Gradient Descent

Gradient Descent is an optimization algorithm used in linear regression to find the best-fit line for the data. It works by gradually adjusting the line's slope and intercept to reduce the difference between actual and predicted values. This process helps the model make accurate predictions by minimizing errors step by step.

Linear regression finds the best-fit line for a dataset by minimizing the error between the actual and predicted values. This error is measured using the cost function usually Mean Squared Error (MSE). The goal is to find the model parameters i.e. the slope m and the intercept b that minimize this cost function.

Gradient Descent Work in Linear Regression

Initializing Parameters: Start with random initial values for the slope (w) and intercept (*b*).

Calculate the Cost Function: Measure the error using the Mean Squared Error (MSE):
$J(m,b)=1/n\sum_{i=1}^{n}((wx_i+b)-y_i)^2$
Compute the Gradient: Calculate how much the cost function changes with respect to w and *b*.

- For slope w :

$$\frac{\partial J}{\partial m} = -\frac{2}{n}\sum_{i=1}^{n} x_i(y_i - (mx_i + b))$$

- For intercept *b*:

$$\frac{\partial J}{\partial b} = -\frac{2}{n}\sum_{i=1}^{n}(y_i - (mx_i + b))$$

Update Parameters: Change w and *b* to reduce the error:

- For slope w :

$$m = m - \alpha \cdot \frac{\partial J}{\partial m}$$

- For intercept *b*:

$$b = b - \alpha \cdot \frac{\partial J}{\partial b}$$

Iteration 1 (by hand)

Initial parameters: $m = 0, b = 0.$

1. **Calculate Predictions ($\hat{y}_i$) for current `m` and `b`:**
   1. For (x=1, y=2): $\hat{y}_1 = (0 \cdot 1) + 0 = 0$
   2. For (x=2, y=4): $\hat{y}_2 = (0 \cdot 2) + 0 = 0$
   3. For (x=3, y=5): $\hat{y}_3 = (0 \cdot 3) + 0 = 0$
2. **Calculate Gradients (Derivatives):**
   1. **Gradient for `m`:**

$$\frac{\partial J}{\partial m} = -\frac{2}{3}\sum_{i=1}^{3} x_i(y_i - \hat{y}_i)$$

$$\frac{\partial J}{\partial m} = -\frac{2}{3} \cdot [1 \cdot (2-0) + 2 \cdot (4-0) + 3 \cdot (5-0)]$$

$$\frac{\partial J}{\partial m} = -\frac{2}{3} \cdot [2 + 8 + 15] = -\frac{2}{3} \cdot 25 \approx -16.67$$

2. **Gradient for `b`:**

$$\frac{\partial J}{\partial b} = -\frac{2}{3}\sum_{i=1}^{3}(y_i - \hat{y}_i)$$

$$\frac{\partial J}{\partial b} = -\frac{2}{3} \cdot [(2-0)+(4-0)+(5-0)]$$

$$\frac{\partial J}{\partial b} = -\frac{2}{3} \cdot [2+4+5] = -\frac{2}{3} \cdot 11 \approx -7.33$$

3. **Update Parameters using Learning Rate ($\alpha = 0.01$):**

   1. **New `m`:** $m_{new} = 0 - 0.01 \cdot (-16.67) \approx 0.1667$
   2. **New `b`:** $b_{new} = 0 - 0.01 \cdot (-7.33) \approx 0.0733$

Iteration 2 (by hand)

Updated parameters: $m \approx 0.1667$, $b \approx 0.0733$.

1. **Calculate Predictions ($\hat{y}_i$) for new `m` and `b`:**

   1. For (x=1, y=2): $\hat{y}_1 = (0.1667 \cdot 1) + 0.0733 \approx 0.24$
   2. For (x=2, y=4): $\hat{y}_2 = (0.1667 \cdot 2) + 0.0733 \approx 0.4067$
   3. For (x=3, y=5): $\hat{y}_3 = (0.1667 \cdot 3) + 0.0733 \approx 0.5734$

2. **Calculate Gradients:**

   1. **Gradient for `m`:**

$$\frac{\partial J}{\partial m} = -\frac{2}{3} \cdot [1 \cdot (2 - 0.24) + 2 \cdot (4 - 0.4067) + 3 \cdot (5 - 0.5734)]$$

$$\frac{\partial J}{\partial m} = -\frac{2}{3} \cdot [1.76 + 7.1866 + 13.28] \approx -14.82$$

   2. **Gradient for `b`:**

$$\frac{\partial J}{\partial b} = -\frac{2}{3} \cdot [(2 - 0.24) + (4 - 0.4067) + (5 - 0.5734)]$$

$$\frac{\partial J}{\partial b} = -\frac{2}{3} \cdot [1.76 + 3.5933 + 4.4266] \approx -6.52$$

3. **Update Parameters:**

   1. **New `m`:** $m_{new} = 0.1667 - 0.01 \cdot (-14.82) \approx 0.3149$

   2. **New `b`:** $b_{new} = 0.0733 - 0.01 \cdot (-6.52) \approx 0.1385$