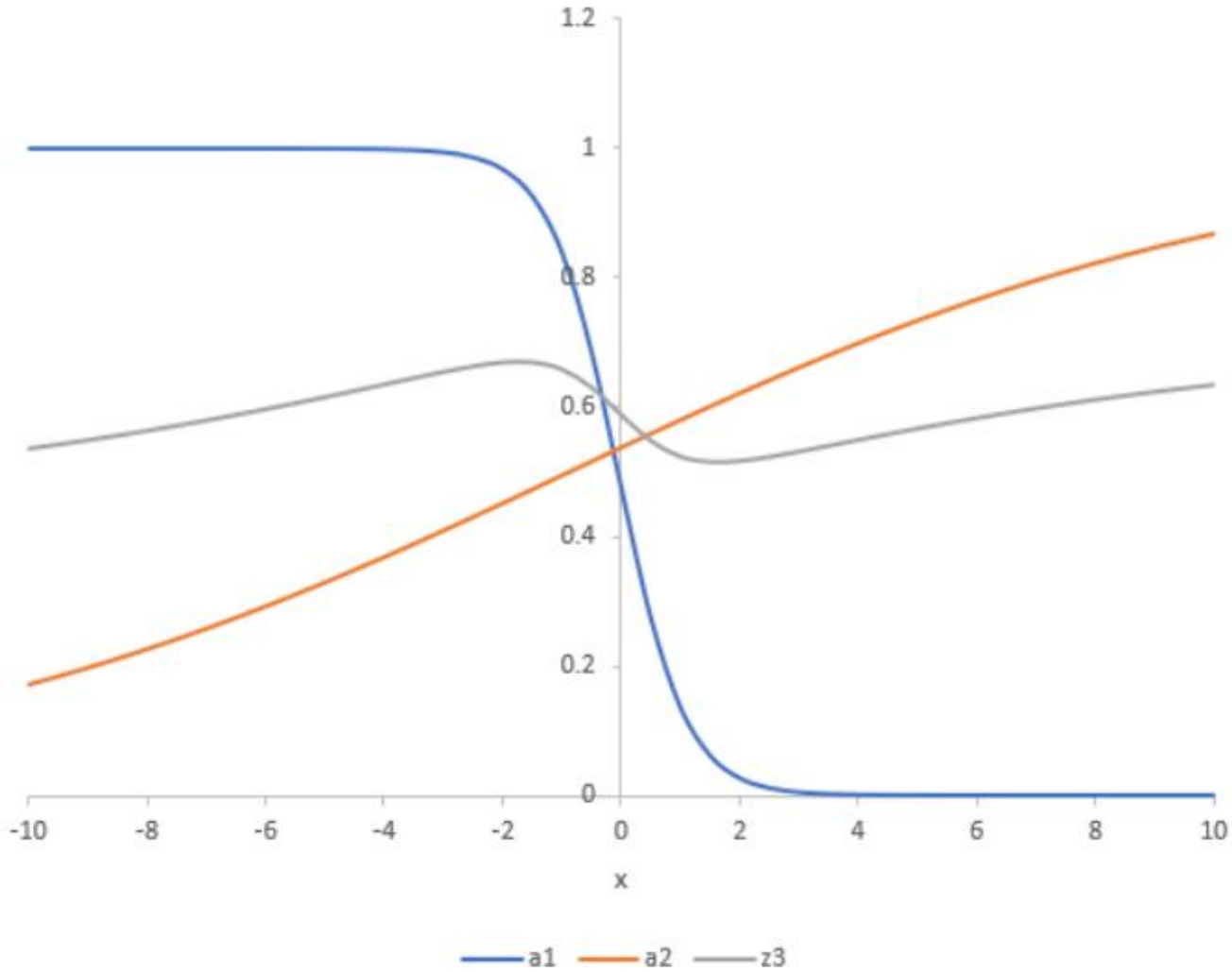
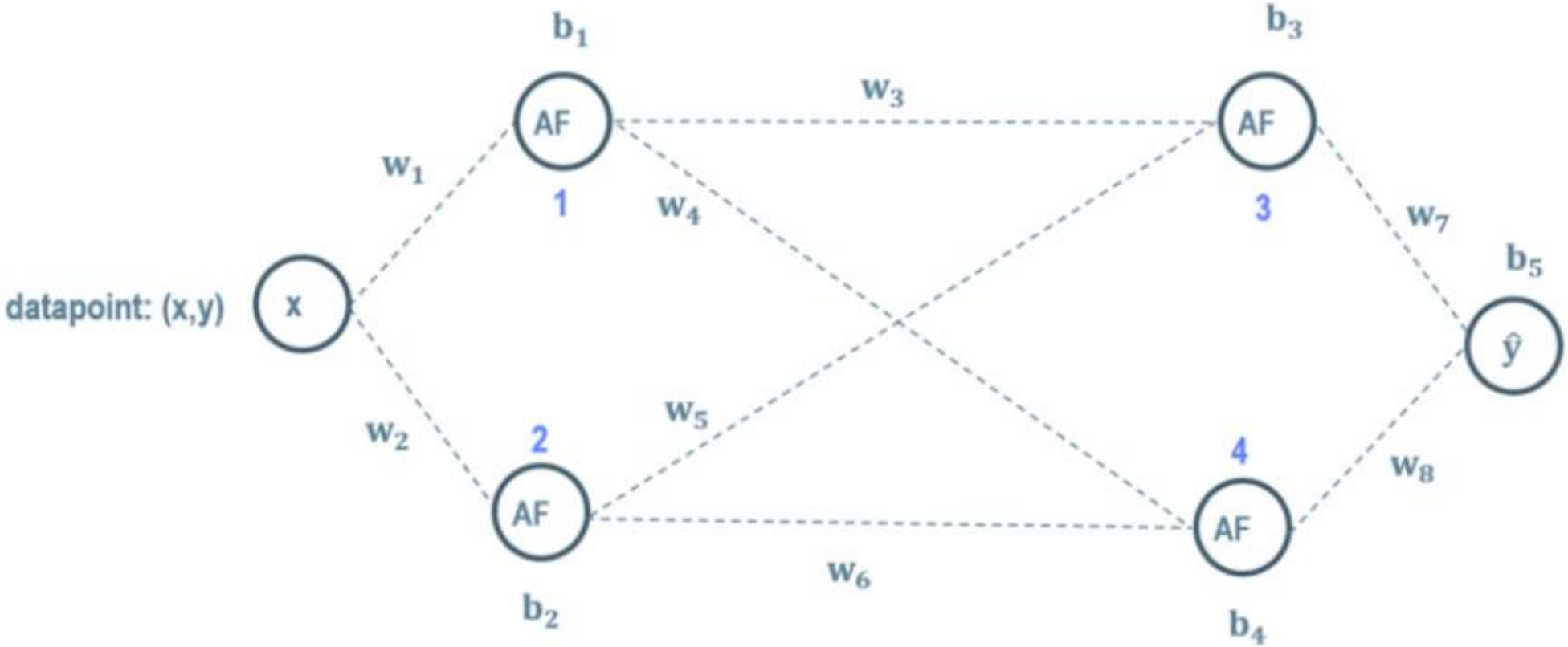
$$z1(x) = -1.75x - 0.1 \text{ and } z2(x) = 0.172x + 0.15$$

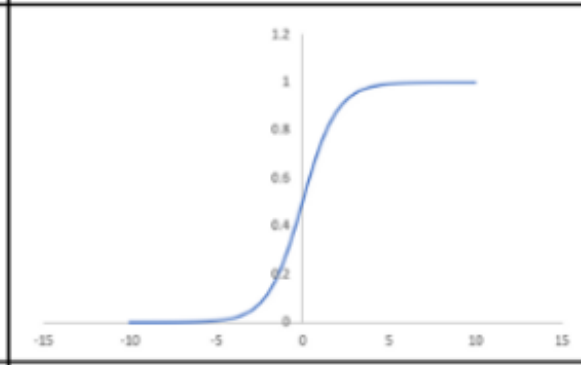
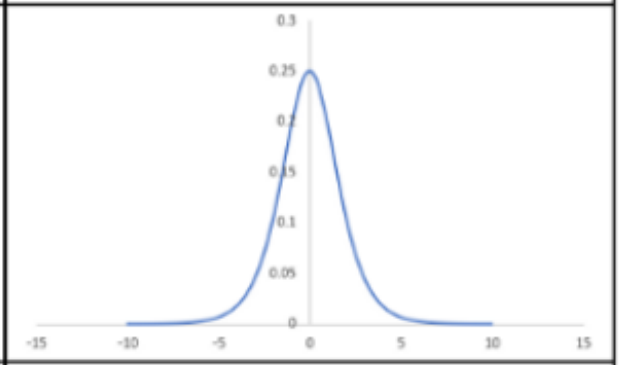
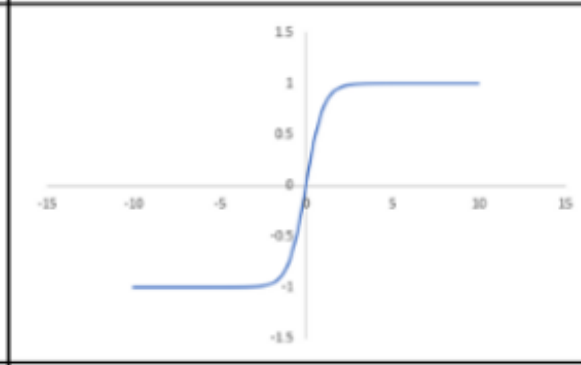
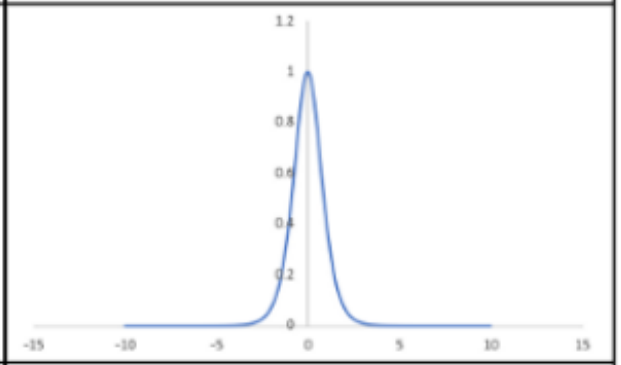
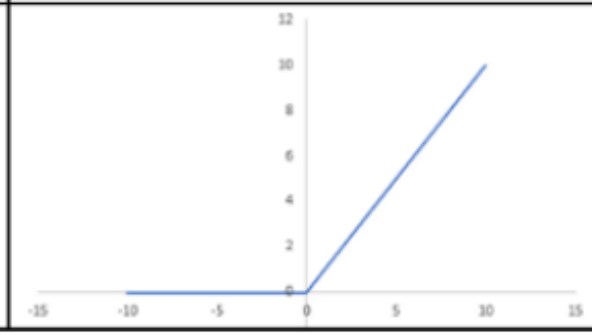
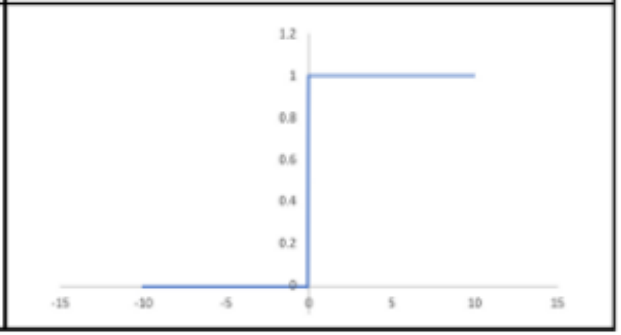
$$f(x) = \frac{1}{1+e^{-x}}$$

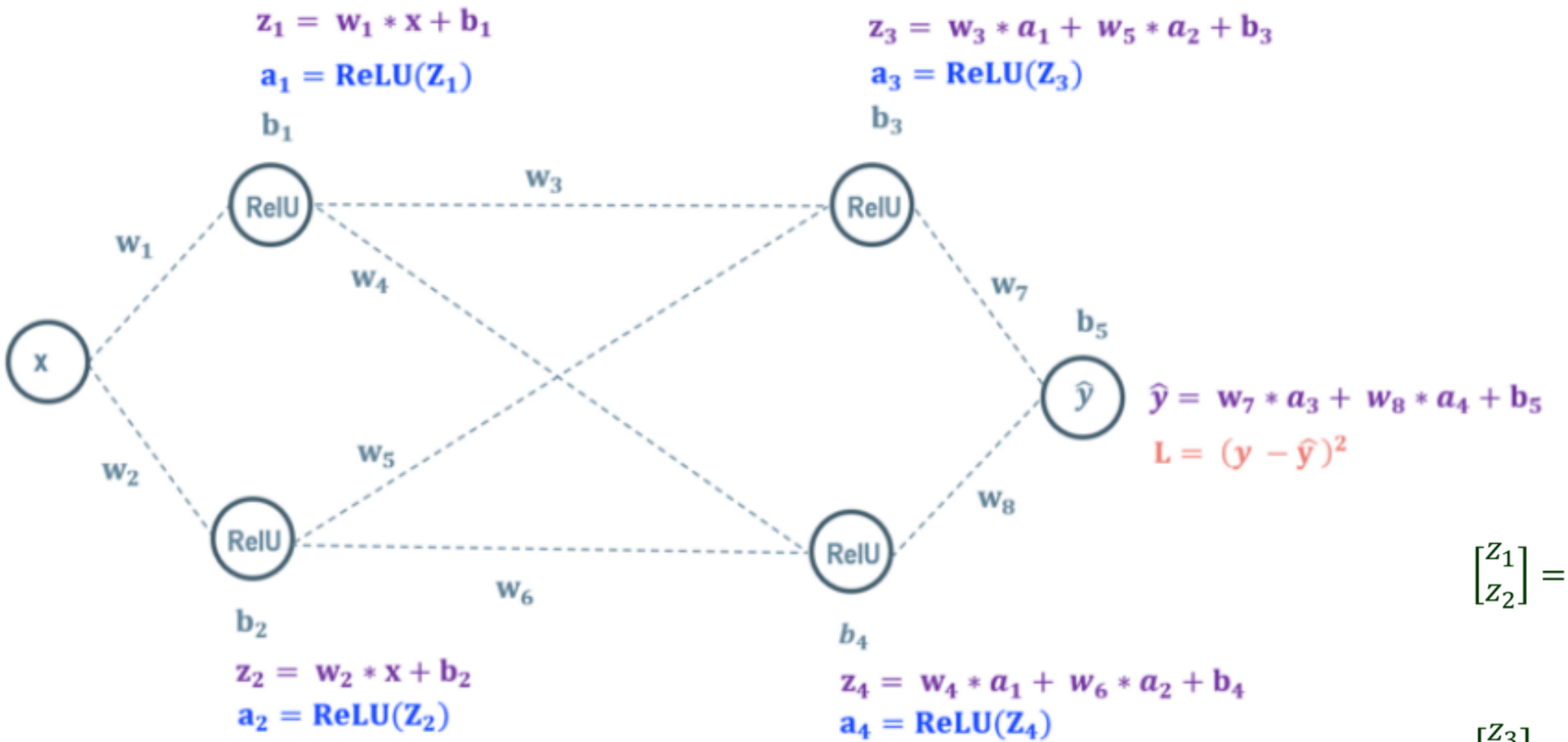
$$a1(z1) = \frac{1}{1+ e^{-z1}} \text{ and } a2(z2) = \frac{1}{1+ e^{-z2}}$$

$$z3 = 0.25 * a1 + 0.5 * a2 + 0.2$$




b_1
b_3
AF
AF
w_3
w_1
w_4
w_7
b_5
1
3
datapoint: (x,y)
x
$\hat{y}$
w_5
w_2
2
4
w_8
AF
AF
w_6
b_2
b_4

Activation function	f(x)	f(x)	$\frac{df(x)}{dx}$
sigmoid	$\frac{1}{1+e^{-x}}$		
tanh	$tanh(x)$		
RelU	max(0,x)		



$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} w_1 & b_1 \\ w_2 & b_2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

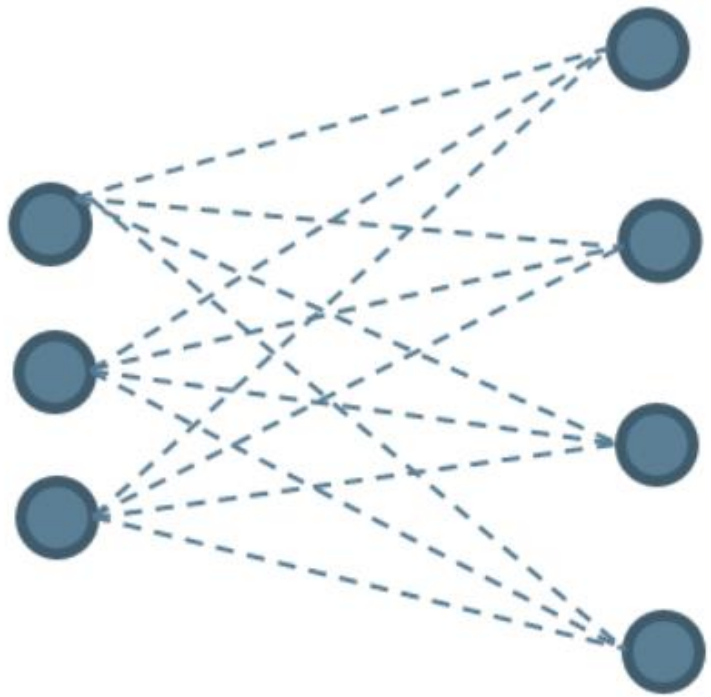
(2 X 2)*(2 X 1)

$$\begin{bmatrix} z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} w_3 & w_5 & b_3 \\ w_4 & w_6 & b_6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ 1 \end{bmatrix}$$

(2 X 2+1)*(2+1 X 1)

$$[\hat{y}] = [w_7 \quad w_8 \quad b_5] \begin{bmatrix} a_3 \\ a_4 \\ 1 \end{bmatrix}$$

(1 X 2+1)*(2+1 X 1)



r nodes s nodes

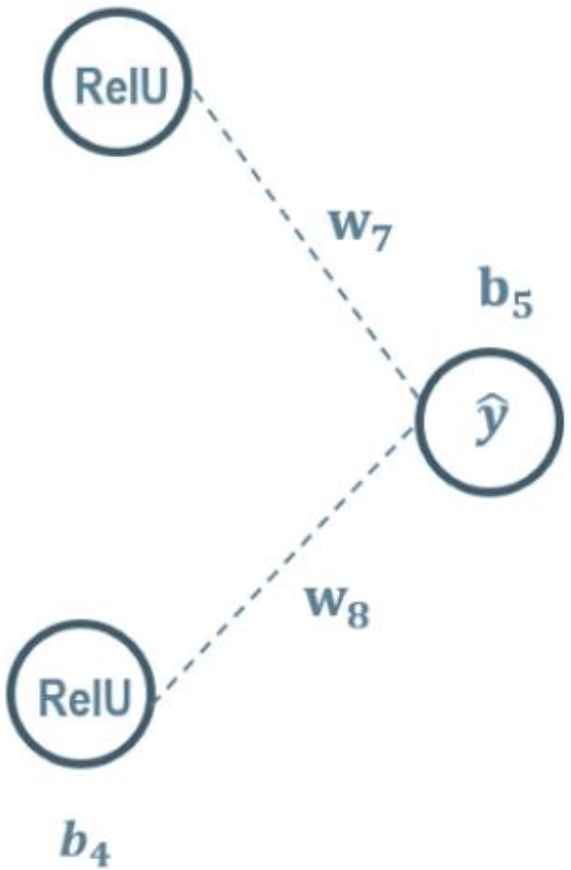
$$w_i^{n+1} = w_i^n - \eta \nabla_w L, i = 1 \dots m.$$

$$\nabla_w L = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \vdots \\ \frac{\partial L}{\partial w_m} \end{bmatrix}$$

$$z_3 = w_3 * a_1 + w_5 * a_2 + b_3$$

$$a_3 = \mathrm{ReLU}(Z_3)$$

b_3



$$\hat{y} = w_7 * a_3 + w_8 * a_4 + b_5$$

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

$$z_4 = w_4 * a_1 + w_6 * a_2 + b_4$$

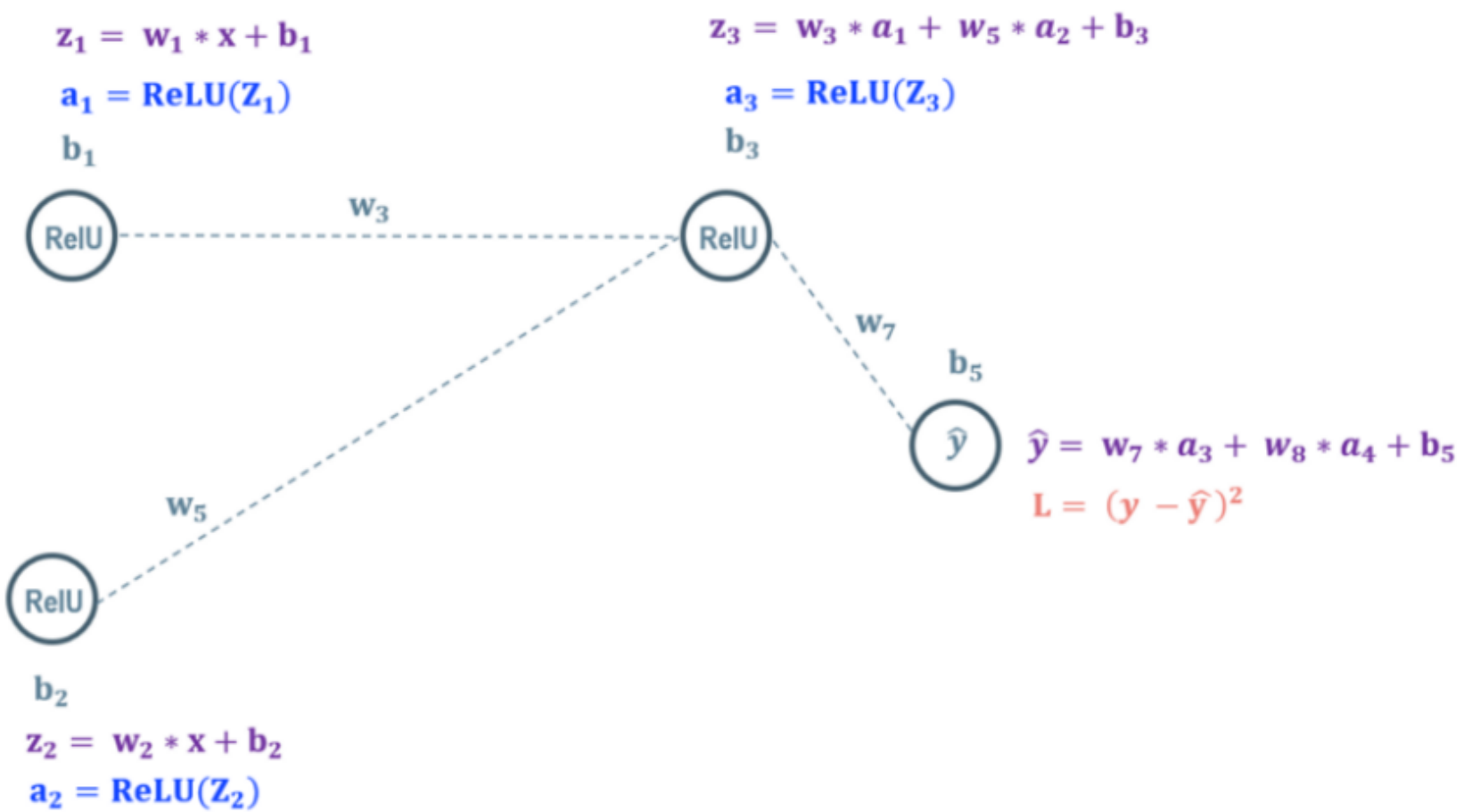
$$a_4 = \mathrm{ReLU}(Z_4)$$

$$L_y = \frac{\partial L}{\partial \hat{y}} = -2(y - \hat{y})$$

$$\frac{\partial L}{\partial w_7} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_7} = L_y a_3$$

$$\frac{\partial L}{\partial w_8} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_8} = L_y a_4$$

$$\frac{\partial L}{\partial b_5} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b_5} = L_y$$



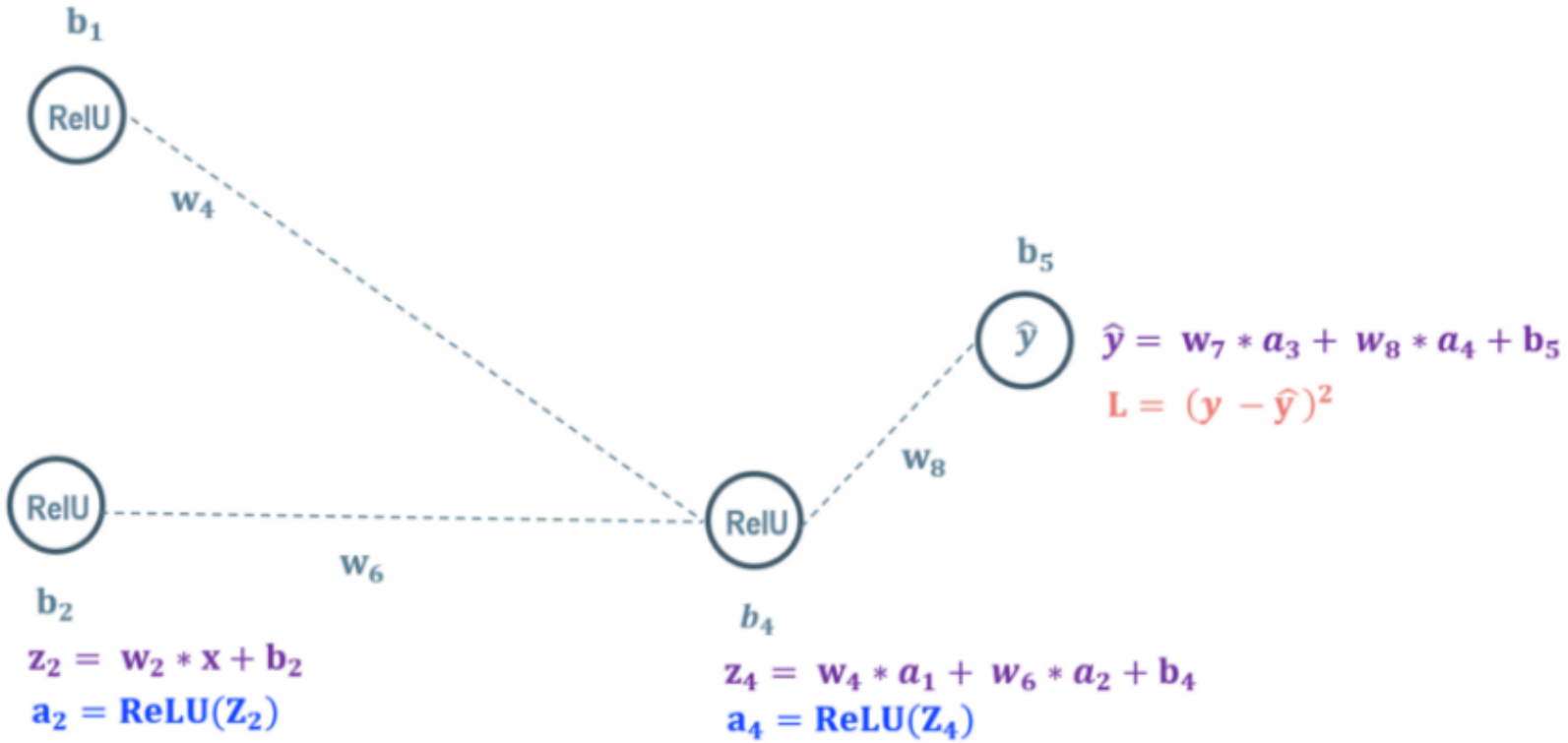
$$\frac{\partial L}{\partial w_3} = \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial w_3} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial w_3} = L_y w_7\, dRL(z_3)\, a_1, \quad \text{where } dRL(z_3) = \frac{dRelU(z_3)}{dz_3}$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial w_5} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial w_5} = L_y w_7\, dRL(z_3)\, a_2, \quad \text{where } dRL(z_3) = \frac{dRelU(z_3)}{dz_3}$$

$$\frac{\partial L}{\partial b_3} = \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial b_3} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial b_3} = L_y w_7\, dRL(z_3), \quad \text{where } dRL(z_3) = \frac{dRelU(z_3)}{dz_3}$$

$z_1 = w_1 * x + b_1$

$a_1 = \mathbf{ReLU}(Z_1)$



$z_2 = w_2 * x + b_2$

$a_2 = \mathbf{ReLU}(Z_2)$

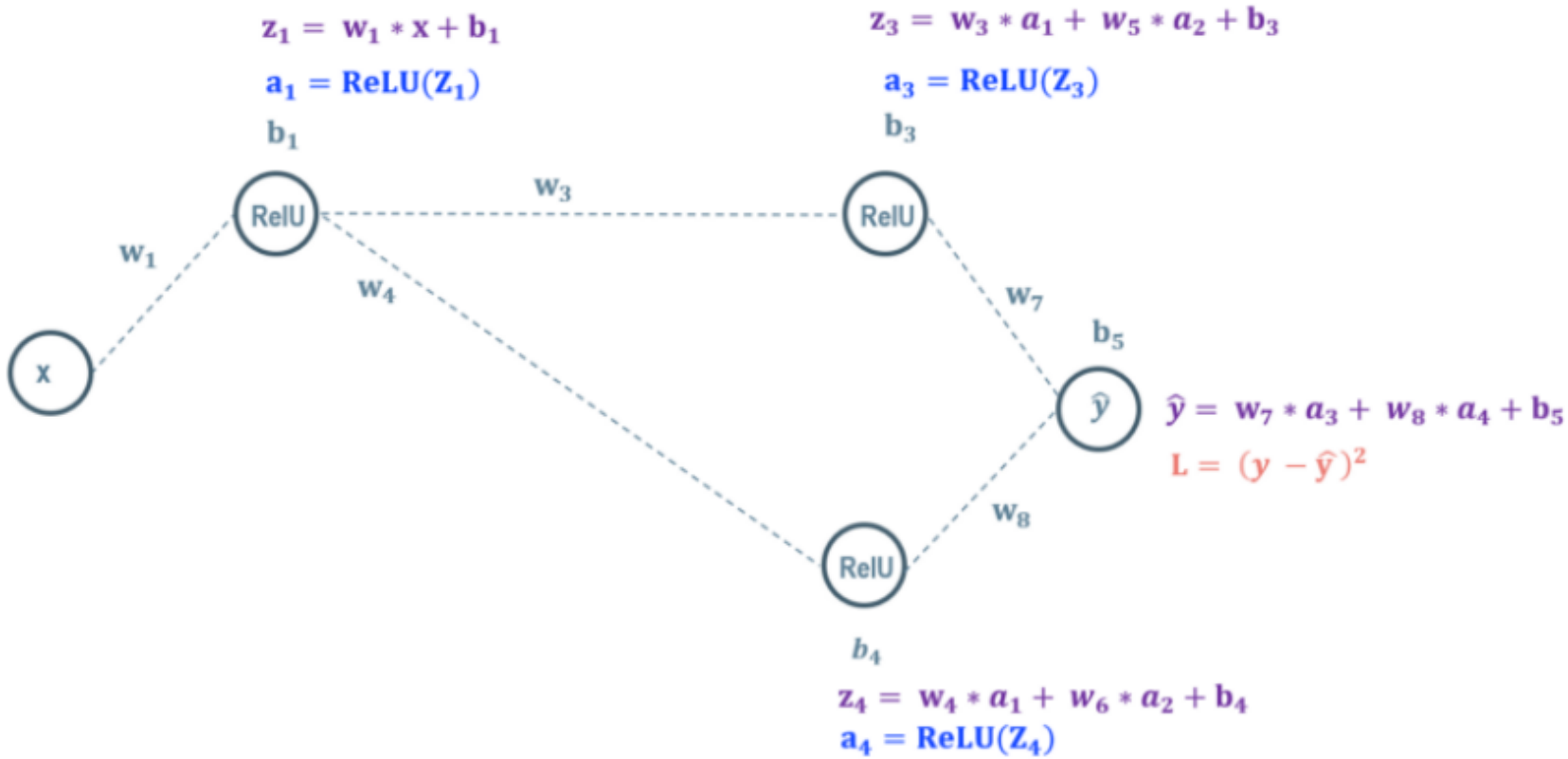
$z_4 = w_4 * a_1 + w_6 * a_2 + b_4$

$a_4 = \mathbf{ReLU}(Z_4)$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial w_6} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial w_6} = L_y w_8\, dRL(z_4)\, a_2, \quad \text{where } dRL(z_4) = \frac{dRelU(z_4)}{dz_4}$$

$$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial w_4} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial w_4} = L_y w_8\, dRL(z_4)\, a_1, \quad \text{where } dRL(z_4) = \frac{dRelU(z_4)}{dz_4}$$

$$\frac{\partial L}{\partial b_4} = \frac{\partial L}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial b_4} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial b_4} = L_y w_8\, dRL(z_4), \quad \text{where } dRL(z_4) = \frac{\mathbf{dRelU(z_4)}}{\mathbf{dz_4}}$$

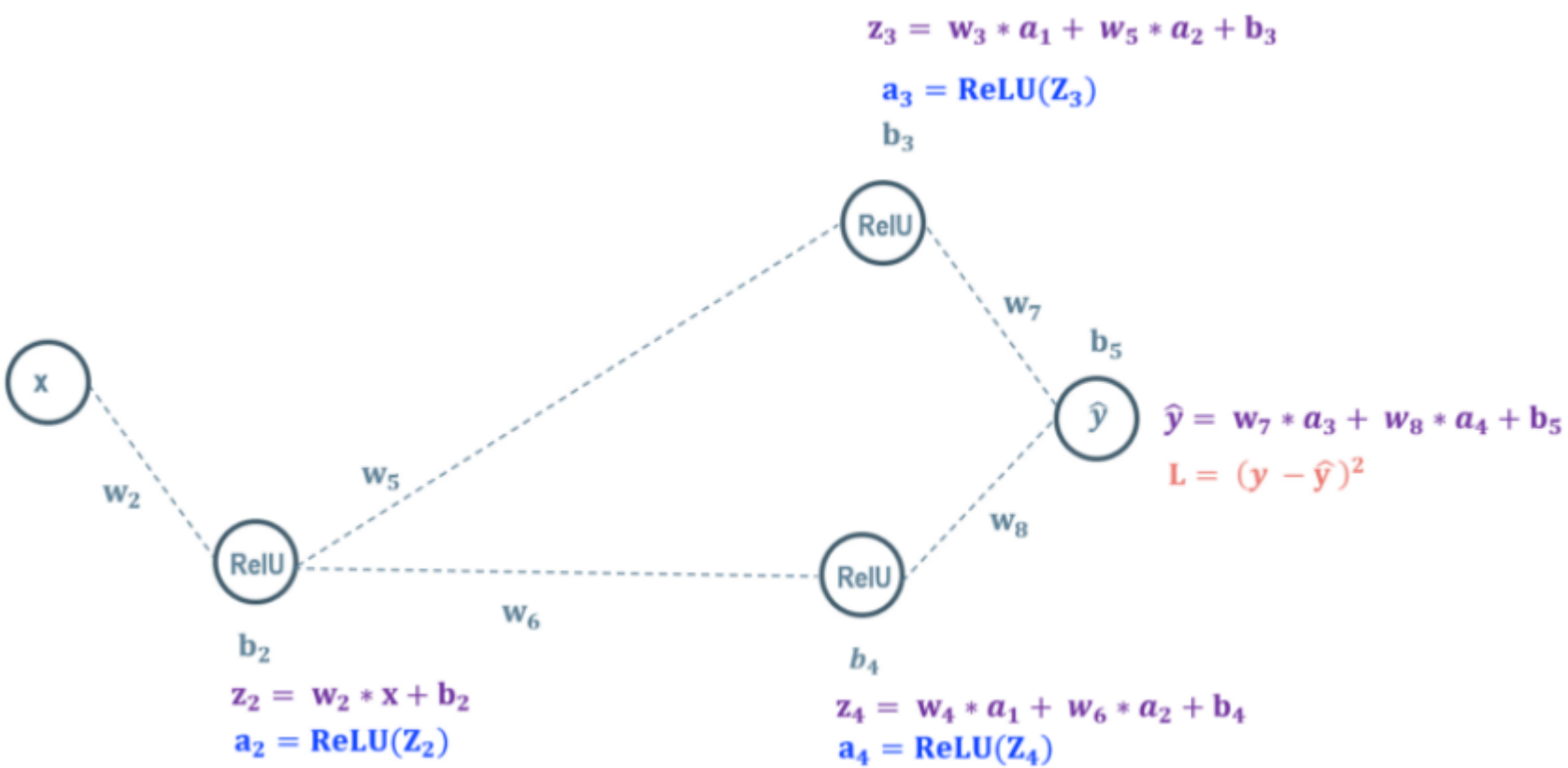


$$\hat{y} = g + h + b_5 \text{ where } g = w_7 * a_3 \text{ and } h = w_8 * a_4$$

$$\frac{\partial \hat{y}}{\partial a_1} = \frac{\partial g}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial a_1} + \frac{\partial h}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial a_1} = w_7 dRL(z_3) w_3 + w_8 dRL(z_4) w_4$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial z_1}\frac{\partial z_1}{\partial w_1} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_1}\frac{\partial a_1}{\partial z_1}\frac{\partial z_1}{\partial w_1} = L_y\{w_7 dRL(z_3) w_3 + w_8 dRL(z_4) w_4\}\, dRL(z_1) x$$

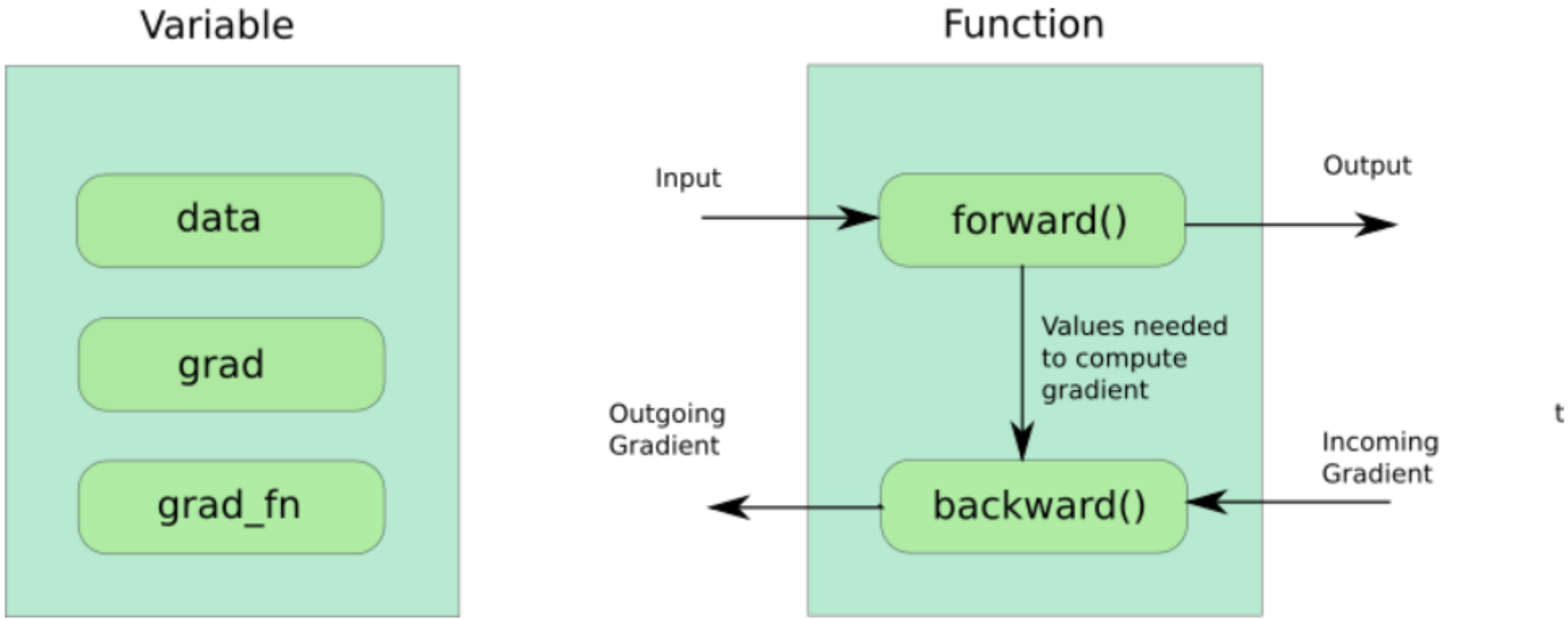
$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial z_1}\frac{\partial z_1}{\partial b_1} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_1}\frac{\partial a_1}{\partial z_1}\frac{\partial z_1}{\partial b_1} = L_y\{w_7 dRL(z_3) w_3 + w_8 dRL(z_4) w_4\}\, dRL(z_1)$$



$$\frac{\partial \hat{y}}{\partial a_2} = \frac{\partial g}{\partial a_3}\frac{\partial a_3}{\partial z_3}\frac{\partial z_3}{\partial a_2} + \frac{\partial h}{\partial a_4}\frac{\partial a_4}{\partial z_4}\frac{\partial z_4}{\partial a_2} = w_7 dRL(z_3) w_5 + w_8 dRL(z_4) w_6$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial w_2} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial w_2} = L_y\{w_7 dRL(z_3) w_5 + w_8 dRL(z_4) w_6\}\, dRL(z_2) x$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial b_2} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial b_2} = L_y\{w_7 dRL(z_3) w_5 + w_8 dRL(z_4) w_6\}\, dRL(z_2)$$


Variable
data
grad
grad_fn
Function
Input
forward()
Output
Values needed
to compute
gradient
Outgoing
Gradient
backward()
Incoming
Gradient