Linear Regression Model

The linear regression model is:

YPred=wX+b

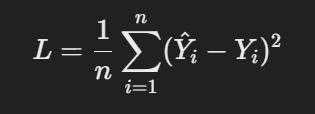
Where: YPred is the Predicted value of X which is an Input variable.

It has coefficients: w: Weight (slope), b: Bias (intercept)

In Linear Regression with **Stochastic Gradient Descent (SGD)**, instead of calculating the MSE for all data points at once (as in Batch Gradient Descent), we calculate the squared error for **one data point at a time**.

Loss Function

The loss function for linear regression is the Mean Squared Error (MSE):



For SGD, we update w and b for each individual data point, rather than the entire dataset at once.

**Why MSE is used?**

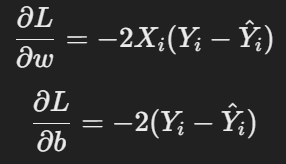
1. **Penalizes larger errors**: The squared difference amplifies larger errors, making the optimization sensitive to significant deviations.
2. **Continuous and differentiable**: This allows us to use gradient-based optimization methods like **Stochastic Gradient Descent (SGD)**.

How Gradient Descent Minimizes the Loss

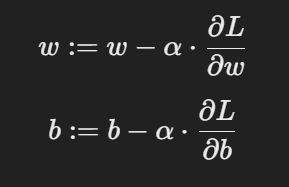
The goal of SGD is to minimize the **loss function** by updating the model parameters(w and b) in the direction of the negative gradient of the loss function.

**Gradient Descent Update Rules for updating model parameters by taking Partial Derivative :**

The gradients of w and b with respect to the loss function are:



Using these gradients, the parameters w and b are updated as follows:



Where:

* α: Learning rate

**Initialization and Parameters :**

1. Initialize w and b (e.g., w=0, b=0).
2. Set a learning rate (e.g., α=0.01).
3. Iterate over each data point in X and Y for a given number of epochs.

Where, epochs are the number of iterations.

**Step-by-Step Calculation (4 Epochs)**

**Epoch 1**

Start with w = 0 and b=0:

1. **For (X=1, Y=3):**

YPred=w⋅X+b=0⋅1+0=0

Error=Y−YPred=3−0=3

Update:

∂L/∂w=−2⋅1⋅3=−6, ∂L/∂b=−2⋅3=−6

w = 0−0.01⋅(−6) = 0.06 , b = 0−0.01⋅(−6)=0.06

1. **For (X=2, Y=4):**

YPred=w⋅X+b =0.06⋅2+0.06=0.18

Error=Y−YPred =4−0.18=3.82

Update:

∂L/∂w = −2⋅2⋅3.82= −15.28 , ∂L/∂b = −2⋅3.82=−7.64

W =0.06−0.01⋅(−15.28) = 0.2128 , b =0.06−0.01⋅(−7.64)=0.1364

1. **For (X=3, Y=2)**

YPred=w⋅X+b=0.2128⋅3+0.1364=0.7748

Error=Y−YPred=2−0.7748=1.2252

Update:

∂L/∂w=−2⋅3⋅1.2252=−7.3512 , ∂L/∂b=−2⋅1.2252=−2.4504

w=0.2128−0.01⋅(−7.3512)=0.2863, b=0.1364−0.01⋅(−2.4504)=0.1619

1. **For (X=4, Y=4)**

YPred =w⋅X+b=0.2863⋅4+0.1619=1.3069

Error=Y−YPred=4−1.3069=2.6931

Update:

∂L/∂w=−2⋅4⋅2.6931=−21.5448, ∂L/∂b=−2⋅2.6931=−5.3862

w =0.2863−0.01⋅(−21.5448)=0.5017, b =0.1619−0.01⋅(−5.3862)=0. 2157

1. **For(X=5, Y=5)**

YPred=w⋅X+b=0.5017⋅5+0.2157=2.7232

Error=Y−YPred=5−2.7232=2.2768

Update:

∂L/∂w=−2⋅5⋅2.2768=−22.768, ∂L/∂b=−2⋅2.2768=−4.5536

w =0.5017−0.01⋅(−22.768)=0.7294, b =0.2157−0.01⋅(−4.5536)=0.2613

**So, by the End of Epoch 1**

* w=0.7294w = 0.7294w=0.7294
* b=0.2613b = 0.2613b=0.2613

**Epoch 2**

We repeat the process with updated w and b. Let’s calculate w and b for each data point in this epoch:

**Step 1: (X=1, Y=3)**

YPred=0.7294⋅1+0.2613=0.9907

Error=3−0.9907=2.0093

Update:

∂L/∂w=−2⋅1⋅2.0093=−4.0186, ∂L/∂b=−2⋅2.0093=−4.0186

w:=0.7294−0.01⋅(−4.0186)=0.7696,b:=0.2613−0.01⋅(−4.0186)=0.3015

**Step 2: (X=2, Y=4)**

YPred=0.7696⋅2+0.3015=1.8407

Error=4−1.8407=2.1593

Update:

∂L/∂w=−2⋅2⋅2.1593=−8.6372, ∂L/∂b=−2⋅2.1593=−4.3186 w:=0.7696−0.01⋅(−8.6372)=0.8550,b:=0.3015−0.01⋅(−4.3186)=0.3447

**Step 3: (X=3, Y=2)**

YPred=0.8550⋅3+0.3447=2.9097

Error=2−2.9097=−0.9097

Update:

∂L/∂w=−2⋅3⋅(−0.9097)=5.4582,∂L/∂b=−2⋅(−0.9097)=1.8194

w:=0.8550−0.01⋅5.4582=0.8004,b:=0.3447−0.01⋅1.8194=0.3265

**Step 4: (X=4, Y=4)**

**YPred=w⋅X+b=0.8004⋅4+0.3265=3.5281**

**Error=Y−YPred=4−3.5281=0.4719**

**Update:**

**∂L/∂w=−2⋅4⋅0.4719=−3.7752,∂L/ ∂b=−2⋅0.4719=−0.9438 w:=0.8004−0.01⋅(−3.7752)=0.8382,b:=0.3265−0.01⋅(−0.9438)=0.3360**

**Step 5: (X=5, Y=5)**

**YPred=w⋅X+b=0.8382⋅5+0.3360=4.5270**

**Error=Y−YPred=5−4.5270=0.4730**

**Update:**

**∂L/∂w=−2⋅5⋅0.4730=−4.7300,∂L/∂b=−2⋅0.4730=−0.9460 w:=0.8382−0.01⋅(−4.7300)=0.8855,b:=0.3360−0.01⋅(−0.9460)=0.3455**

**End of Epoch 2**

* **w=0.8855w = 0.8855w=0.8855**
* **b=0.3455b = 0.3455b=0.3455**

**Epoch 3**

**Step 1: (X=1, Y=3)**

**YPred=w⋅X+b=0.8855⋅1+0.3455=1.2310**

**Error=Y−Y^=3−1.2310=1.7690**

**Update:**

**∂L/∂w=−2⋅1⋅1.7690=−3.5380,∂L/∂b=−2⋅1.7690=−3.5380**

**w:=0.8855−0.01⋅(−3.5380)=0.9208,b:=0.3455−0.01⋅(−3.5380)=0.3808**

**Step 2: (X=2, Y=4)**

**YPred=0.9208⋅2+0.3808=2.2224**

**Error=Y−YPred=4−2.2224=1.7776**

**Update:**

**∂L/∂w=−2⋅2⋅1.7776=−7.1104,∂L/∂b=−2⋅1.7776 w:=0.9208−0.01⋅(−7.1104)=0.9919,b:=0.3808−0.01⋅(−3.5552)=0.4164**

**Step 3: (X=3, Y=2)**

**YPred=0.9919⋅3+0.4164=3.3919**

**Error=Y−YPred=2−3.3919=−1.3919**

**Update:**

**∂L/∂w=−2⋅3⋅(−1.3919)=8.3514,∂L/∂b=−2⋅(−1.3919)=2.7838 w:=0.9919−0.01⋅8.3514=0.9084,b:=0.4164−0.01⋅2.7838=0.3885**

**Steps 4 and 5**

**Repeat the same procedure as before for X=4 and X=5.**

**Epoch 4**

**Repeat the same steps as before, with w and b values carried forward from Epoch 3. As we iterate, the values of w and b will converge further, reducing the error.**