Basics of Neural Network Programming

deeplearning.ai

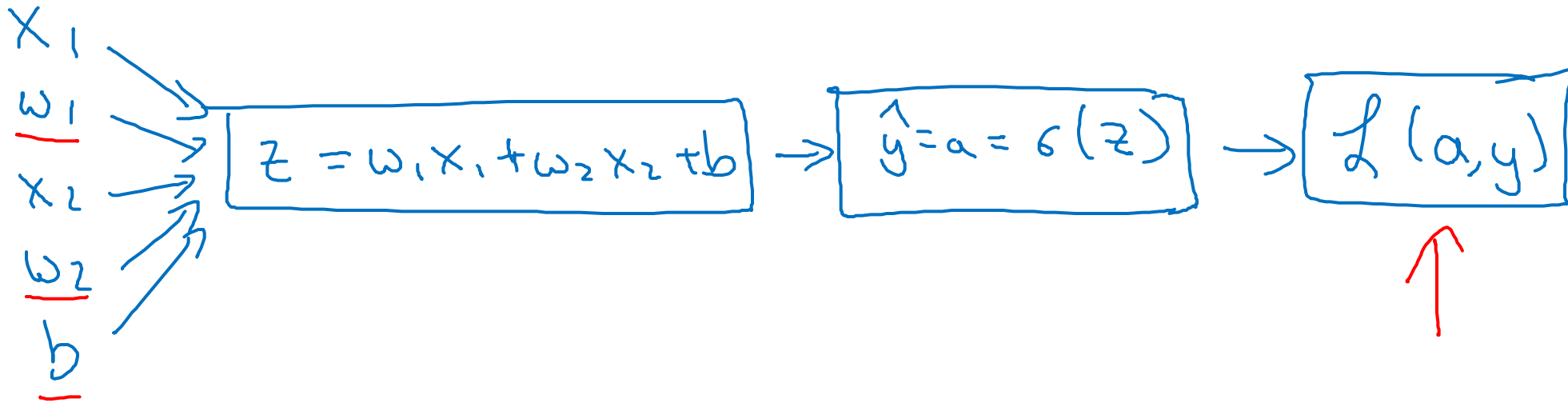## Logistic Regression Gradient descent

Logistic regression recap

$$z = w^T x + b$$

$$\hat{y} = a = \sigma(z)$$

$$\mathcal{L}(a, y) = -(y \log(a) + (1 - y) \log(1 - a))$$

$$z = w_1 x_1 + w_2 x_2 + b \rightarrow \hat{y} = a = \sigma(z) \rightarrow \mathcal{L}(a, y)$$

Logistic regression derivatives

x_1

w_1

x_2

w_2

b

$$z = w_1 x_1 + w_2 x_2 + b$$

$$a = \sigma(z)$$

$$\mathcal{L}(a, y)$$

$$dz = \frac{d\mathcal{L}}{dz} = \frac{d\mathcal{L}(a,y)}{dz}$$

$$"da" = \frac{d\mathcal{L}(a,y)}{da}$$

$$= a - y$$

$$a(1-a)$$

$$= -\frac{y}{a} + \frac{1-y}{1-a}$$

$$= \frac{d\mathcal{L}}{da} \cdot \frac{da}{dz}$$

$$\frac{d\mathcal{L}}{dw_1} = "dw_1" = x_1 \cdot dz.$$

$$dw_2 = x_2 \cdot dz. \quad db = dz.$$

$$w_1 := w_1 - \alpha \, dw_1$$

$$w_2 := w_2 - \alpha \, dw_2$$

$$b := b - \alpha \, db.$$

Andrew Ng
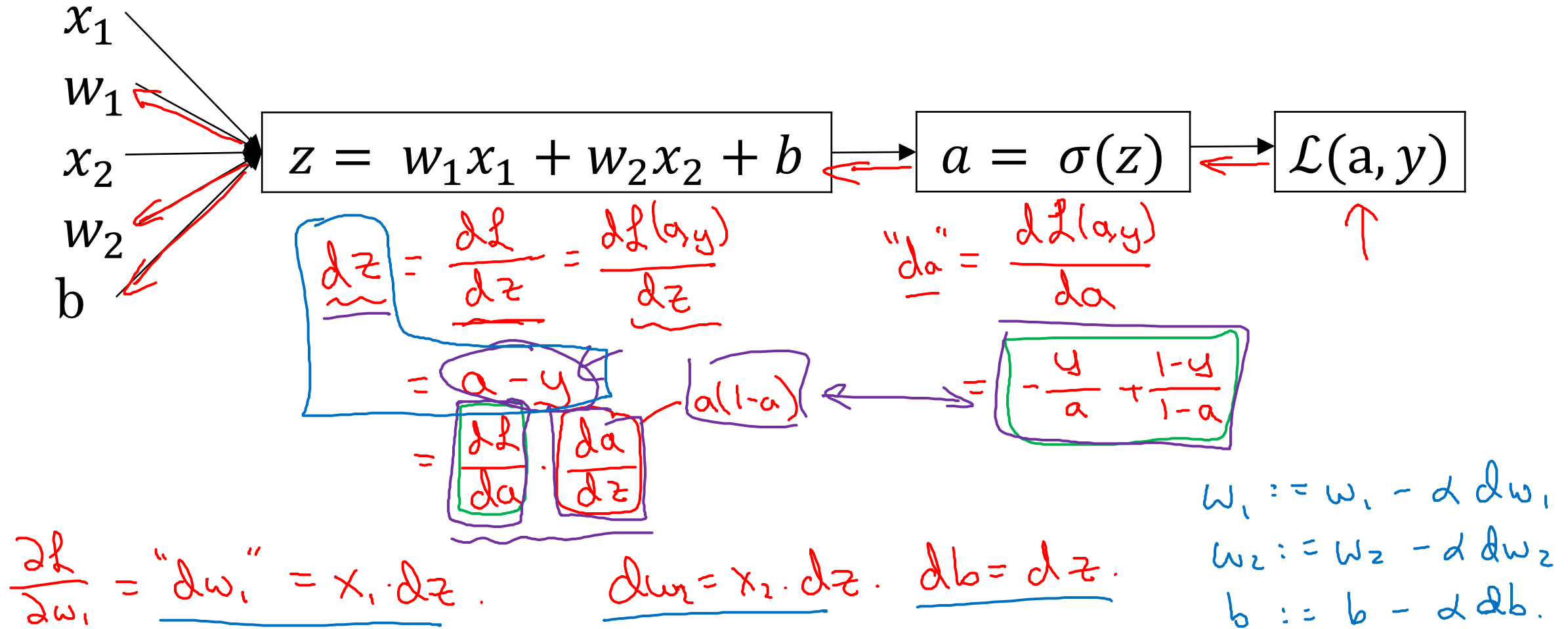
Basics of Neural Network Programming

Gradient descent on m examples

deeplearning.ai

Logistic regression on m examples

$$J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(a^{(i)}, y^{(i)})$$

$$\rightarrow a^{(i)} = \hat{y}^{(i)} = \sigma(z^{(i)}) = \sigma(w^T x^{(i)} + b)$$

$$(x^{(i)}, y^{(i)})$$

$$dw_1^{(i)}, \, dw_2^{(i)}, \, db^{(i)}$$

$$\frac{\partial}{\partial w_1} J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\frac{\partial}{\partial w_1} \mathcal{L}(a^{(i)}, y^{(i)})}_{dw_1^{(i)} - (x^{(i)}, y^{(i)})}$$

Logistic regression on m examples

$J = 0$; $dw_1 = 0$; $dw_2 = 0$; $db = 0$

$\rightarrow$ For $i = 1$ to m

$\qquad z^{(i)} = w^T x^{(i)} + b$

$\qquad a^{(i)} = \sigma(z^{(i)})$

$\qquad J \mathrel{+}= -\left[y^{(i)} \log a^{(i)} + (1 - y^{(i)}) \log(1 - a^{(i)}) \right]$

$\qquad dz^{(i)} = a^{(i)} - y^{(i)}$

$\qquad dw_1 \mathrel{+}= x_1^{(i)} dz^{(i)}$

$\qquad dw_2 \mathrel{+}= x_2^{(i)} dz^{(i)}$ $\qquad n = 2$

dw_3
$\vdots$
dw_n

$\qquad db \mathrel{+}= dz^{(i)}$

$J /= m$ $\Leftarrow$

$dw_1 /= m$; $dw_2 /= m$; $db /= m.$ $\Leftarrow$

$dw_1 = \dfrac{\partial J}{\partial w_1}$

$w_1 := w_1 - \alpha\, dw_1$

$w_2 := w_2 - \alpha\, dw_2$

$b := b - \alpha\, db.$

Vectorization

Andrew Ng