Due Oct 11, 2022 (Tuesday) by 11:59pm

Team can get together and discuss. But, each member is expected to write up their own derivations. Simply copying and pasting the work of other(s) is considered no effort.

Prob. Let $J(\theta)$ be the cost objective function of parameters $\theta = (\theta_0, \theta_1, \dots, \theta_n)$ in a hypothesis/regression model. It is to be minimized through learning. Specifically,

$$J(\theta) = -\frac{1}{m}\{\sum_{i=1}^{m} y^{(i)} \log\left[h_{(\theta)}\left(x^{(i)}\right)\right] + (1 - y^{(i)}) \log\left[1 - h_{(\theta)}\left(x^{(i)}\right)\right]\},$$

where

$h_{(\theta)}(x) = \frac{1}{1+e^{-\theta^T x}}$.

Derive the gradient of $J(\theta)$ along each parameter, or specifically

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots, \theta_n)$$

for j=0, 1, 2, …, n.

Hint: you may find this identify useful:

$$\frac{\partial h_\theta(x)}{\partial \theta_j} = \frac{1 + e^{-\theta^T x} - 1}{\left(1 + e^{-\theta^T x}\right)^2} . x_j = h_\theta(x)[1 - h_\theta(x)]x_j$$