Linear Regression

Abstract

This report consists of my basic understanding of one of the popular Ml methods ”Linear Regression”.

1 Linear Regression

Linear Regression is an old and standard techniques to understand and predict data not only in machine learning but also in statistics.Over the years this method is thoroughly studied and have different flavours of it.

2 What it will do?

Basically it is one of the techniques developed to solve a regression problem,i.e,the output that we want to predict is a real number.As the name suggests, linear regression assumes a **linear relationship** between the input and output parameters.It means that we are sure that $y(output\,variable)$ can be calculated by linear combiantion of input variables($\mathbf{x}$).The following is the key equation of linear regression,

$$y = \mathbf{w}^{\mathbf{T}}\mathbf{x} + c \tag{2.1}$$

where,$\mathbf{w}$ is known as weight vector and c is known as bias.

3 What does the equation signifies?

To understand that, we will take an example, let say we want to predict height of a child w.r.t the age of the child. In this case, the output variable(y) here is height and the input variable(x) is age.Since we have only one input parameter, the method is known as **Simple linear Regression**. Now from 3.1, we can relate x and y as,

$$y - 75 = \frac{90 - 75}{4 - 2}(x - 2) \tag{3.1}$$

$$y = \frac{25}{2}x + 100 \tag{3.2}$$

Here we are able to find a linear relationship between age and height with $w = 7.5$ as coefficient/weight of age and $c = 120cm$ as bias. And with this equation we will generally predict the height using age as a parameter.

But not always things are simple,we may have more input variables which can alter the value of y as in here we can take gender also as an parameter.And also we can’t say for

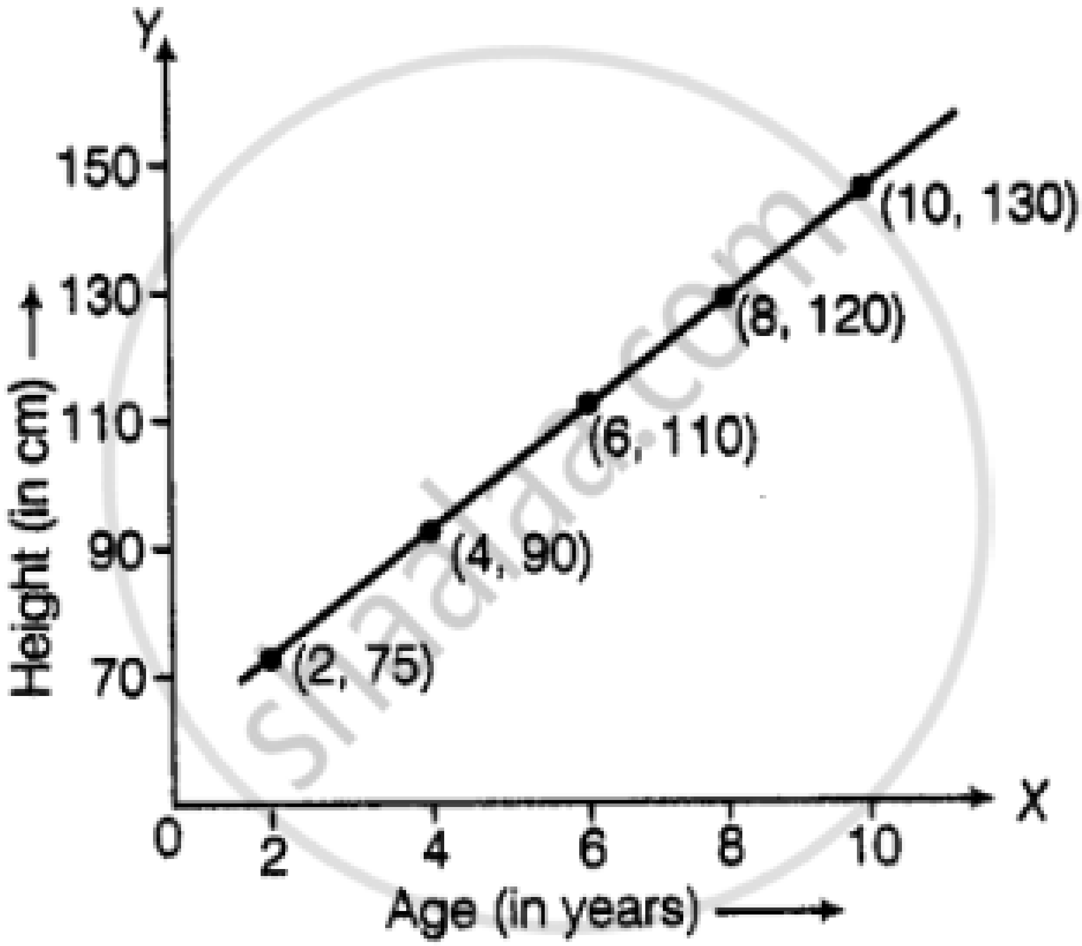

Fig. 3.1: Height vs Age

sure that the input variables and output variable linearly dependent. The task of linear regression is to find a best line for the given data which suits the training data and as well as testing data.By this we can understand that the weights and bias play a major role in predicting the output.

4 The Cost Function

As we discussed earlier that the output that we predict may not match the actual output.In that case, we want modify the weights so that the output that we predict is as close as possible.For that we should minimize the error in predicting the output.In general,for linear regression we will use the below cost function,

$$J = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \left(w^T X_i + c\right)\right)^2 \tag{4.1}$$

where, N are total pairs of (X_i, y_i) and $\mathbf{w}$ be the weight vector and c is bias. Now the cost function is a function/value that we want minimize in a model.So to minimize this cost function we will follow a standard method known as ”Gradient Descent”.

5 Gradient Descent

Before going to the gradient descent method, first we will see what actually gradient will do. Let us take a function f, the gradient of f

$$\nabla f = \frac{\partial f}{\partial x}\hat{\imath} + \frac{\partial f}{\partial y}\hat{\jmath} + \frac{\partial f}{\partial k}\hat{z} \tag{5.1}$$

will give the direction in which f will increase.
So we will modify the weights and bias in such a way that it will minimize the cost function, i.e.,

$$w_k = w_k - \alpha\frac{\partial J(w)}{\partial w_k} \tag{5.2}$$

$$c_k = c_k - \alpha\frac{\partial J(c)}{\partial c_k} \tag{5.3}$$

where w_k and c_k be the weight vector and bias after k^{th} iteration and α be the learning rate or explicitly it is distance we want to move the parameter in each iteration in order to minimize the cost function.

6 Regularization

In this process, we want minimize the mean square error and absolute sum of squares of weights and bias. This practice is carried out to avoid overfitting.As we assumed that there linear dependency between input parameters and the output,in search of best line the model may fail to generalize the unseen data.The new cost function looks like,

$$J = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \left(w^T X_i + c\right)\right)^2 + ||\beta|| \tag{6.1}$$

where $\beta = \begin{pmatrix}\theta \\ w\end{pmatrix}$

7 Summary

In this report we discussed,

1) What is linear Regression?
2) How does linear Regression works?
3) Cost function of linear regression
4) Gradient Descent Method
5) Regularization

8 Questions:

1) What is the basic assumption made in Linear Regression?
2) What is the general cost function used in linear Regression?
3) What parameter decides the size of improvement in each step in gradient descent method?
4) Why we should do regularisation?
5) What are the disadvantages of linear Regression?

9 Solutions:

1) The basic assumption made in linear regression is that there is linear relationship between dependent and independent variables.
2) The cost function used in linear regression is known as,Root Mean Squared Error which is given by

$$J = \frac{1}{2N} \sum_{i=1}^{N} (y_i - y_i pred)^2 \tag{9.1}$$

   where,
   - N total number data samples
   - y_i - actual label of i^{th} input vector.
   - $y_i pred$ - predicted label of i^{th} input vector
3) The learning rate(α) decides size of improvement in each step in gradient descent method.
4) To avoid overfitting, we should do regulariation. It penalizes the higher magnitude coefficients.
5) The disadvantages of linear Regression are
   a) The assumption of linear relationship between dependent and independent variables.
   b) Quite sensitive to outliers and causes overfitting