In this derivation, I use the convention: superscripts are rows and subscripts are columns. We have a matrix X, where each row corresponds to a sample, y is a column vector of target values, and θ is a column vector of weights.
To make it easier to read, let the hypothesis of the i^{th} sample of X be written h^i,

$$h^i = h_\theta(X^i) = \sigma(\theta^\top X^i).$$

$$h = \begin{bmatrix} \sigma(\theta^\top X^1) \\ \sigma(\theta^\top X^2) \\ \vdots \\ \sigma(\theta^\top X^n) \end{bmatrix},\ log(h) = \begin{bmatrix} log(h^1) \\ log(h^2) \\ \vdots \\ log(h^n) \end{bmatrix}$$

Cost Function

$$\begin{aligned} J(\theta) &= -\frac{1}{m}\sum_{i=1}^{m}\left[y^i log(h^i) + (1-y^i)log(1-h^i)\right] \\ &= -\frac{1}{m}\left[y \cdot log(h) + (1-y)\cdot h\right] && \text{dot product} \\ &= -\frac{1}{m}\left[y^\top log(h) + (1-y)^\top h\right] && \text{matrix form} \end{aligned}$$

Gradient of Cost Function
Notation, $\partial_k = \frac{\partial}{\partial\theta_k}$. The derivative of $\sigma(z) = \sigma(z)(1-\sigma(z))$, so that

$$\begin{aligned} \partial_k h^i &= \sigma(\theta^\top X^i)[1-\sigma(\theta^\top X^i)]X_k^i \\ &= h^i[1-h^i]X_k^i. \end{aligned}$$

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_k} &= -\frac{1}{m}\left[\sum_{i=1}^{m} y^i(\frac{1}{h^i})\partial_k h^i + (1-y^i)\frac{1}{1-h^i}\partial_k(1-h^i)\right] \\ &= -\frac{1}{m}\left[\sum_{i=1}^{m}\frac{y^i}{h^i}h^i(1-h^i)X_k^i + \frac{1-y^i}{1-h^i}(-h^i(1-h^i))X_k^i\right] \\ &= -\frac{1}{m}\left[\sum_{i=1}^{m} y^i(1-h^i) - (1-y^i)h^i\right]X_k^i \\ &= -\frac{1}{m}\sum_{i=1}^{m}[y^i - h^i]X_k^i \\ &= -\frac{1}{m}X^\top(y-h) && \text{matrix form} \end{aligned}$$