deeplearning.ai

Basics of Neural Network Programming

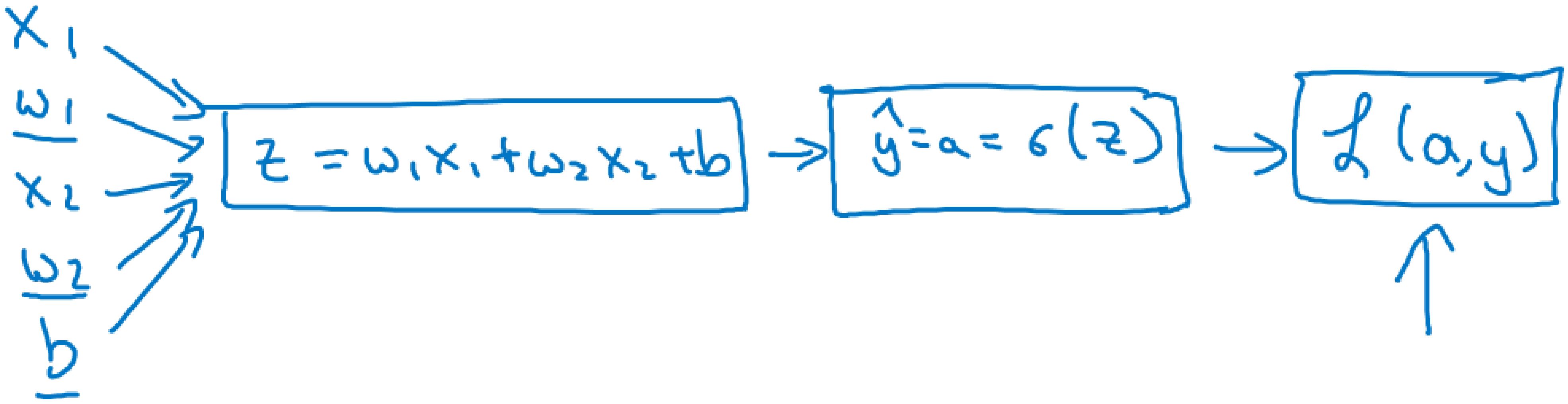
Logistic Regression Gradient descent

Logistic regression recap

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

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

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



Logistic regression derivatives

$$
\left.\begin{matrix} x_1 \\ w_1 \\ x_2 \\ w_2 \\ b \end{matrix}\right\} \rightarrow \boxed{z = w_1 x_1 + w_2 x_2 + b} \rightarrow \boxed{a = \sigma(z)} \rightarrow \boxed{\mathcal{L}(a, y)}
$$

$$
\boxed{dz} = \frac{d\mathcal{L}}{dz} = \frac{d\mathcal{L}(a,y)}{dz} = a - y = \boxed{\frac{d\mathcal{L}}{da}} \cdot \boxed{\frac{da}{dz}}
$$

$$
\text{"}da\text{"} = \frac{d\mathcal{L}(a,y)}{da} = \boxed{-\frac{y}{a} + \frac{1-y}{1-a}}
$$

$$
\boxed{a(1-a)}
$$

$$
\frac{\partial \mathcal{L}}{\partial w_1} = \text{"}dw_1\text{"} = x_1 \cdot dz. \qquad dw_2 = x_2\, dz. \qquad db = dz.
$$

$$
\begin{aligned}
w_1 &:= w_1 - \alpha\, dw_1 \\
w_2 &:= w_2 - \alpha\, dw_2 \\
b &:= b - \alpha\, db.
\end{aligned}
$$

Basics of Neural Network Programming

Gradient descent on m examples

deeplearning.ai

Logistic regression on m examples

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

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

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

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

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

Logistic regression on m examples

$J=0;\ \underline{dw_1}=0;\ \underline{dw_2}=0;\ \underline{db}=0$

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

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

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

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

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

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

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

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

$\underline{n}=2$

dw_3
$\vdots$
dw_n

$J /= m \leftarrow$

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

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

$$w_1 := w_1 - \alpha\, \underline{dw_1}$$

$$w_2 := w_2 - \alpha\, \underline{dw_2}$$

$$b := b - \alpha\, \underline{db}.$$

Vectorization

Andrew Ng