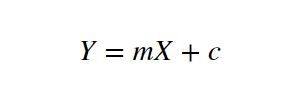
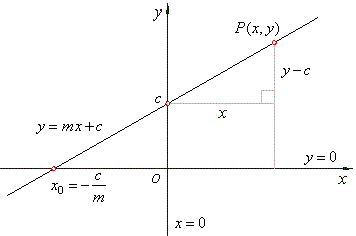
**Linear Regression**

In statistics, linear regression is a linear approach to modelling the relationship between a dependent variable and one or more independent variables.

Let **X** be the independent variable and **Y** be the dependent variable. We will define a linear relationship between these two variables as follows:





* Here **m** is the slope of the line
* **c** is the y intercept.

we will use this equation to train our model with a given dataset and predict the value of **Y** for any given value of **X**. Our challenge is to determine the value of **m** and **c**, such that the line corresponding to those values is the best fitting line or gives the minimum error.

Difference between cost function and loss function:

A **Cost Function/Loss Function** evaluates the performance of our Machine Learning Algorithm. The**Loss function** computes the error for a single training example while the **Cost functi**on is the average of the loss functions for all the training examples. Henceforth, I shall be using both the terms interchangeably.

**“A Cost function basically tells us ‘ how good’ our model is at making predictions for a given value of m and c.”**

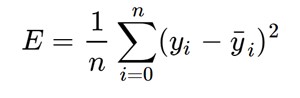
**Loss Function:**

The loss is the error in our predicted value of **m** and **c**. Our goal is to minimize this error to obtain the most accurate value of **m** and **c**.  
We will use the Mean Squared Error function to calculate the loss.

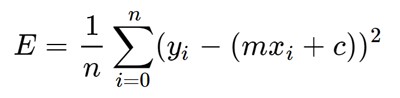
There are three steps in this function:

* Find the difference between the actual y and predicted y value(y = mx + c), for a given x.
* Square this difference.
* Find the mean of the squares for every value in X.

Mean Squared Error Equation is



Here yᵢ is the actual value and ȳᵢ is the predicted value. Let’s substitute the value of ȳᵢ:



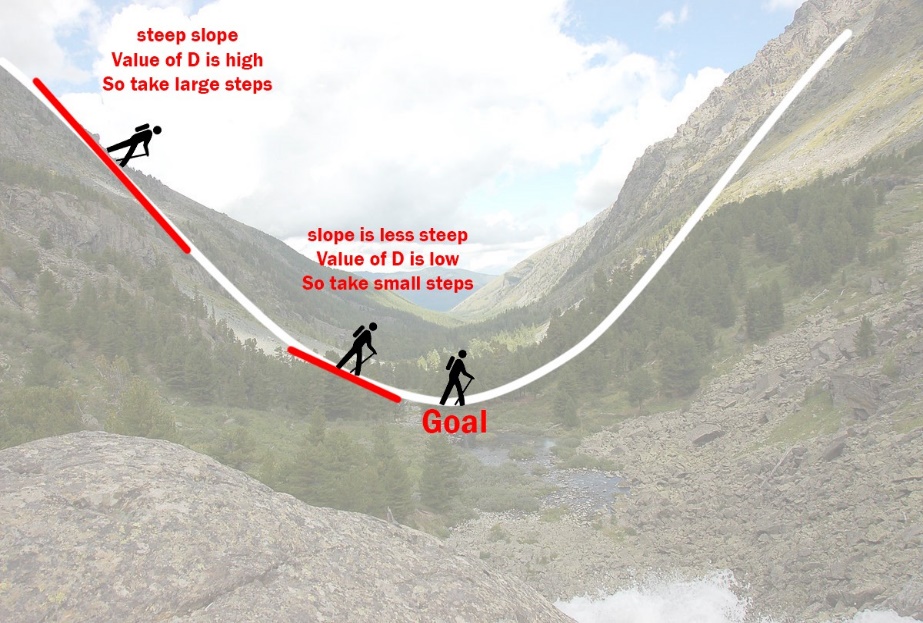
Substituting the value of ȳᵢ

So we square the error and find the mean. hence the name Mean Squared Error. Now that we have defined the loss function, lets get into the interesting part — minimizing it and finding **m** and **c.**

**The Gradient Descent Algorithm**

Gradient descent is an iterative optimization algorithm to find the minimum of a function. Here that function is our Loss Function.

**Understanding Gradient Descent**

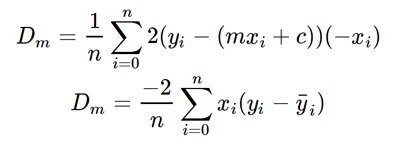


Example: Imagine a valley and a person with no sense of direction who wants to get to the bottom of the valley. He goes down the slope and takes large steps when the slope is steep and small steps when the slope is less steep. He decides his next position based on his current position and stops when he gets to the bottom of the valley which was his goal.

I got one other example from our team is how an aeroplane lands.

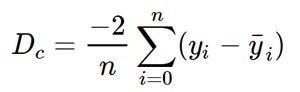
Let’s try applying gradient descent to **m** and **c** and approach it step by step:

1. Initially let m = 0 and c = 0. Let L be our learning rate. This controls how much the value of **m** changes with each step. L could be a small value like 0.0001 for good accuracy.
2. Calculate the partial derivative of the loss function with respect to m, and plug in the current values of x, y, m and c in it to obtain the derivative value **D**.



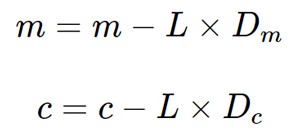
Derivative with respect to **m**

Dₘ is the value of the partial derivative with respect to **m**. Similarly lets find the partial derivative with respect to **c**, Dc :



Derivative with respect to **c**

3. Now we update the current value of **m** and **c** using the following equation:



4. We repeat this process until our loss function is a very small value or ideally 0 (which means 0 error or 100% accuracy). The value of **m** and **c** that we are left with now will be the optimum values.

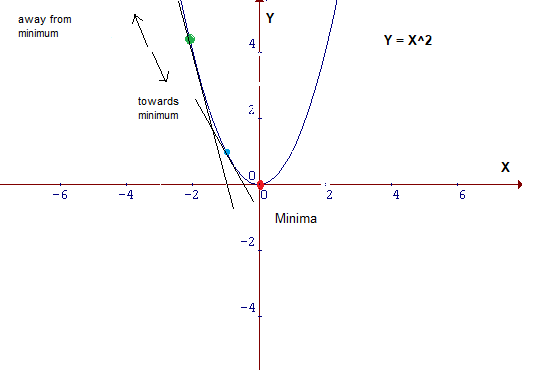
Now going back to our analogy,

* **m** can be considered the current position of the person.
* **D** is equivalent to the steepness of the slope
* **L** can be the speed with which he moves.

Now the new value of **m** that we calculate using the above equation will be his next position, and **L×D** will be the size of the steps he will take. When the slope is more steep (**D** is more) he takes longer steps and when it is less steep (**D** is less), he takes smaller steps.

Finally he arrives at the bottom of the valley which corresponds to our loss = 0.  
Now with the optimum value of **m** and **c** our model is ready to make predictions.

## In another way of explanation to find the Minimum Value or Minima is:



To find the minimum value, if we draw a tangent at the green point, we know that if we are moving upwards, we are moving away from the minima and vice versa. Also, the tangent gives us a sense of the steepness of the slope.

The slope at the blue point is less steep than that at the green point which means it will take much smaller **steps**to reach the minimum from the blue point than from the green point.