Solution:

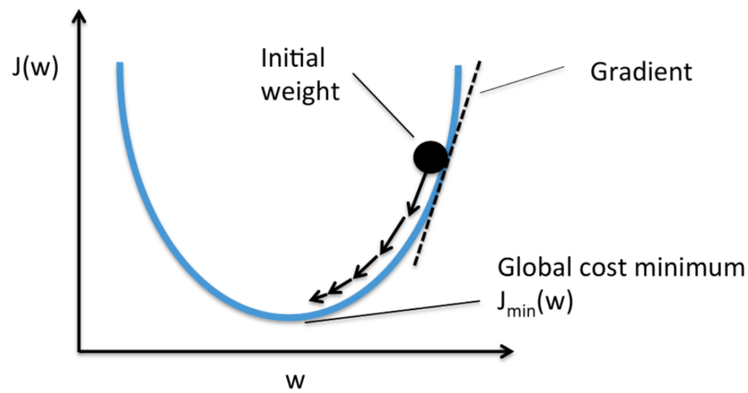
The hypothesis for linear regression model is usually represented as:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

The θvalues are the parameters which need to be optimised in order to minimise the value of MSE. MSE can also be defined as 'Cost Function' already mentioned in the question,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i' - y_i)^2$$

The better way to minimise the MSE is by using Gradient Descent. We start by initialising $\theta_0$and $\theta_1$to any two values say 0 for both. And follow the following algorithm:



$$\theta_j := \theta_j - \alpha\frac{\delta}{\delta\theta_j}J(\theta_0, \theta_1)$$

where α, is our learning rate, or you can say how quickly we want to move towards the minimum. If α is too large, however, we can overfit the model. Now, we have MSE as a cost function which we have to minimise using Gradient Descent. This gives us the value of θ_0 and θ_1 as follows:

- $\theta_0, j = 0: \frac{\delta}{\delta\theta_j}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})$
- $\theta_1, j = 1: \frac{\delta}{\delta\theta_j}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}).\ (X^{(i)})$