Gradient Decent Manual Calculation PART 3

Our objective:

We want to find the best line that fits two points using gradient decent.

Therefore, the line formula is: $\hat{y} = mx + b$

We will improve m (slope) and b (intercept) over 3 iterations using the gradient descent algorithm.

Given:

- start with: $m = -1$, $b = 1$
- Learning rate: $\alpha = 0.1$
- Points: $(1,3)$ and $(3,6)$

Cost function (J):

We use Mean Squared Error (MSE):

$$J(m,b) = \frac{1}{n}\left(\sum (y_i - \hat{y}_i)^2\right)$$

This helps us measure how wrong our line is. So our goal is to minimize J.

Gradient Formulas:

These formulas will help us know how much to change m and b.

$$\frac{\partial J}{\partial m} = -\frac{2}{n}\sum (y_i - \hat{y}_i)x_i$$

$$\frac{\partial J}{\partial b} = -\frac{2}{n}\sum (y_i - \hat{y}_i)$$

The 2 comes from the derivative of a square: $\frac{d}{dx}(x^2) = 2x$

iteration 1.

- Predict values using $\hat{y} = mx + b$

For $(1,3)$: $\hat{y}_1 = (-1)(1) + 1$
$= 0$

For $(3,6)$: $\hat{y}_2 = (-1)(3) + 1$
$= -2$

- Calculate errors:

$$\text{Error} = y - \hat{y}$$

Point $(1,3)$:
$3 - 0 = \underline{\underline{3}}$

Point $(3,6)$:
$6 - (-2) = \underline{\underline{8}}$

- Use gradient formulas

$$\frac{\partial J}{\partial m} = -\frac{2}{n}\sum (y_i - \hat{y}_i)x_i$$

$$\frac{\partial J}{\partial m} = -\frac{2}{2}[(3)(1) + (8)(3)]$$

$$= -1(3+24)$$

$$= \underline{\underline{-27}}$$

$$\frac{\partial J}{\partial b} = -\frac{2}{n}\sum (y_i - \hat{y}_i)$$

$$= -\frac{2}{2}(3+8) = \underline{\underline{-11}}$$

- Update m and b

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

$$m = -1 - 0.1(-27)$$

$$= -1 + 2.7$$

$$= \underline{\underline{1.7}}$$

$$b_{new} = b_{old} - \alpha \times \frac{\partial J}{\partial b}$$

$$= 1 - 0.1(-11)$$

$$= 1 + 1.1$$

$$= \underline{\underline{2.1}}$$

$\therefore$ After Iteration 1

$$m = 1.7$$

$$b = 2.1$$

Iteration 2.

- Predict values

$$\hat{y}_1 = 1.7(1) + 2.1$$

$$= \underline{\underline{3.8}}$$

$$\hat{y}_2 = 1.7(3) + 2.1$$

$$= \underline{\underline{7.2}}$$

- Errors

$y - \hat{y}_1 = 3 - 3.8 = \underline{-0.8}$

$y - \hat{y}_2 = 6 - 7.2 = \underline{\underline{-1.2}}$

- Gradients decent formula

$$\frac{\partial J}{\partial m} = -\frac{2}{2}\left[(-0.8)(1) + (-1.2)(3)\right]$$
$$= -1(-0.8 - 3.6)$$
$$= \underline{\underline{4.4}}$$

$$\frac{\partial J}{\partial b} = -\frac{2}{2}(-0.8 + -1.2)$$
$$= \underline{\underline{2}}$$

- Update

$$m = 1.7 - 0.1(4.4)$$
$$= 1.7 - 0.44$$
$$= \underline{\underline{1.26}}$$

$$b = 2.1 - 0.1(2)$$
$$= 2.1 - 0.2$$
$$= \underline{\underline{1.9}}$$

$\therefore$ After iteration 2

$m = 1.26$

$b = 1.9$

---

iteration 3.

- Predict values

$\hat{y}_1 = 1.26(1) + 1.9$

$= \underline{\underline{3.16}}$

$\hat{y}_2 = 1.26(3) + 1.9$

$= \underline{\underline{5.68}}$

- Errors

$$y - \hat{y}_1 = 3 - 3.16$$
$$= \underline{-0.16}$$

$$y - y_2 = 6 - 5.68$$
$$= \underline{0.32}$$

- Gradients

$$\frac{\partial J}{\partial m} = -\frac{2}{2}\left[(-0.16)(1) + (0.32)(3)\right]$$
$$= -1(-0.16 + 0.96)$$
$$= \underline{-0.8}$$

$$\frac{\partial J}{\partial b} = -\frac{2}{2}(-0.16 + 0.32)$$
$$= \underline{-0.16}$$

- Update

$$m = 1.26 - 0.1(-0.8)$$
$$= 1.26 + 0.08$$
$$= \underline{1.34}$$

$$b = 1.9 - 0.1(-0.016)$$
$$= 1.9 + 0.016$$
$$= \underline{1.916}$$

$\therefore$ After iteration 3

$$m = 1.34$$

$$b = 1.916$$

Final Results (After 3 iterations)

Iteration	m	b.
Start	-1	1
Iteration 1	1.7	2.1
Iteration 2	1.26	1.9
Iteration 3	1.34	1.916

Description:

- We started with a bad guess: $m = -1$, $b = 1$
- We use gradient descent to adjust m and b little by little.
- After 3 steps, our line is getting much closer to fitting the points.
- This is how machines learn from data by reducing the error step by step.