we define a notation:

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

2. $a^{(i)} = \sigma(z^{(i)}) = \dfrac{1}{1+e^{-z^{(i)}}} \quad (a^{(i)} = \Phi(x^{(i)}))$

3. $L(a^{(i)}, y^{(i)}) = -y^{(i)} \log a^{(i)} - (1-y^{(i)}) \log(1-a^{(i)})$

4. $C(w,b) = \sum_{i=1}^{m} L(a^{(i)}, y^{(i)})$

we know that:

$$\frac{\partial C(w,b)}{\partial w_j} = \sum_{i=1}^{m} \frac{\partial}{\partial w_j} L(a^{(i)}, y^{(i)})$$

so we use chain rule to calculate $\dfrac{\partial}{\partial w_j} L(a^{(i)}, y^{(i)})$:

$$\frac{\partial}{\partial w_j} L(a^{(i)}, y^{(i)}) = \frac{\partial L(a^{(i)}, y^{(i)})}{\partial a^{(i)}} \times \frac{\partial a^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial w_j}$$

1. $\dfrac{\partial L(a^{(i)}, y^{(i)})}{\partial a^{(i)}} = \boxed{\left(\dfrac{-y^{(i)}}{a^{(i)}} + \dfrac{1-y^{(i)}}{1-a^{(i)}}\right)\dfrac{1}{\ln 10}}$

2. $\dfrac{\partial a^{(i)}}{\partial z^{(i)}} = \dfrac{e^{-z^{(i)}}}{(1+e^{-z^{(i)}})^2} = \dfrac{1}{1+e^{-z^{(i)}}} \times \dfrac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}} =$

$$\frac{1}{1+e^{-z^{(i)}}} \times \left(\frac{1+e^{-z^{(i)}}-1}{1+e^{-z^{(i)}}}\right) = \frac{1}{1+e^{-z^{(i)}}} \times \left(1-\frac{1}{1+e^{-z^{(i)}}}\right)$$

$$= \boxed{a^{(i)}(1-a^{(i)})}$$

3.1. $\frac{\partial z^{(i)}}{\partial w_j} = \boxed{x_j^{(i)}}$ 3.2. $\frac{\partial z^{(i)}}{\partial b} = \boxed{1}$

so:

$$\frac{\partial}{\partial w_j} L(a^{(i)}, y^{(i)}) = \frac{1}{\ln 10}\left(\frac{-y^{(i)}}{a^{(i)}} + \frac{1-y^{(i)}}{1-a^{(i)}}\right)\left(a^{(i)}(1-a^{(i)})\right) x_j^{(i)}$$

$$= \frac{1}{\ln 10}\left(-y^{(i)}(1-a^{(i)}) + (1-y^{(i)})a^{(i)}\right) x_j^{(i)}$$

$$= \frac{1}{\ln 10}\left(-y^{(i)} + a^{(i)}y^{(i)} + a^{(i)} - a^{(i)}y^{(i)}\right) x_j^{(i)} = \boxed{\frac{1}{\ln(10)}(a^{(i)}-y^{(i)})x_j^{(i)}}$$

$$\frac{\partial}{\partial b} L(a^{(i)},y^{(i)}) = \boxed{\frac{1}{\ln 10}(a^{(i)}-y^{(i)})}$$

$$\frac{\partial C(w,b)}{\partial b} = \boxed{\sum_{i=1}^{m} \frac{1}{\ln 10}(a^{(i)}-y^{(i)})}$$

$$\frac{\partial C(w,b)}{\partial w_j} = \sum_{i=1}^{m} \frac{\partial}{\partial w_j} L(a^{(i)},y^{(i)}) = \frac{1}{\ln 10}\sum_{i=1}^{m} (a^{(i)}-y^{(i)})x_j^{(i)}$$

« Vectorization »

define: $X = \left[x^{(1)}\ x^{(2)} \dots x^{(m)}\right]$, $x^{(i)} \in R^n$

$W = \begin{bmatrix} w_1\\ w_2\\ \vdots\\ w_n\end{bmatrix}$, $w_i \in R$

$Y = \left[y^{(1)}\ y^{(2)} \dots y^{(n)}\right]$ $y^{(i)} \in R$

Subject:
Year: Month: Day: ( )

$$Z = [z^{(1)}\ z^{(2)}\ \dots\ z^{(m)}],\ z^{(i)} \in R$$

$$Z = W^T X + b$$

$$A = [a^{(1)}\ a^{(2)}\ \dots\ a^{(m)}] \quad A^{(i)} \in R$$

$$A = \sigma(Z) \longrightarrow \text{elementwise}$$

$$\frac{\partial}{\partial w_j} C(w,b) = \frac{1}{\ln 10} \sum_{i=1}^{m} (a^{(i)} - y^{(i)}) x_j^{(i)} = X[:,j](A-Y)^T$$

$$dW = \begin{bmatrix} \frac{\partial C(w,b)}{\partial w_1} \\ \frac{\partial C(w,b)}{\partial w_2} \\ \vdots \\ \frac{\partial C(w,b)}{\partial w_n} \end{bmatrix}$$

$$dW = \boxed{X(A-Y)^T}, \quad db = \sum_{i=1}^{m} (a^{(i)} - y^{(i)}) = \boxed{\mathrm{Sum}(A-Y)}$$

w, b, X