# Linear Regression and Gradient Descent

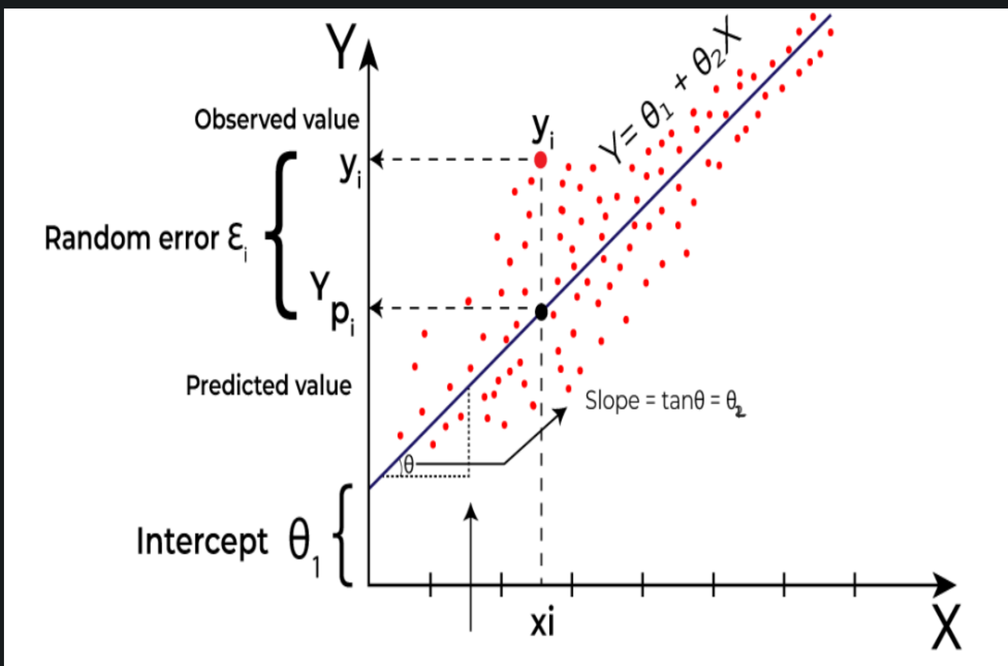
The goal of the Linear regression algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.

In regression set of records are present with X and Y values and these values are used to learn a function so if you want to predict Y from an unknown X this learned function can be used. In regression we have to find the value of Y, So, a function is required that predicts continuous Y in the case of regression given X as independent features.

### Best Fit Line

Our primary objective while using linear regression is to locate the best-fit line, which implies that the error between the predicted and actual values should be kept to a minimum. There will be the least error in the best-fit line. The best fit line is determined by finding the coefficients ( $\beta_0$ and $\beta_1$) that minimize the sum of squared differences between the observed values of the dependent variable and the values predicted by the regression equation. This process is often achieved through the method of least squares.

The best Fit Line equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable(s).



### Use of Best Fit Line

**Prediction:** Once the best fit line is established, it can be used to make predictions. Given a value of the independent variable(s), the equation allows you to estimate the corresponding value of the dependent variable. This is particularly useful for forecasting and understanding trends in data.

**Understanding Relationships:** The slope ($\beta_1$) of the best fit line indicates the strength and direction of the relationship between the variables. A positive slope suggests a positive correlation, while a negative slope indicates a negative correlation.

**Visual Representation:** The best fit line is often plotted on a scatterplot of the data points. It visually represents the linear trend in the data and helps assess how well the model fits the observed data.

### Cost Function of Linear Regression

In regression, the difference between the observed value of the dependent variable(yi) and the predicted value(predicted) is called the residuals.

$\varepsilon_i = y_{predicted} – y^i = \hat{Y} - y_i$

where $y_{predicted} = \hat{Y} = \Theta_1 + \Theta_2$ Xi

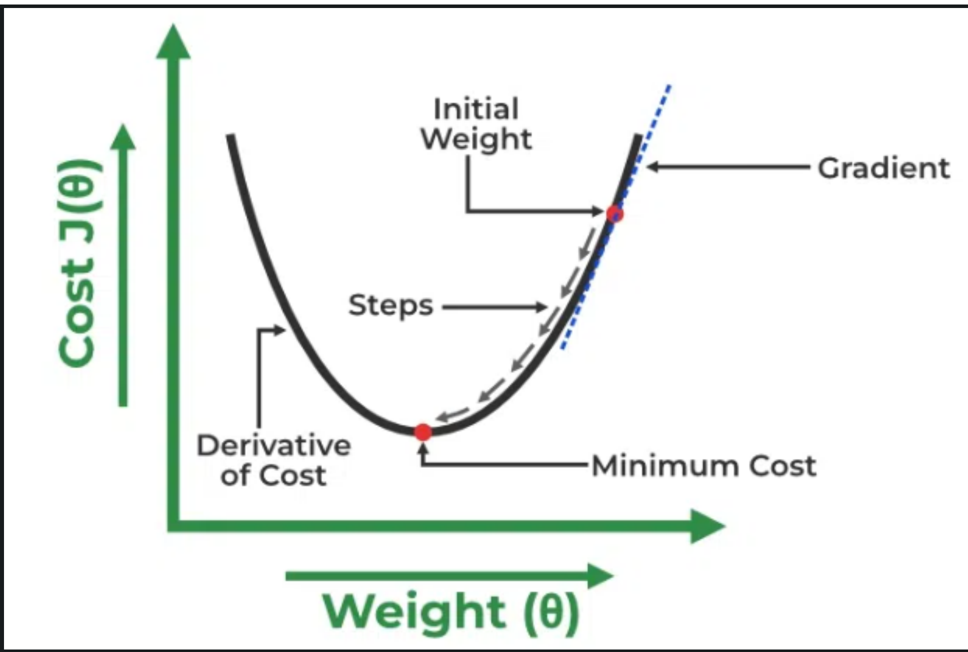
The cost function helps to work out the optimal values for B0 and B1, which provides the best fit line for the data points.

In Linear Regression, generally Mean Squared Error (MSE) cost function is used, which is the average of squared error that occurred between the $y_{predicted}$ and yi. This is the specified cost function.

$$minimize \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

### Gradient Descent

A regression model optimizes the gradient descent algorithm to update the coefficients of the line by reducing the cost function by randomly selecting coefficient values and then iteratively updating the values to reach the minimum cost function.




To update $\Theta_1$ and $\Theta_2$ values in order to reduce the Cost function (minimizing MSE value) and achieve the best-fit line the model uses Gradient Descent.

**$\theta_1$:** intercept
**$\theta_2$:** coefficient of x

Let's differentiate the cost function(J) with respect to Θ1 :

$$\begin{aligned}
J'_{\theta_1} &= \frac{\partial J(\theta_1, \theta_2)}{\partial \theta_1} \\
&= \frac{\partial}{\partial \theta_1}\left[\frac{1}{n}\left(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial \theta_1}(\hat{y}_i - y_i)\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial \theta_1}(\theta_1 + \theta_2 x_i - y_i)\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\,(1 + 0 - 0)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)\,(2)\right] \\
&= \frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)
\end{aligned}$$

$$
\begin{aligned}
J'_{\theta_2} &= \frac{\partial J(\theta_1, \theta_2)}{\partial \theta_2} \\
&= \frac{\partial}{\partial \theta_2}\left[\frac{1}{n}\left(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial \theta_2}(\hat{y}_i - y_i)\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial \theta_2}(\theta_1 + \theta_2 x_i - y_i)\right)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\,(0 + x_i - 0)\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)\,(2x_i)\right] \\
&= \frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)\cdot x_i
\end{aligned}
$$

Now for the new best fit line the new ϴ1 and ϴ2 are :

$$
\begin{aligned}
\theta_1 &= \theta_1 - \alpha\left(J'_{\theta_1}\right) \\
&= \theta_1 - \alpha\left(\frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)\right) \\
\theta_2 &= \theta_2 - \alpha\left(J'_{\theta_2}\right) \\
&= \theta_2 - \alpha\left(\frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)\cdot x_i\right)
\end{aligned}
$$

where, alpha is the Learning rate

[redacted]