Linear Regression

Aaryan
CO21BTECH11001

Let number of features of dataset = n

Let number of sets of features = m

Data consists of matrices X and y where i^{th} column of X represents the i^{th} feature of dataset and i^{th} element of y represents the value of variable dependent on set of features listed in i^{th} row of X.

Linear regression is a type of regression algorithm where it assumes a linear relationship between dependent (X) and independent (y) variables.

Let $X^{(i)} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ . \\ . \\ . \\ X_n \end{bmatrix} \quad and \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ . \\ . \\ \theta_n \end{bmatrix}$

where θ is known as parameter.

We define a hypothesis function $h_\theta(x)$ as follows –

$$h_\theta\left(X^{(i)}\right) = \theta^T X^{(i)} = \sum_{j=0}^{n} \theta_j X_j^{(i)}$$

where $X_0^{(i)} = 1$

We will calculate a value of θ which best fits the approximation –

$$h_\theta\left(X^{(i)}\right) \approx y_i$$

To do this, we will define a cost function $J(\theta)$ as follows –

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta X^{(i)} - y_i\right)^2$$

We can see from here that when $J(\theta) \to 0$, our assumption is satisfied.

Objective – Minimize or Converge the cost function.

There are two approaches to do this –

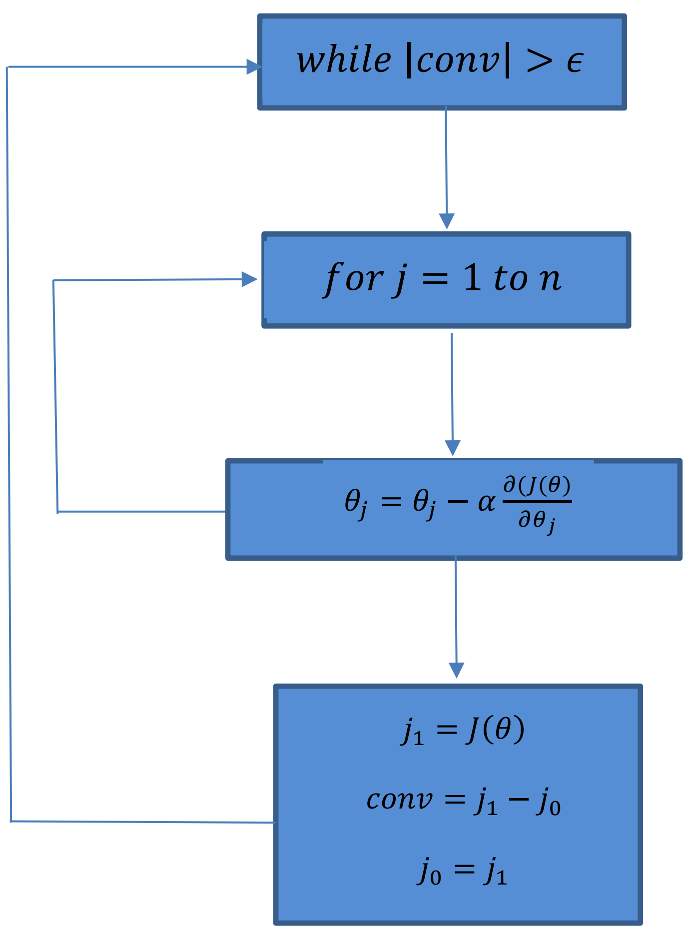
Gradient Descent Algorithm –

$$Convergence\ limit = \epsilon = 10^{-10}$$

$$Initialize\ \theta = \vec{0}$$

$$j_0 = J(\theta)$$

$$conv = \infty$$



Normal Equation Method –

$$\nabla_\theta J(\theta) = \begin{bmatrix} \frac{\partial J}{\partial \theta_0} \\ \frac{\partial J}{\partial \theta_1} \\ \frac{\partial J}{\partial \theta_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial J}{\partial \theta_n} \end{bmatrix}$$

If $\nabla_\theta J(\theta) = 0$ then $J(\theta)$ will get minimized.

Solving for that, we get –

$$\theta = (X^T X)^{-1} X^T y$$

After getting optimal θ, we can get the value corresponding to a new data D as

$$Val = h_\theta(D) = \theta^T D$$

Questions –

1. Is Linear Regression a regression algorithm or classification algorithm?
   Ans. Regression algorithm
2. Why do we need to take $X_0^{(i)} = 1 \; \forall \; i$?
   Ans. Because in the hypothesis function there is a constant term apart from the linear combination of $X^{(i)}$ and θ, which is θ_0, so the multiplier of θ_0 can be any value. For simplicity, we take it as 1.

3. How can we increase the accuracy of linear regression?
   Ans. We can increase the accuracy of linear regression by outlier treatment, i.e. we have to remove the values of sparse feature which are usually less than 1 percentile or more than 99 percentile.
4. What are the disadvantages of linear regression model?
   Ans. It is sensitive and dependent on outliers, which effect the overall accuracy of the model.
5. What are the fields where linear regression is usually used?
   Ans. It is usually used in Businesses, Statistics, Medical Science etc.