Gradient Descent Example with House Prices

Let's walk through an example of applying gradient descent to a simple linear regression problem using a small dataset of predicted house prices based on floor space.

Example Dataset

Consider the following dataset where $x^{(i)}$ is the floor space (in square feet) and $y^{(i)}$ is the actual house price (in thousands of dollars):

Floor Space $x^{(i)}$	House Price $y^{(i)}$
650	150
800	200
1200	250
1500	300
1800	350

Step 1: Initialize Parameters

Let's start with initial guesses for w and b:

$$w = 0.1, \quad b = 0.1$$

Assume a learning rate α of 0.01.

Step 2: Linear Regression Equation

The prediction for each house price based on the current w and b is:

$$\hat{y}^{(i)} = w \cdot x^{(i)} + b$$

Let's compute the predictions for each data point using the initial parameters:

$$\hat{y}^{(1)} = 0.1 \times 650 + 0.1 = 65.1$$

$$\hat{y}^{(2)} = 0.1 \times 800 + 0.1 = 80.1$$

$$\hat{y}^{(3)} = 0.1 \times 1200 + 0.1 = 120.1$$

$$\hat{y}^{(4)} = 0.1 \times 1500 + 0.1 = 150.1$$

$$\hat{y}^{(5)} = 0.1 \times 1800 + 0.1 = 180.1$$

Step 3: Compute the Cost Function

The cost function $J(w, b)$ is:

$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$$

For our dataset:

$$J(0.1, 0.1) = \frac{1}{10} \left[(65.1 - 150)^2 + (80.1 - 200)^2 + (120.1 - 250)^2 + (150.1 - 300)^2 + (180.1 - 350)^2 \right]$$

Let's compute each term:

$$(65.1 - 150)^2 = 7196.01$$
$$(80.1 - 200)^2 = 14376.01$$
$$(120.1 - 250)^2 = 16980.01$$
$$(150.1 - 300)^2 = 22470.01$$
$$(180.1 - 350)^2 = 28812.01$$

Summing these:

$$J(0.1, 0.1) = \frac{1}{10} \times (7196.01 + 14376.01 + 16980.01 + 22470.01 + 28812.01) = \frac{1}{10} \times 89834.05 = 8983.405$$

Step 4: Compute the Gradients

Next, compute the gradients with respect to w and b:

$$\frac{\partial J(w,b)}{\partial w} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right) \cdot x^{(i)}$$
$$\frac{\partial J(w,b)}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)$$

Let's calculate the gradient for w:

$$\frac{\partial J(w,b)}{\partial w} = \frac{1}{5} [(65.1 - 150) \times 650 + (80.1 - 200) \times 800 + (120.1 - 250) \times 1200 + (150.1 - 300) \times 1500 + (18$$

Simplifying each term:

$$= \frac{1}{5} \times [(-84.9) \times 650 + (-119.9) \times 800 + (-129.9) \times 1200 + (-149.9) \times 1500 + (-169.9) \times 1800]$$

$$= \frac{1}{5} \times [-55285 - 95920 - 155880 - 224850 - 305820] = \frac{1}{5} \times (-838755) = -167751$$

Now for b:

$$\frac{\partial J(w,b)}{\partial b} = \frac{1}{5} [(65.1 - 150) + (80.1 - 200) + (120.1 - 250) + (150.1 - 300) + (180.1 - 350)]$$

Simplifying:

$$= \frac{1}{5} \times [-84.9 - 119.9 - 129.9 - 149.9 - 169.9] = \frac{1}{5} \times (-654.5) = -130.9$$

Step 5: Update the Parameters

Finally, update the parameters using the gradients:

$$w := w - \alpha \cdot \frac{\partial J(w,b)}{\partial w}$$

$$b := b - \alpha \cdot \frac{\partial J(w,b)}{\partial b}$$

Substituting in the values:

$$w := 0.1 - 0.01 \times (-167751) = 0.1 + 1677.51 = 1677.61$$

$$b := 0.1 - 0.01 \times (-130.9) = 0.1 + 1.309 = 1.409$$

Summary

After one iteration of gradient descent, the updated parameters are:

$$w = 1677.61, \quad b = 1.409$$

These updated values would then be used in the next iteration, and the process would continue until convergence, gradually minimizing the cost function $J(w,b)$.