Gradient Descent – Solved Example

x	y
0	1
1	3
2	7
3	13
4	21

In the data set given above x is the input feature and y is the target variable. For the above dataset, compute the model parameters using Gradient descent for 1 iteration
Note: Initialize the model parameters to 0, take the learning rate parameter value as 0.005, and tolerance = 0.1

Use the following information:

$$\begin{bmatrix} w_0^{t+1} \\ w_1^{t+1} \end{bmatrix} = \begin{bmatrix} w_0^t - \eta(-\sum_{i=1}^{n}[y_i - (w_0^t + w_1^t x_i)]) \\ w_1^t - \eta(-\sum_{i=1}^{n}[y_i - (w_0^t + w_1^t x_i)]x_i) \end{bmatrix}$$

In each step of the Gradient Descent we will do the following:

1. Compute the predicted values with the given the current slope (w1) and intercept(w0)

2. Compute the prediction errors (y - prediction)

3. Update the intercept(w0):

compute the derivative: sum(errors)

compute the adjustment as learning rate times the derivative

decrease the intercept by the adjustment

4. Update the slope:

compute the derivative: sum(errors*input)

compute the adjustment as learning rate times the derivative

decrease the slope by the adjustment

5. Compute the magnitude of the gradient (sqrt(($\sum_{i=1}^{n}[y_i - (w_0 + w_1 x_i)]$)^2)+ ($\sum_{i=1}^{n}[y_i - (w_0 + w_1 x_i)]x_i$)^2)

6. Check for convergence (magnitude of Gradient < tolerance)

initial_intercept (w0)= 0
initial_slope (w1) = 0
Learning rate = 0.05
tolerance = 0.01

First step:

Intercept = 0

Slope = 0

1. predictions = [0, 0, 0, 0, 0]

2. errors = [1, 3, 7, 13, 21]

3. update Intercept

- sum([1, 3, 7, 13, 21]) = 45
- adjustment = 0.005 * 45 = 0.225
- new_intercept = 0 – (-0.225) = 0.225

4. update Slope

- sum([0, 1, 2, 3, 4] * [1, 3, 7, 13, 21]) = 140
- adjustment = 0.005 * 140 = 0.7
- new_slope = 0 – (-0.7) = 0.7

5. magnitude = sqrt((45)^2 + (140)^2) = 147.05

6. magnitude > tolerance: not converged

Iteration2:

Intercept = 0.225

Slope = 0.7

1. predictions = [0.225, 0.925, 1.625, 2.325, 3.025]

2. errors = [0.775, 2.075, 5.375, 10.675, 17.975]

3. update Intercept

- sum([0.775, 2.075, 5.375, 10.675, 17.975]) = 36.875
- adjustment = 0.005 * 36.875 = 0.1843
- new_intercept = 0.225-(-0.1843) = 0.4093

4. update Slope

- sum([0, 1, 2, 3, 4] * [0.775, 2.075, 5.375, 10.675, 17.975]) = 116.75
- adjustment = 0.005 * 116.75 = 0.58375

- new_slope = 0.7 – (-0.58375) = 1.28375

5. magnitude = sqrt((36.87)^2 + (116.75)^2) = 122.435

6. magnitude > tolerance: not converged

Proceeding this way we will arrive at the solution