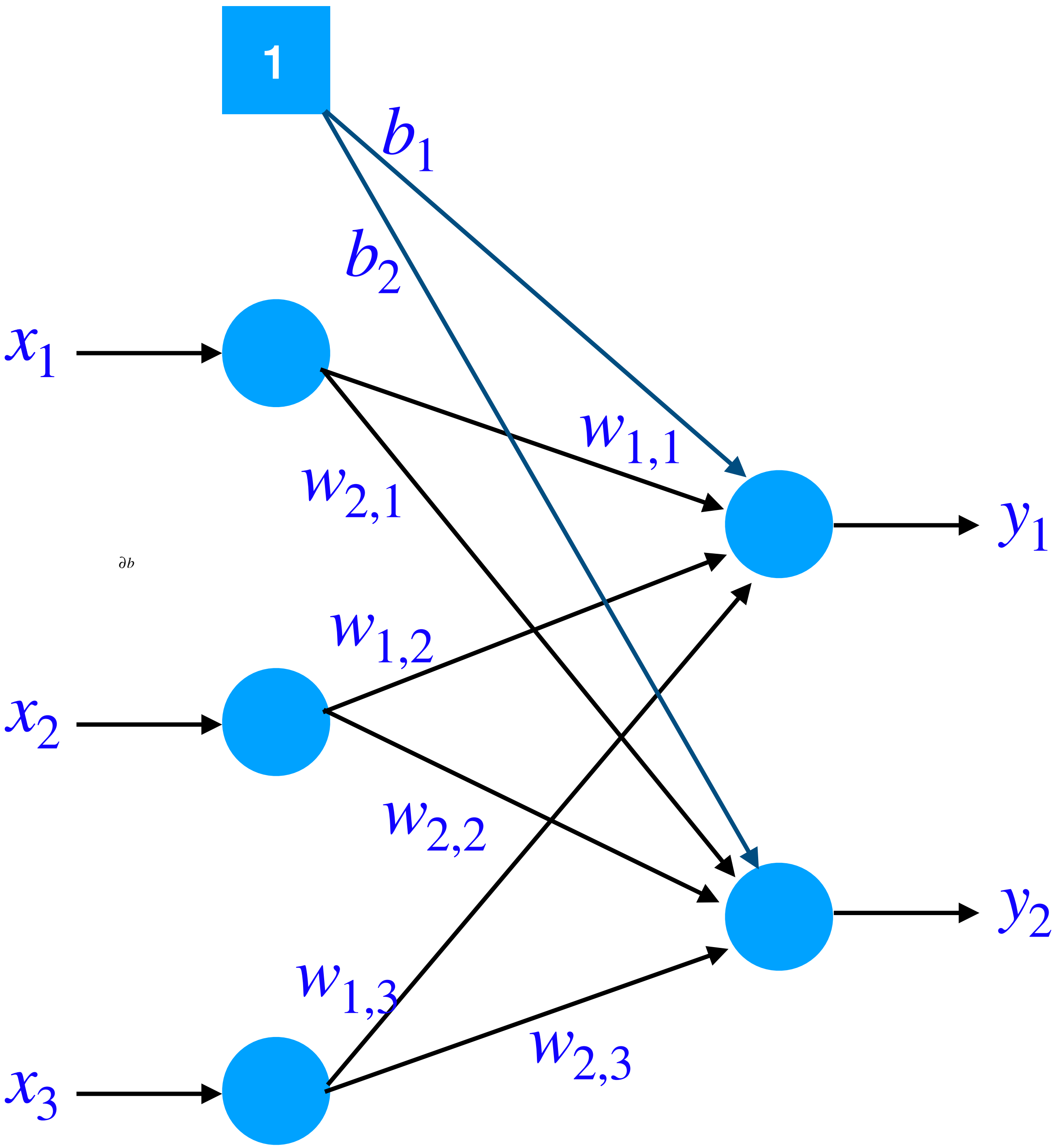
$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

$$\delta w_{n,m} = \frac{\partial J}{\partial w_{n,m}} = ?$$

Back-propagation through FC-Layer

Derivation of ΔW



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Chain rule:

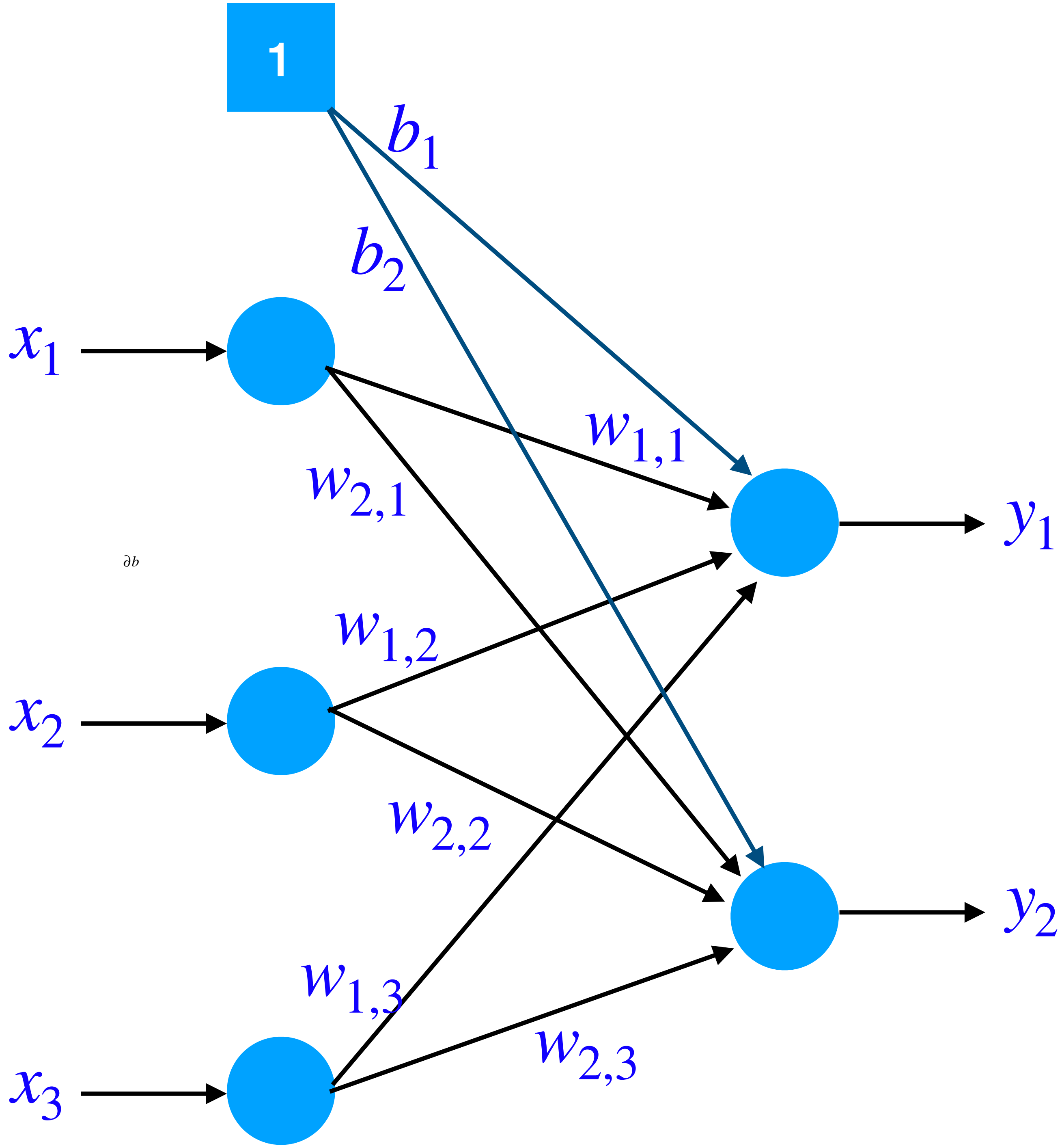
$$\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,1}}$$

$$\frac{\partial J}{\partial w_{1,2}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,2}}$$

$$\frac{\partial J}{\partial w_{1,3}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,3}}$$

Back-propagation through FC-Layer

Derivation of ΔW



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Chain rule:

$$\frac{\partial J}{\partial w_{2,1}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,1}}$$

$$\frac{\partial J}{\partial w_{2,2}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,2}}$$

$$\frac{\partial J}{\partial w_{2,3}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,3}}$$

Back-propagation through FC-Layer

Derivation of ΔW

$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

∂b

$$\frac{\partial J}{\partial w_{1,1}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,1}} = \delta y_1 x_1 \qquad \frac{\partial J}{\partial w_{2,1}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,1}} = \delta y_2 x_1$$

$$\frac{\partial J}{\partial w_{1,2}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,2}} = \delta y_1 x_2 \qquad \frac{\partial J}{\partial w_{2,2}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,2}} = \delta y_2 x_2$$

$$\frac{\partial J}{\partial w_{1,3}} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial w_{1,3}} = \delta y_1 x_3 \qquad \frac{\partial J}{\partial w_{2,3}} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial w_{2,3}} = \delta y_2 x_3$$

Back-propagation through FC-Layer
Derivation of ΔW

$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

∂b

$\delta w_{1,1} = \delta y_1 x_1$	$\delta w_{1,2} = \delta y_1 x_2$	$\delta w_{1,3} = \delta y_1 x_3$
$\delta w_{2,1} = \delta y_2 x_1$	$\delta w_{2,2} = \delta y_2 x_2$	$\delta w_{2,3} = \delta y_2 x_3$

Back-propagation through FC-Layer

Derivation of ΔW

Notation:

$$\Delta\mathbf{y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \end{bmatrix}$$

partial derivatives of loss function with respect to each value in the output of fully-connected layer

∂b

$$\Delta\mathbf{W} = \begin{bmatrix} \delta w_{1,1} & \delta w_{1,2} & \delta w_{1,3} \\ \delta w_{2,1} & \delta w_{2,2} & \delta w_{2,3} \end{bmatrix}$$

partial derivatives of loss function with respect to each value the weight matrix

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \longrightarrow \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

Back-propagation through FC-Layer

Derivation of ΔW

$$\boxed{\mathbf{\Delta W} = \mathbf{\Delta y} * \mathbf{x}^T}$$

$\mathbf{x}^T =$

x_1	x_2	x_3

∂b

$\mathbf{\Delta y} =$

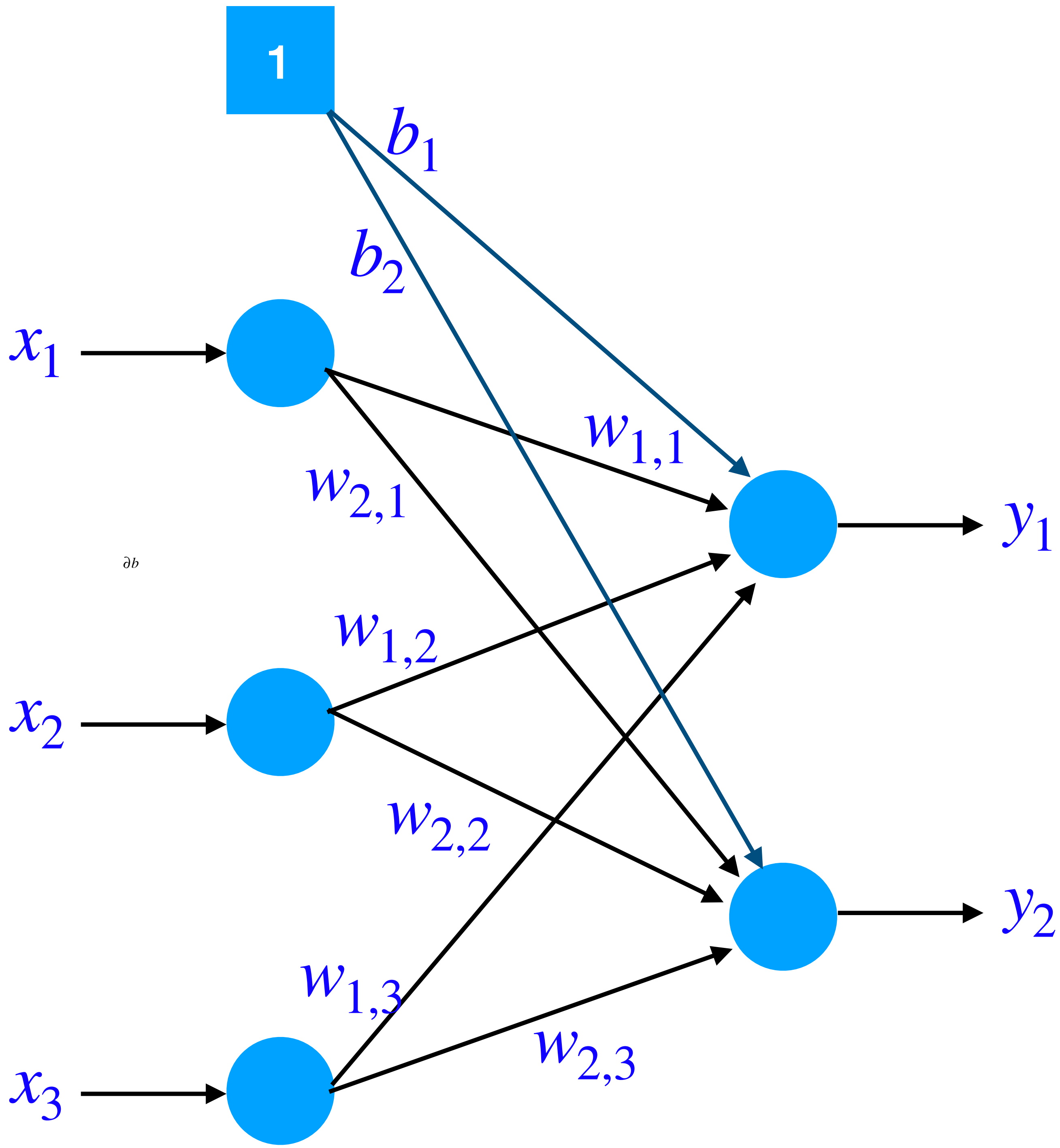
δy_1
δy_2

$\delta w_{1,1} = \delta y_1 x_1$	$\delta w_{1,2} = \delta y_1 x_2$	$\delta w_{1,3} = \delta y_1 x_3$
$\delta w_{2,1} = \delta y_2 x_1$	$\delta w_{2,2} = \delta y_2 x_2$	$\delta w_{2,3} = \delta y_2 x_3$

$\mathbf{\Delta W}$

Back-propagation through FC-Layer

Derivation of Δb



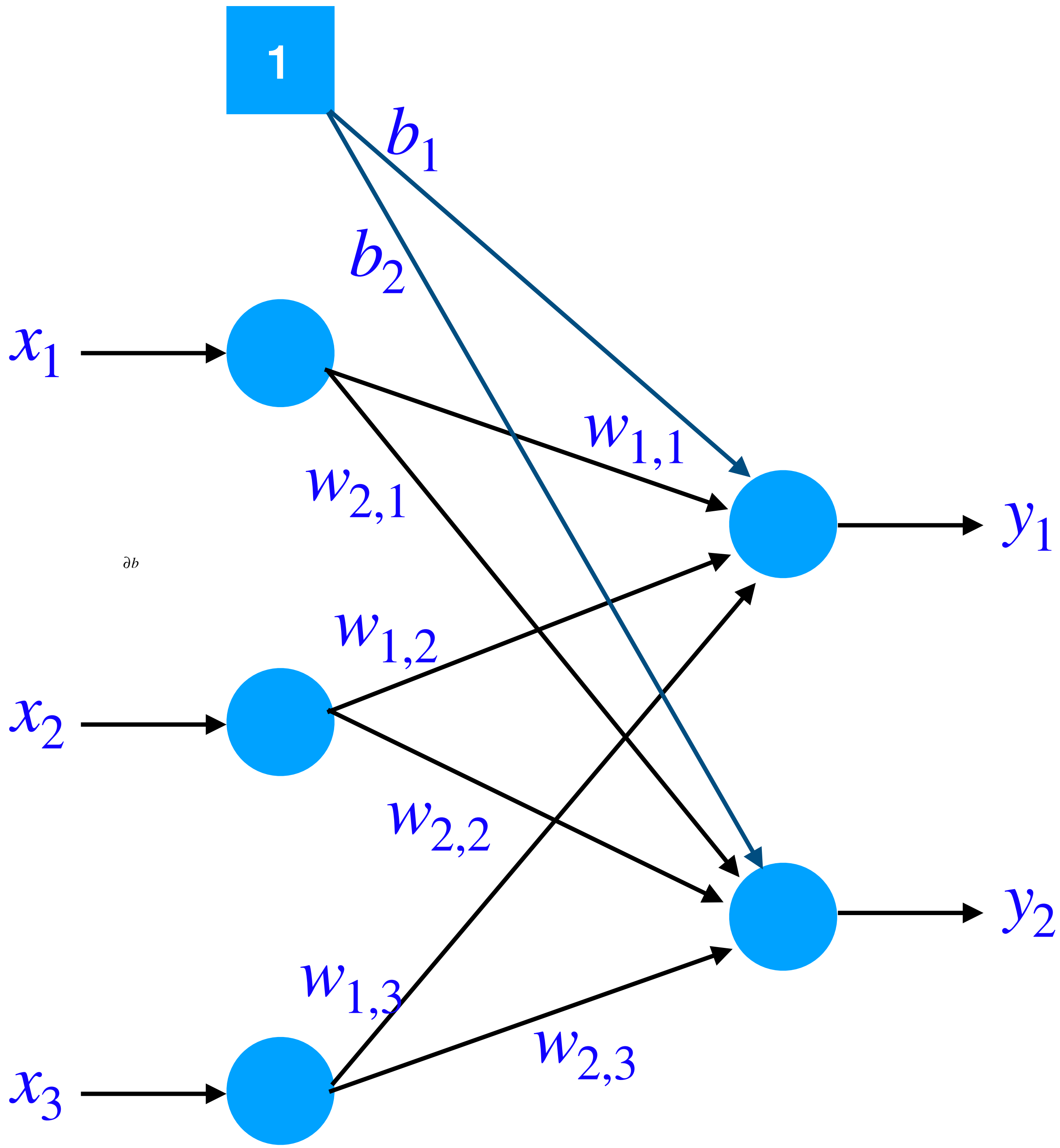
$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

$$\delta b_n = \frac{\partial J}{\partial b_n} = ?$$

Back-propagation through FC-Layer

Derivation of Δb



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

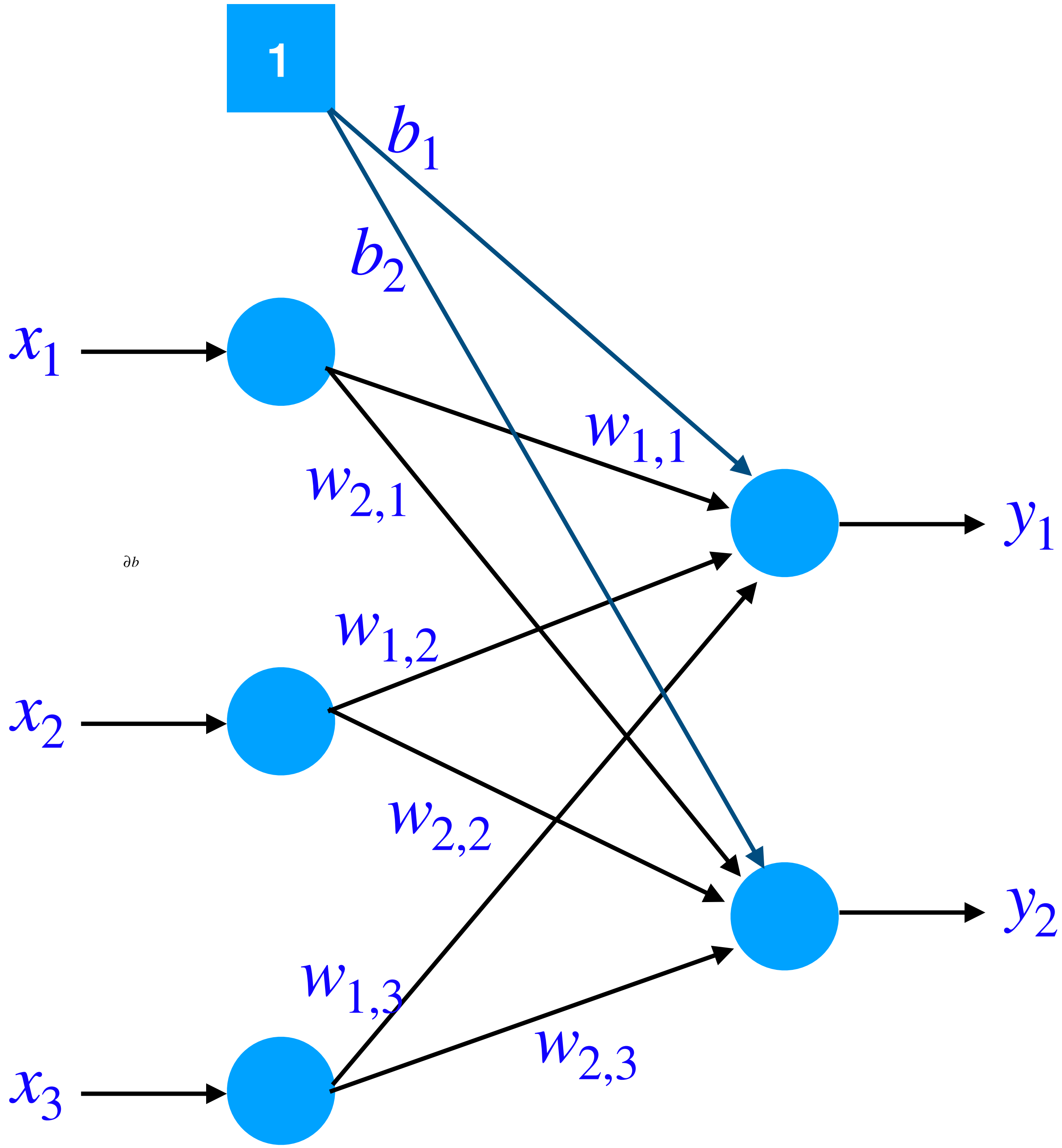
Chain rule:

$$\frac{\partial J}{b_1} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial b_1} = \delta y_1$$

$$\frac{\partial J}{b_2} = \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial b_2} = \delta y_2$$

Back-propagation through FC-Layer

Derivation of Δb



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Chain rule:

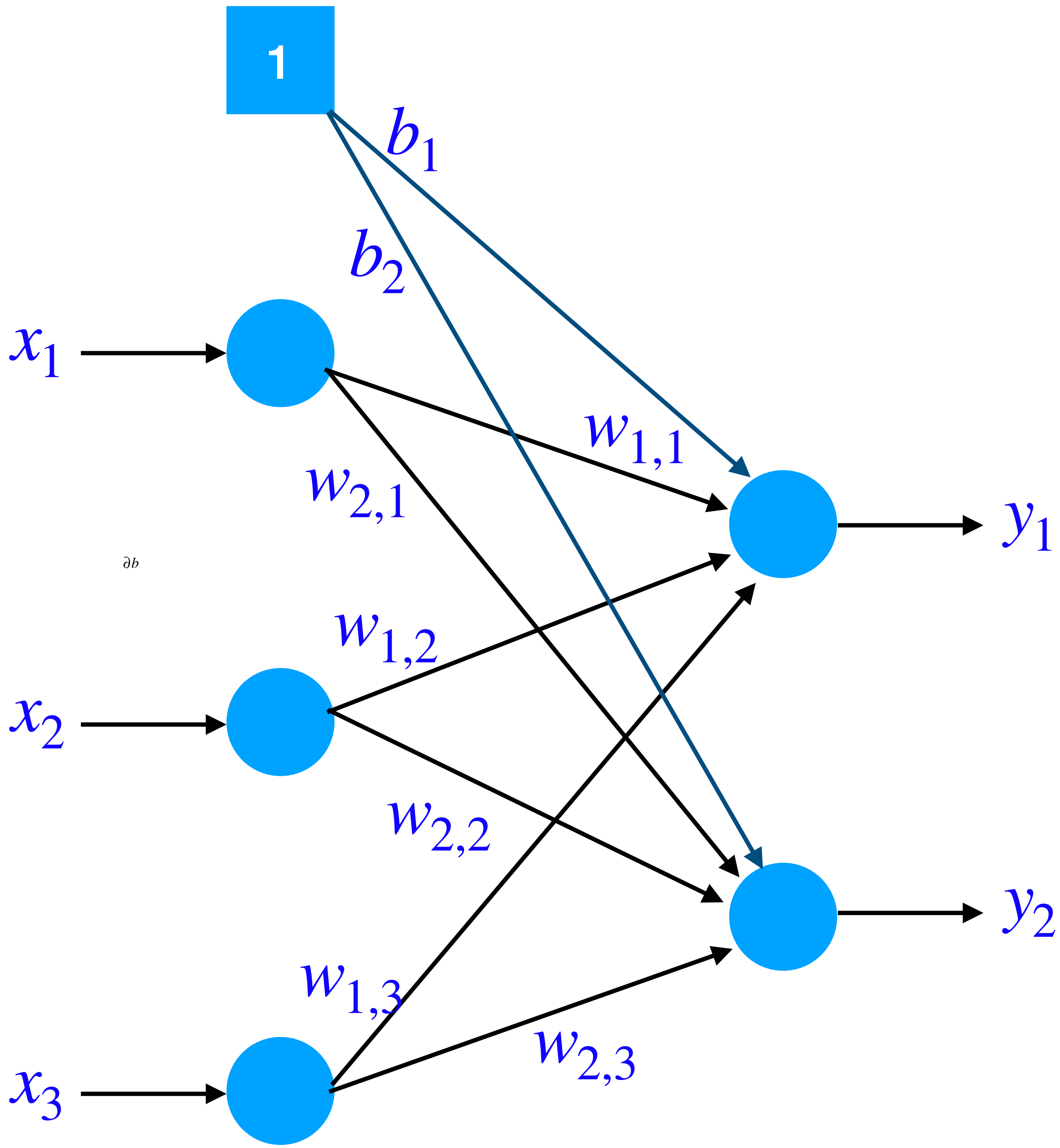
$$\delta b_1 = \delta y_1$$

$$\delta b_2 = \delta y_2$$

$$\boxed{\mathbf{\Delta b} = \mathbf{\Delta y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \end{bmatrix}}$$

Back-propagation through FC-Layer

Derivation of ΔX



$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

$$\delta x_m = \frac{\partial J}{\partial x_m} = ?$$

Back-propagation through FC-Layer

Derivation of ΔX

GV LAB

$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Chain rule:

∂b

$$\frac{\partial J}{\partial x_1} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial x_1} = \delta y_1 w_{1,1} + \delta y_2 w_{2,1}$$

$$\frac{\partial J}{\partial x_2} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial x_2} + \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial x_2} = \delta y_1 w_{1,2} + \delta y_2 w_{2,2}$$

$$\frac{\partial J}{\partial x_3} = \frac{\partial J}{\partial y_1}\frac{\partial y_1}{\partial x_3} + \frac{\partial J}{\partial y_2}\frac{\partial y_2}{\partial x_3} = \delta y_1 w_{1,3} + \delta y_2 w_{2,3}$$

Back-propagation through FC-Layer

Derivation of ΔX

$$y_1 = w_{1,1}x_1 + w_{1,2}x_2 + w_{1,3}x_3 + b_1$$

$$y_2 = w_{2,1}x_1 + w_{2,2}x_2 + w_{2,3}x_3 + b_2$$

Chain rule:

∂b

$$\delta x_1 = \delta y_1 w_{1,1} + \delta y_2 w_{2,1}$$

$$\delta x_2 = \delta y_1 w_{1,2} + \delta y_2 w_{2,2}$$

$$\delta x_3 = \delta y_1 w_{1,3} + \delta y_2 w_{2,3}$$

Back-propagation through FC-Layer

Derivation of ΔX

Notation:

$$\Delta \mathbf{y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \end{bmatrix}$$

∂b

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,2} \\ w_{2,1} & w_{2,2} & w_{2,3} \end{bmatrix} \longrightarrow \mathbf{W}^T = \begin{bmatrix} w_{1,1} & w_{2,1} \\ w_{1,2} & w_{2,2} \\ w_{1,2} & w_{2,3} \end{bmatrix}$$

Back-propagation through FC-Layer

Derivation of ΔX

$$\boxed{\Delta\mathbf{X} = \mathbf{W}^T * \Delta\mathbf{y}}$$

$$\Delta\mathbf{y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \end{bmatrix}$$

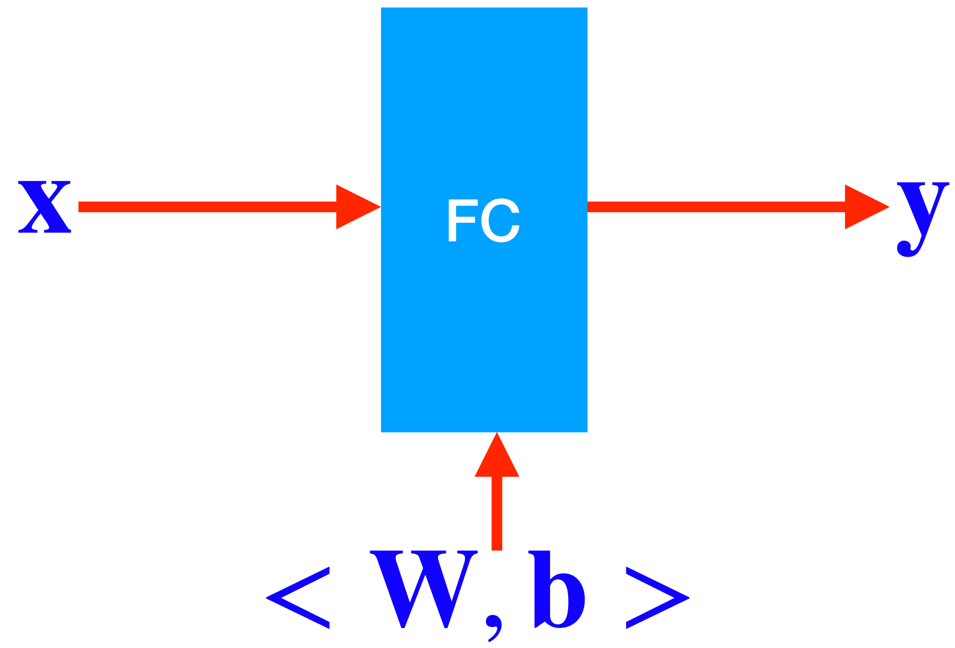
∂b

$$\mathbf{W}^T = \begin{bmatrix} w_{1,1} & w_{2,1} \\ w_{1,2} & w_{2,2} \\ w_{1,2} & w_{2,3} \end{bmatrix}$$

$$\begin{bmatrix} \delta x_1 = \delta y_1 w_{1,1} + \delta y_2 w_{2,1} \\ \delta x_2 = \delta y_1 w_{1,2} + \delta y_2 w_{2,2} \\ \delta x_3 = \delta y_1 w_{1,3} + \delta y_2 w_{2,3} \end{bmatrix}$$

$$\Delta\mathbf{X}$$

Summary



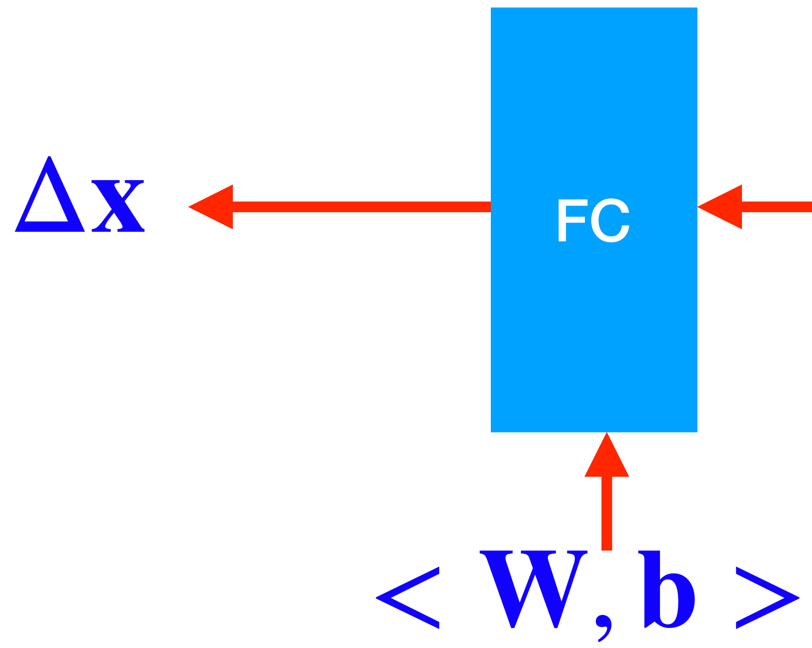
∂b

Forward-pass:

$$\boxed{\mathbf{y} = \mathbf{W} * \mathbf{x} + \mathbf{b}}$$

(Matrix multiplication)

Summary

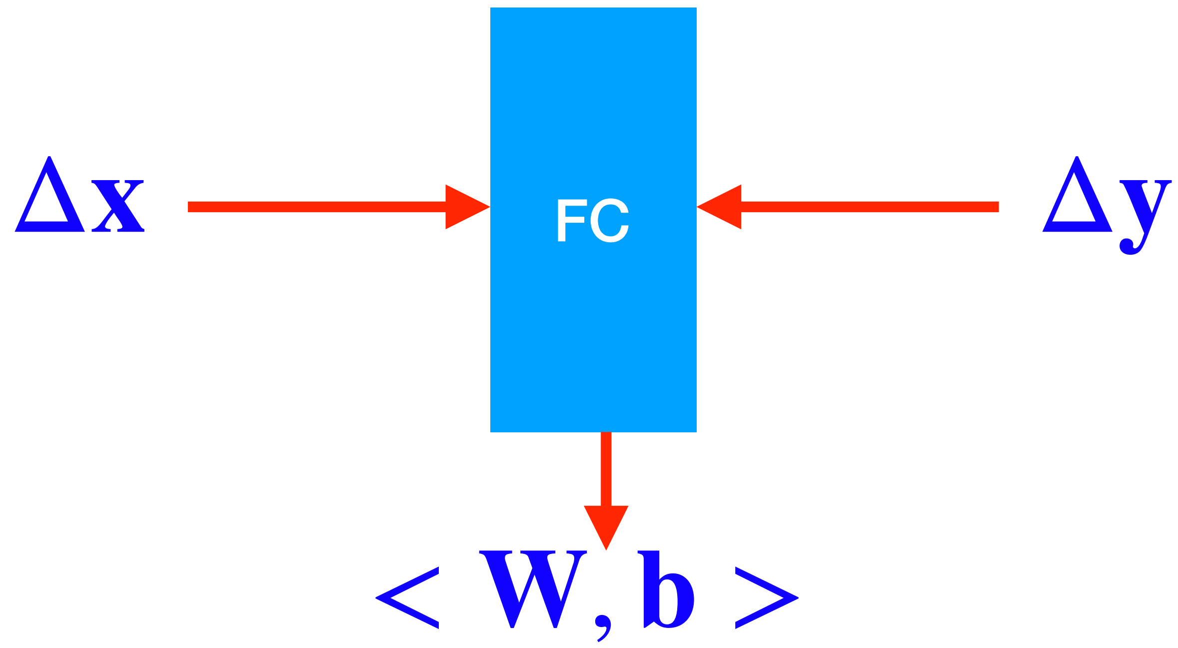


∂b

Backward-pass: $$\boxed{\Delta\mathbf{X} = \mathbf{W}^T * \Delta\mathbf{y}}$$

(Matrix multiplication)

Summary



∂b

Backward-pass:

$$\Delta \mathbf{W} = \Delta \mathbf{y} * \mathbf{x}^T$$

$$\Delta \mathbf{b} = \Delta \mathbf{y}$$

(Matrix multiplication)