Gradient Descent Algorithm :-

Cost function :-

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \hat{\theta}_3 x_3 + \cdots + \hat{\theta}_k x_k$$

$$\hat{y} = \hat{\theta}_0 x_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \cdots + \hat{\theta}_k x_k$$

SL No	x_0	x_1	x_2	$\cdots$	x_k
1.	1	$x_1^{(1)}$	$x_2^{(1)}$	$\cdots$	$x_k^{(1)}$
2.	1	$x_1^{(2)}$	$x_2^{(2)}$	$\cdots$	$x_k^{(2)}$
3.	1	$x_1^{(3)}$	$x_2^{(3)}$	$\cdots$	$x_k^{(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m-1	1	$x_1^{(m-1)}$	$x_2^{(m-1)}$	$\cdots$	$x_k^{(m-1)}$
m	1	$x_1^{(m)}$	$x_2^{(m)}$	$\cdots$	$x_k^{(m)}$

$x_0 = 1$ for all rows.

for any row i

$$\left[\overset{1}{\cancel{x_0^{(i)}}},\ x_1^{(i)},\ x_2^{(i)},\ \cdots,\ x_k^{(i)} \right]$$

$$\vec{x}^{(i)} = \left[1,\ x_1^{(i)},\ x_2^{(i)},\ \cdots,\ x_k^{(i)} \right]$$

Sl No.	x_0	x_1	x_2	$\cdots$	x_K	y
1.	1	$x_1^{(1)}$	$x_2^{(1)}$	$\cdots$	$x_K^{(1)}$	$y^{(1)}$
2.	1	$x_1^{(2)}$	$x_2^{(2)}$	$\cdots$	$x_K^{(2)}$	$y^{(2)}$
3.	1	$x_1^{(3)}$	$x_2^{(3)}$	$\cdots$	$x_K^{(3)}$	$y^{(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m.	1	$x_1^{(m)}$	$x_2^{(m)}$	$\cdots$	$x_K^{(m)}$	$y^{(m)}$

target.

$m \rightarrow$ total number of training samples.

$K \rightarrow$ number of features.

$$\hat{y}^{(i)} = \hat{\theta}_0 x_0^{(i)} + \hat{\theta}_1 x_1^{(i)} + \hat{\theta}_2 x_2^{(i)} + \cdots + \hat{\theta}_K x_K^{(i)} = \sum_{j=0}^{K} \hat{\theta}_j \cdot x_j^{(i)}$$

$$= \left[\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K\right] \cdot \left[x_0^{(i)}, x_1^{(i)}, x_2^{(i)}, \ldots, x_K^{(i)}\right]$$

$$\boxed{\hat{y}^{(i)} = \vec{\hat{\theta}} \cdot \vec{x}^{(i)}}$$

$\vec{\hat{\theta}} \rightarrow$ Estimated model parameter vector.

$\vec{x}^{(i)} \rightarrow$ i^{th} observation (datapoint) vector.

$\hat{y}^{(i)} \rightarrow$ Predicted value of $y^{(i)}$. $\rightarrow$ Actual value.

$e^{(i)} = \hat{y}^{(i)} - y^{(i)} \longrightarrow$ Error for i^{th} observati

Residual Sum of Square (RSS)

$$RSS = \sum_{i=1}^{m} \left[e^{(i)}\right]^2 = \sum_{i=1}^{m} \left[\hat{y}^{(i)} - y^{(i)}\right]^2$$

Cost function:- $J = \frac{1}{2m}(RSS)$.

$$\therefore J = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{\hat{y}^{(i)}}_{\sum_{j=0}^{k} \hat{\theta}_j x_j^{(i)}} - y^{(i)}\Big)^2 = \frac{1}{2m} \sum_{i=1}^{m} \left(\sum_{j=0}^{k} \hat{\theta}_j x_j^{(i)} - y^{(i)}\right)^2$$

$$J = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{\sum_{j=0}^{k} \hat{\theta}_j \cdot x_j^{(i)} - y^{(i)}}_{z^{(i)}} \Big)^2$$

$$J = \frac{1}{2m} \sum_{i=1}^{m} \left(z^{(i)}\right)^2$$

$$z^{(i)} = \underbrace{\sum_{j=0}^{k} \hat{\theta}_j \cdot x_j^{(i)}}_{\hat{y}^{(i)}} - y^{(i)}$$

$$\frac{\partial J}{\partial \hat{\theta}_j} = \frac{\partial J}{\partial z^{(i)}} \cdot \frac{\partial z^{(i)}}{\partial \hat{\theta}_j}$$

$$\frac{\partial z^{(i)}}{\partial \hat{\theta}_j} = x_j^{(i)}$$

$$= \frac{1}{m} \sum_{i=1}^{m} z^{(i)} \cdot x_j^{(i)} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) x_j^{(i)}$$

$$\Delta\hat{\theta}_j = -\alpha \frac{\partial J}{\partial \hat{\theta}_j} \quad \& \quad \hat{\theta}_j(t+1) = \hat{\theta}_j(t) + \Delta\hat{\theta}_j$$

Gradient descent update rule.

$$\underline{\underline{\Delta\hat{\theta}_j}} = -\frac{\alpha}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right) \underline{\underline{x_j^{(i)}}}$$

$$\|\Delta\hat{\theta}_j\|$$

$$\hat{y}^{(i)} = \vec{\hat{\theta}} \cdot \vec{x}^{(i)}$$

$$\vec{\hat{\theta}}_{k\times 1} \qquad \vec{\hat{\theta}}_{1\times k}$$

$$X\_train = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & & x_k^{(2)} \\ \vdots & \vdots & & \vdots \\ x_1^{(i)} & x_2^{(i)} & & x_k^{(i)} \\ \vdots & \vdots & & \vdots \\ x_1^{(m)} & x_2^{(m)} & & x_k^{(m)} \end{bmatrix}_{m\times k} \begin{array}{l} \to \vec{x}^{(1)} \\ \\ \\ \to \vec{x}^{(i)} \\ \\ \\ \end{array}$$

$$X\_train = \begin{bmatrix} \vec{x}^{(1)} \\ \vec{x}^{(2)} \\ \vdots \\ \vec{x}^{(i)} \\ \vdots \\ \vec{x}^{(m)} \end{bmatrix}_{m\times k}$$

$$\vec{\hat{y}} = \vec{\hat{\theta}}_{1\times k} \cdot (X\_train)^T$$

$$(1\times \cancel{k}) \times (\cancel{k}\times m)$$

$$1\times m$$