MULTIVARIATE LINEAR REGRESSION

Multivariate Regression :

Linear Regression with multiple variables is also known as multivariate regression.

Consider our same example in Univariate Linear Regression, but in here we have additional features along with size($feet^2$).

Size in sq. Ft (X_1)	Number of bedrooms (X_2)	Number of Floors (X_3)	Age of house(years) (X_4)	Price($) in 1000's (y)
2104	**5**	**1**	**45**	**460**
1416	**3**	**2**	**40**	**232**
1534	**3**	**2**	**30**	**315**
852	**2**	**1**	**36**	**178**
..	..	..	..	..
..	..	..	..	..

Notation:

m → number of data samples.

n → number of features.

x^i → input features of i^{th} training example(vector of n-dimension).

$X_j^{(i)}$ →value of feature j of i^{th} training example.

Hypothesis :

$$h_\theta(\mathrm{x}) = \theta_0 + \theta_1 \mathrm{x1} + \theta_2 x2 + \theta_3 x3 + \theta_4 x4$$

consider X_0 =1 then our hypothesis becomes,

$$h_\theta(\mathrm{x}) = \theta_0 \,\mathrm{x0} + \theta_1 \mathrm{x1} + \theta_2 x2 + \theta_3 x3 + \theta_4 x4$$

and the size of x^i becomes (n+1).

Now we can perform a matrix multiplication to get the above equation.

$$h_\theta(\mathrm{x}) = \theta^T X$$

Gradient Descent for Multiple Variables :

Cost function is defined as

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

$$= \frac{1}{2m} * \sum_{i=1}^{m} (\theta^T(x^i) - y^i)^2$$

$$= \frac{1}{2m} * \sum_{i=1}^{m} ((\sum_{j=0}^{n} \theta_j(x_j^i)) - y^i)^2$$

Gradient descent :

Repeat until converges {

$$\Theta\mathrm{j} = \Theta\mathrm{j} - \alpha \frac{\partial \mathrm{J}}{\partial \theta} J(\theta_{0,} \theta_1, \dots \dots, \theta_n)$$

}

$$\frac{\partial J}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \cdot \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \frac{1}{2m} \cdot \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^i - y^i\right)^2$$

$$= \frac{\partial}{\partial \theta_0} \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \cancel{2}\left(\theta_0 + \theta_1 x^i - y^i\right) \frac{\partial}{\partial \theta_0}\left(\theta_0 + \theta_1 x^i - y^i\right)$$

$$\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)(1)$$

$$\frac{\partial J}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x^i - y^i\right)^2$$

$$= \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \cancel{2}\left(\theta_0 + \theta_1 x^i - y^i\right) \frac{\partial}{\partial \theta_1} \left(\theta_0 + \theta_1 x^i - y^i\right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)\left(0 + \frac{\partial}{\partial \theta_1}(\theta_1 x^i) + 0\right)$$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)(x^i)$$

Gradient Descent

Previously ($n=1$):

Repeat {

$$\theta_0 := \theta_0 - \alpha \underbrace{\boxed{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)}}_{\frac{\partial}{\partial \theta_0} J(\theta)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$x_1^{(i)}$

(simultaneously update θ_0, θ_1)

}

New algorithm ($n \geq 1$):

Repeat {

$\frac{\partial}{\partial \theta_j} J(\theta)$

$$\theta_j := \theta_j - \alpha \boxed{\frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_j^{(i)}}$$

(simultaneously update θ_j for $j = 0, \ldots, n$)

}

$x_0^{(i)} = 1$

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_0^{(i)}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_2^{(i)}$$

...

Repeat until converges

{

$$\Theta j = \Theta j - \alpha \frac{1}{m} * \sum_{i=1}^{m} (h_{theta}(x^i) - y^i)\, x_j^i$$

}