Gradient Descent for Linear Regression

Manual Calculations and Iteration Table

1. Linear Regression Model

The hypothesis function is:

$$\hat{y} = wx + b$$

The Mean Squared Error (MSE) cost function is:

$$J(w, b) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (wx_i + b))^2$$

2. Gradients

The partial derivatives of the cost function are:

$$\frac{\partial J}{\partial w} = -\frac{2}{n} \sum_{i=1}^{n} x_i \left(y_i - (wx_i + b)\right)$$

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

Gradient descent update rules:

$$w \leftarrow w - \alpha \frac{\partial J}{\partial w}, \qquad b \leftarrow b - \alpha \frac{\partial J}{\partial b}$$

We use:

$$X = [1, 4, 8, 11, 15], \quad Y = [4, 7, 12, 15, 17], \quad n = 5, \quad \alpha = 0.01$$

In the calculations below, numerical values are rounded to 4 decimal places where appropriate.

3. Manual Calculations (First 3 Iterations, Step by Step)

We start with:

$$w_0 = 0, \qquad b_0 = 0$$

3.1 Iteration 1 (From (w_0, b_0) to (w_1, b_1))

Step 1: Predictions

Using

$$\hat{y}_i = w_0 x_i + b_0$$

with $w_0 = 0$, $b_0 = 0$:

$$\hat{y}_1 = (0)(1) + 0 = 0$$
$$\hat{y}_2 = (0)(4) + 0 = 0$$
$$\hat{y}_3 = (0)(8) + 0 = 0$$
$$\hat{y}_4 = (0)(11) + 0 = 0$$
$$\hat{y}_5 = (0)(15) + 0 = 0$$

So:

$$\hat{y} = [0,\ 0,\ 0,\ 0,\ 0]$$

Step 2: Errors

Error for each data point:

$$e_i = y_i - \hat{y}_i$$

$$e_1 = 4 - 0 = 4$$
$$e_2 = 7 - 0 = 7$$
$$e_3 = 12 - 0 = 12$$
$$e_4 = 15 - 0 = 15$$
$$e_5 = 17 - 0 = 17$$

So:

$$e = [4,\ 7,\ 12,\ 15,\ 17]$$

Step 3: Cost J_0

$$J_0 = J(w_0, b_0) = \frac{1}{5} \sum_{i=1}^{5} e_i^2$$

Compute each square:

$$4^2 = 16, \quad 7^2 = 49, \quad 12^2 = 144, \quad 15^2 = 225, \quad 17^2 = 289$$

Sum:

$$16 + 49 + 144 + 225 + 289 = 723$$

So:

$$J_0 = \frac{723}{5} = 144.6$$

Step 4: Gradients at (w_0, b_0)

Gradient w.r.t. w**:**

$$\frac{\partial J}{\partial w} = -\frac{2}{n}\sum_{i=1}^{5} x_i e_i = -\frac{2}{5}\sum_{i=1}^{5} x_i e_i$$

Compute each product $x_i e_i$:

$$x_1 e_1 = 1 \cdot 4 = 4$$
$$x_2 e_2 = 4 \cdot 7 = 28$$
$$x_3 e_3 = 8 \cdot 12 = 96$$
$$x_4 e_4 = 11 \cdot 15 = 165$$
$$x_5 e_5 = 15 \cdot 17 = 255$$

Sum:

$$\sum_{i=1}^{5} x_i e_i = 4 + 28 + 96 + 165 + 255 = 548$$

Now substitute:

$$\frac{\partial J}{\partial w} = -\frac{2}{5} \cdot 548 = -\frac{1096}{5} = -219.2$$

Gradient w.r.t. b**:**

$$\frac{\partial J}{\partial b} = -\frac{2}{n}\sum_{i=1}^{5} e_i = -\frac{2}{5}\sum_{i=1}^{5} e_i$$

Sum of errors:

$$\sum_{i=1}^{5} e_i = 4 + 7 + 12 + 15 + 17 = 55$$

So:

$$\frac{\partial J}{\partial b} = -\frac{2}{5} \cdot 55 = -\frac{110}{5} = -22$$

Step 5: Parameter Update

Using:

$$w_1 = w_0 - \alpha\frac{\partial J}{\partial w}, \qquad b_1 = b_0 - \alpha\frac{\partial J}{\partial b}$$

with $\alpha = 0.01$:

$$w_1 = 0 - 0.01(-219.2) = 0 + 2.192 = 2.192$$
$$b_1 = 0 - 0.01(-22) = 0 + 0.22 = 0.22$$

So after Iteration 1:

$$w_1 = 2.192, \qquad b_1 = 0.22$$

3.2 Iteration 2 (From (w_1, b_1) to (w_2, b_2))

Now we start with:

$$w_1 = 2.192, \qquad b_1 = 0.22$$

Step 1: Predictions

$$\hat{y}_i = w_1 x_i + b_1$$

$$\hat{y}_1 = (2.192)(1) + 0.22 = 2.192 + 0.22 = 2.412$$
$$\hat{y}_2 = (2.192)(4) + 0.22 = 8.768 + 0.22 = 8.988$$
$$\hat{y}_3 = (2.192)(8) + 0.22 = 17.536 + 0.22 = 17.756$$
$$\hat{y}_4 = (2.192)(11) + 0.22 = 24.112 + 0.22 = 24.332$$
$$\hat{y}_5 = (2.192)(15) + 0.22 = 32.880 + 0.22 = 33.100$$

So:

$$\hat{y} = [2.412,\ 8.988,\ 17.756,\ 24.332,\ 33.100]$$

Step 2: Errors

$$e_i = y_i - \hat{y}_i$$

$$e_1 = 4 - 2.412 = 1.588$$
$$e_2 = 7 - 8.988 = -1.988$$
$$e_3 = 12 - 17.756 = -5.756$$
$$e_4 = 15 - 24.332 = -9.332$$
$$e_5 = 17 - 33.100 = -16.100$$

So:

$$e = [1.588,\ -1.988,\ -5.756,\ -9.332,\ -16.100]$$

Step 3: Cost J_1

$$J_1 = J(w_1, b_1) = \frac{1}{5} \sum_{i=1}^{5} e_i^2$$

Compute each square:

$$(1.588)^2 = 2.521744$$
$$(-1.988)^2 = 3.952144$$
$$(-5.756)^2 = 33.131536$$
$$(-9.332)^2 = 87.086224$$
$$(-16.100)^2 = 259.210000$$

Sum:

$$2.521744 + 3.952144 + 33.131536 + 87.086224 + 259.210000 = 385.901648$$

So:

$$J_1 = \frac{385.901648}{5} \approx 77.1803$$

Step 4: Gradients at (w_1, b_1)

Gradient w.r.t. w**:**

$$\frac{\partial J}{\partial w} = -\frac{2}{5}\sum_{i=1}^{5} x_i e_i$$

Compute $x_i e_i$:

$$x_1 e_1 = 1 \cdot 1.588 = 1.588$$
$$x_2 e_2 = 4 \cdot (-1.988) = -7.952$$
$$x_3 e_3 = 8 \cdot (-5.756) = -46.048$$
$$x_4 e_4 = 11 \cdot (-9.332) = -102.652$$
$$x_5 e_5 = 15 \cdot (-16.100) = -241.500$$

Sum:

$$\sum_{i=1}^{5} x_i e_i = 1.588 - 7.952 - 46.048 - 102.652 - 241.500 = -396.564$$

Now:

$$\frac{\partial J}{\partial w} = -\frac{2}{5}(-396.564) = \frac{2}{5} \cdot 396.564 = 0.4 \cdot 396.564 = 158.6256$$

Gradient w.r.t. b**:**

$$\frac{\partial J}{\partial b} = -\frac{2}{5}\sum_{i=1}^{5} e_i$$

Sum of errors:

$$\sum_{i=1}^{5} e_i = 1.588 + (-1.988) + (-5.756) + (-9.332) + (-16.100)$$

$$= 1.588 - 1.988 - 5.756 - 9.332 - 16.100 = -31.588$$

So:

$$\frac{\partial J}{\partial b} = -\frac{2}{5}(-31.588) = \frac{2}{5} \cdot 31.588 = 0.4 \cdot 31.588 = 12.6352$$

Step 5: Parameter Update

$$w_2 = w_1 - \alpha \frac{\partial J}{\partial w} = 2.192 - 0.01(158.6256) = 2.192 - 1.586256 = 0.605744$$

$$b_2 = b_1 - \alpha \frac{\partial J}{\partial b} = 0.22 - 0.01(12.6352) = 0.22 - 0.126352 = 0.093648$$

So after Iteration 2:

$$w_2 \approx 0.6057, \qquad b_2 \approx 0.0936$$

3.3 Iteration 3 (From (w_2, b_2) to (w_3, b_3))

Start with:

$$w_2 = 0.605744, \qquad b_2 = 0.093648$$

Step 1: Predictions

$$\hat{y}_i = w_2 x_i + b_2$$

$$\hat{y}_1 = (0.605744)(1) + 0.093648 = 0.605744 + 0.093648 = 0.699392$$

$$\hat{y}_2 = (0.605744)(4) + 0.093648 = 2.422976 + 0.093648 = 2.516624$$

$$\hat{y}_3 = (0.605744)(8) + 0.093648 = 4.845952 + 0.093648 = 4.939600$$

$$\hat{y}_4 = (0.605744)(11) + 0.093648 = 6.663184 + 0.093648 = 6.756832$$

$$\hat{y}_5 = (0.605744)(15) + 0.093648 = 9.086160 + 0.093648 = 9.179808$$

So:

$$\hat{y} = [0.699392,\ 2.516624,\ 4.939600,\ 6.756832,\ 9.179808]$$

Step 2: Errors

$$e_i = y_i - \hat{y}_i$$

$$e_1 = 4 - 0.699392 = 3.300608$$
$$e_2 = 7 - 2.516624 = 4.483376$$
$$e_3 = 12 - 4.939600 = 7.060400$$
$$e_4 = 15 - 6.756832 = 8.243168$$
$$e_5 = 17 - 9.179808 = 7.820192$$

So:

$$e = [3.300608,\ 4.483376,\ 7.060400,\ 8.243168,\ 7.820192]$$

Step 3: Cost J_2

$$J_2 = J(w_2, b_2) = \frac{1}{5}\sum_{i=1}^{5} e_i^2$$

Compute each square (rounded):

$$(3.300608)^2 \approx 10.894013$$

$$(4.483376)^2 \approx 20.100660$$

$$(7.060400)^2 \approx 49.849248$$

$$(8.243168)^2 \approx 67.949819$$

$$(7.820192)^2 \approx 61.155403$$

Sum:

$$10.894013 + 20.100660 + 49.849248 + 67.949819 + 61.155403 \approx 209.949143$$

So:

$$J_2 = \frac{209.949143}{5} \approx 41.9898$$

Step 4: Gradients at (w_2, b_2)

Gradient w.r.t. w:

$$\frac{\partial J}{\partial w} = -\frac{2}{5}\sum_{i=1}^{5} x_i e_i$$

Compute $x_i e_i$:

$$x_1 e_1 = 1 \cdot 3.300608 = 3.300608$$

$$x_2 e_2 = 4 \cdot 4.483376 = 17.933504$$

$$x_3 e_3 = 8 \cdot 7.060400 = 56.483200$$

$$x_4 e_4 = 11 \cdot 8.243168 = 90.674848$$

$$x_5 e_5 = 15 \cdot 7.820192 = 117.302880$$

Sum:

$$\sum_{i=1}^{5} x_i e_i = 3.300608 + 17.933504 + 56.483200 + 90.674848 + 117.302880 = 285.695040$$

Thus:

$$\frac{\partial J}{\partial w} = -\frac{2}{5}(285.695040) = -0.4 \cdot 285.695040 = -114.278016$$

Gradient w.r.t. b:

$$\frac{\partial J}{\partial b} = -\frac{2}{5}\sum_{i=1}^{5} e_i$$

Sum of errors:

$$\sum_{i=1}^{5} e_i = 3.300608 + 4.483376 + 7.060400 + 8.243168 + 7.820192 = 30.907744$$

So:

$$\frac{\partial J}{\partial b} = -\frac{2}{5}(30.907744) = -0.4 \cdot 30.907744 = -12.3630976$$

Step 5: Parameter Update

$$w_3 = w_2 - \alpha\frac{\partial J}{\partial w} = 0.605744 - 0.01(-114.278016) = 0.605744 + 1.14278016 = 1.74852416$$

$$b_3 = b_2 - \alpha\frac{\partial J}{\partial b} = 0.093648 - 0.01(-12.3630976) = 0.093648 + 0.123630976 = 0.217278976$$

So after Iteration 3:

$$w_3 \approx 1.7485, \qquad b_3 \approx 0.2173$$

4. Ten Iteration Summary Table

Below is a summary of the first 10 iterations (values rounded to 4 decimal places).

Iter	w_{before}	b_{before}	$\partial J/\partial w$	$\partial J/\partial b$	w_{after}	b_{after}	J_{before}
1	0.0000	0.0000	-219.2000	-22.0000	2.1920	0.2200	144.6000
2	2.1920	0.2200	158.6256	12.6352	0.6057	0.0936	77.1803
3	0.6057	0.0936	-114.2780	-12.3631	1.7485	0.2173	41.9898
4	1.7485	0.2173	82.8375	5.7115	0.9201	0.1602	23.6138
5	0.9201	0.1602	-59.5399	-7.3253	1.5155	0.2334	14.0104
6	1.5155	0.2334	43.2970	2.1094	1.0826	0.2123	8.9838
7	1.0826	0.2123	-30.9834	-4.6871	1.3924	0.2592	6.3453
8	1.3924	0.2592	22.6674	0.2400	1.1657	0.2568	4.9528
9	1.1657	0.2568	-16.0860	-3.3009	1.3266	0.2898	4.2105
10	1.3266	0.2898	11.9038	-0.7255	1.2076	0.2971	3.8077