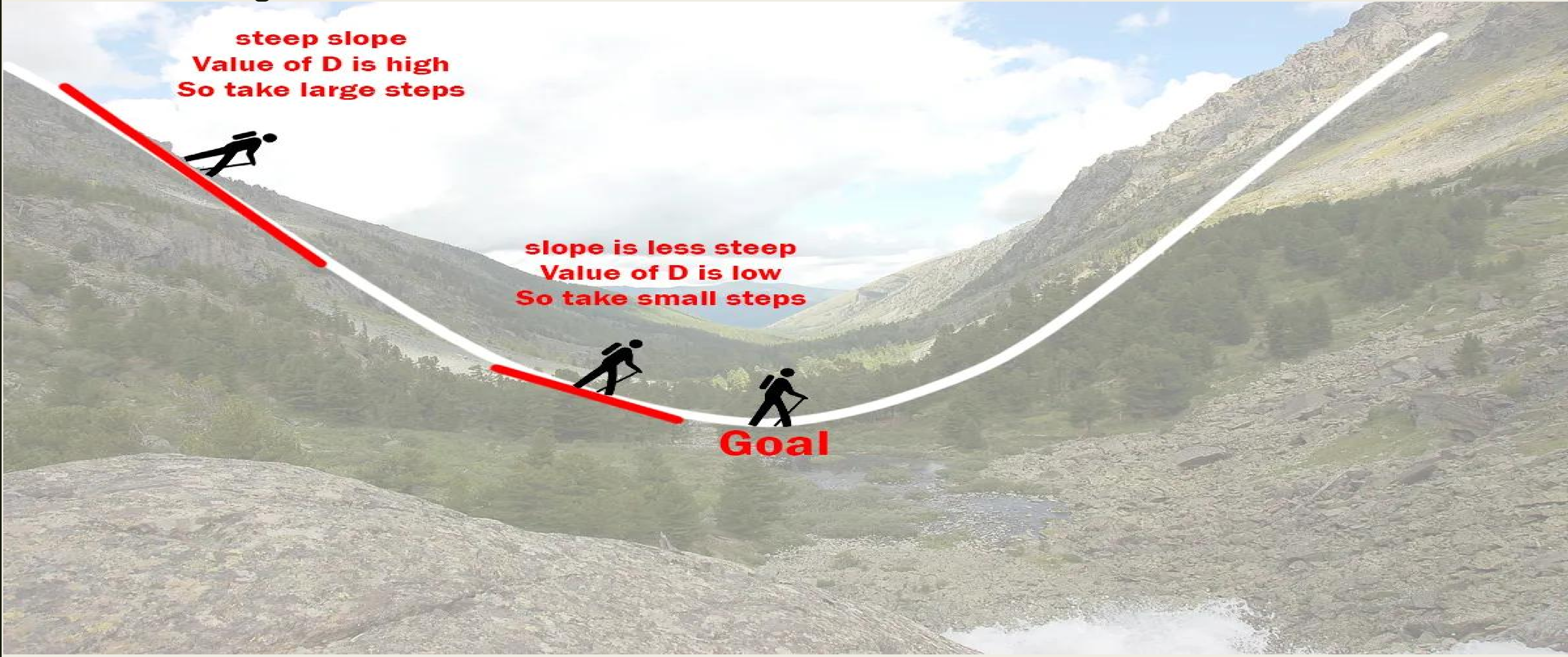
ML
LECTURE-4

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

Linear Regression using Gradient Descent

- **Gradient descent is an iterative optimization algorithm to find the minimum of a function.** Here that function is our Loss Function. We will use **Mean Square Error (MSE)** as Loss Function in this topic which is shown below:
- $\mathbf{E} = \frac{1}{n}\sum_{i=1}^{n}[y_i - (a + bx_i)]^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- Understanding Gradient Descent



Linear Regression using Gradient Descent

- **Mathematical derivation of Gradient Descent in simple Linear Regression :**
- 1. Initially let **a = 0** and **b = 0**. Let **L** be our learning rate. This controls how much the value of **b** 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 **a** and **b**, and plug in the current values of **x, y, b** and **a** in it to obtain the derivative value D.
- $D_b = 2\frac{1}{n}\sum_{i=1}^{n}[y_i - (a + bx_i)](-x_i)$
- $\mathbf{D_b} = \frac{-2}{n}\sum_{i=1}^{n} x_i[y_i - (a + bx_i)]$
- $\mathbf{D_b} = \frac{-2}{n}\sum_{i=1}^{n} x_i(y_i - \hat{y}_i)$
- $\mathbf{D_b}$ is the value of the partial derivative with respect to **b.**

- Similarly, the partial derivative with respect to **a** is $\mathbf{D_a}$**:**
- $D_a = 2\frac{1}{n}\sum_{i=1}^{n}[y_i - (a + bx_i)](-1)$
- $\mathbf{D_a} = \frac{-2}{n}\sum_{i=1}^{n}[y_i - (a + bx_i)]$
- $\mathbf{D_a} = \frac{-2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$

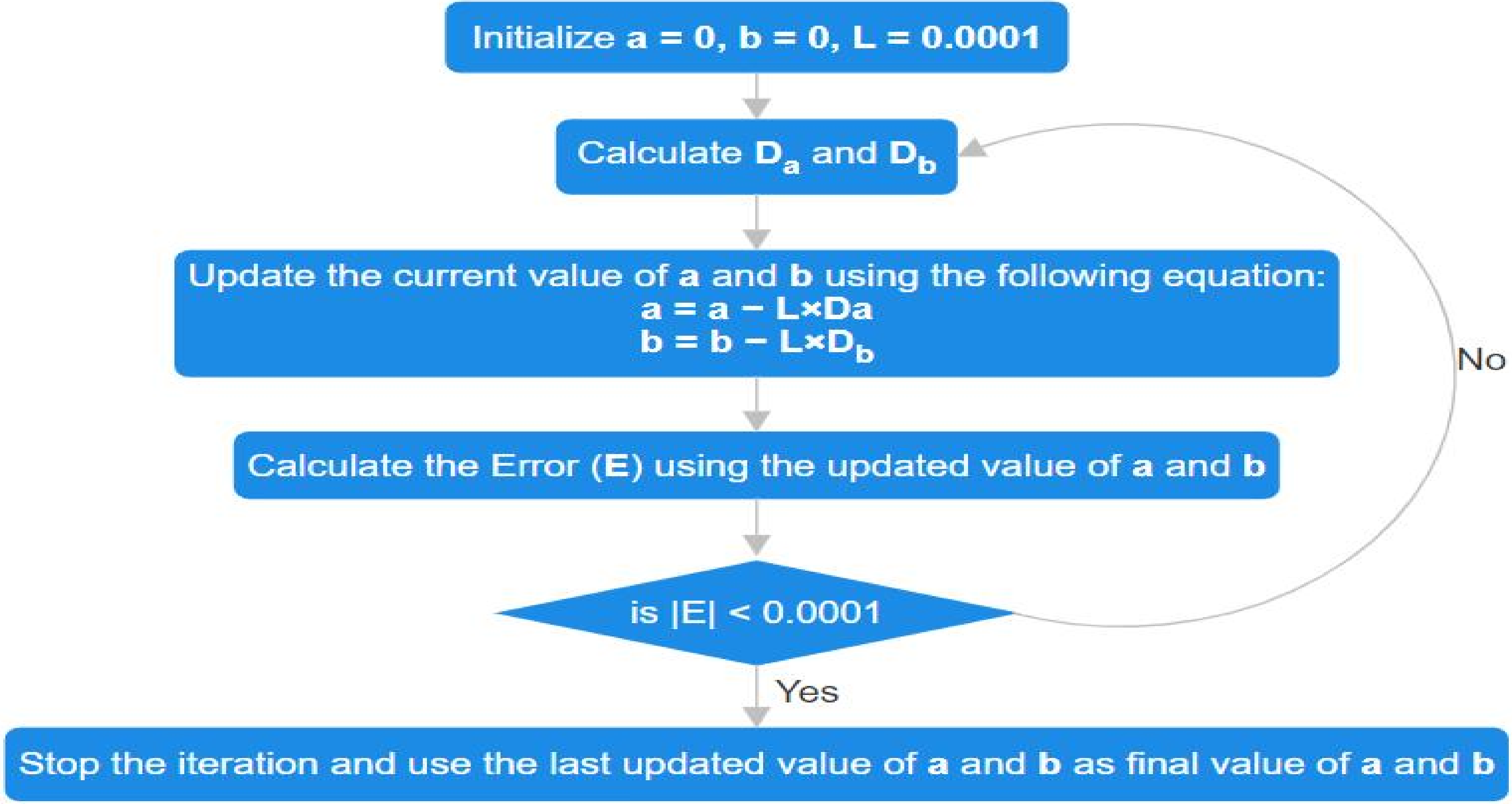
Linear Regression using Gradient Descent

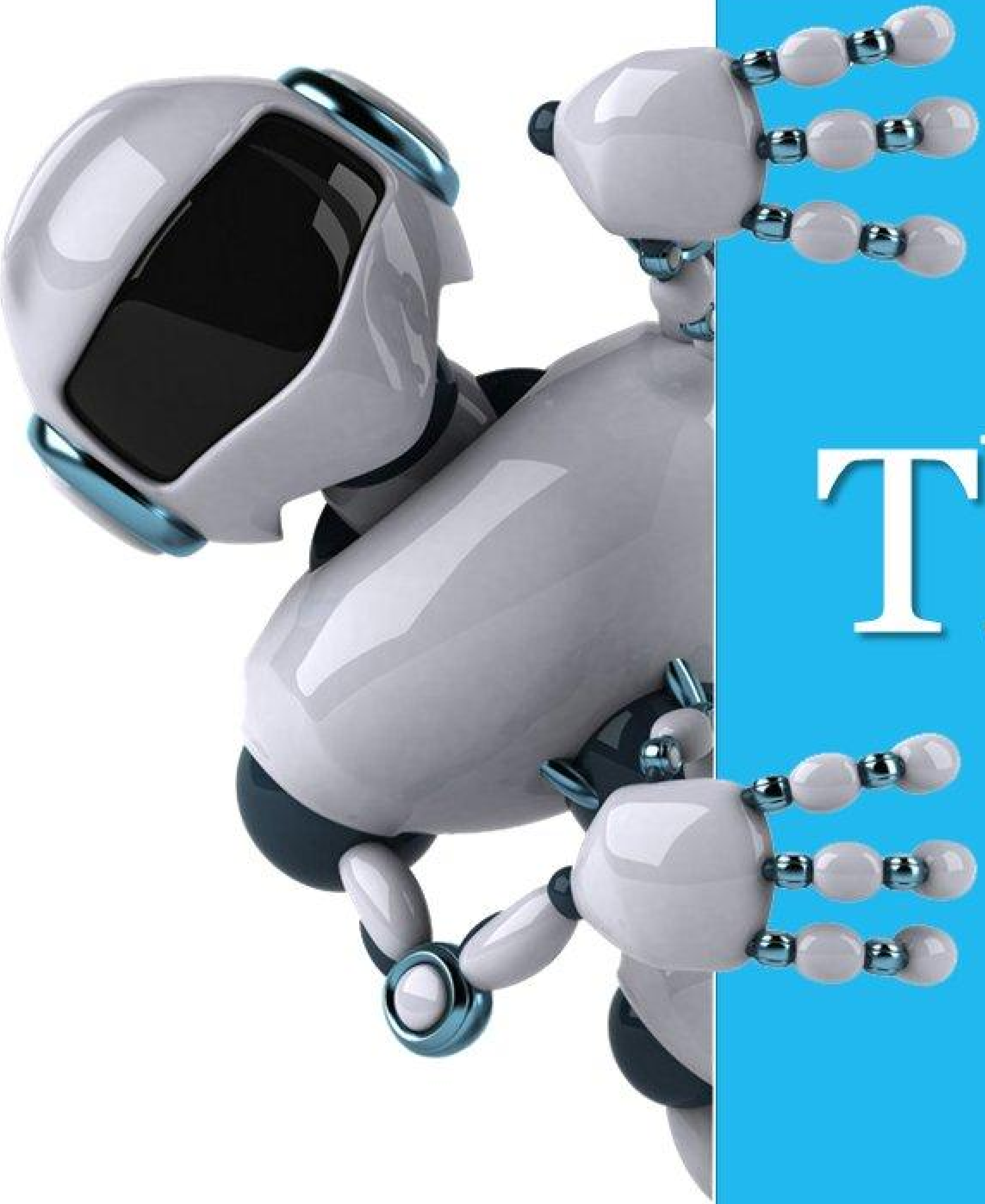
- **Mathematical derivation of Gradient Descent in simple Linear Regression :**
- 3. Now we update the current value of **b** and **a** using the following equation:
- $b = b - L \times D_b$
- $a = a - L \times D_a$

- 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 **b** and **a** that we are left with now will be the optimum values.

- Now going back to our analogy, **b** can be considered the current position of the person. **D is equivalent to the steepness of the slope and L can be the speed with which he moves.** Now the **new value of b 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 **b** and **a** our model is ready to make predictions !

Flowchart of Linear Regression using Gradient Descent





Thank you