

Due Sept 30, 2022 (Friday) by 11:59pm

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

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

where $h_{\theta}(x) = \theta^T x = \theta_0 + \theta_1 x$
and m is the number of training samples.

Derive the gradient of $J(\theta)$ along each parameter, i.e.,

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

for $j=0, 1$.