BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI – HYDERABAD CAMPUS
BITS F464 : MACHINE LEARNING
FIRST SEMESTER 2023-2024

A worked-out example on Linear Regression

Here's a step-by-step solved numerical problem on simple linear regression with gradient descent.

Let's assume the following dataset related to the relationship between the number of hours studied and the corresponding exam scores of 10 students:

Hours Studied (x)	Exam Score (y)
1	53
2	62
3	70
4	65
5	75
6	82
7	90
8	95
9	89
10	100

We want to find the best fit line that represents the relationship between hours studied and exam scores using simple linear regression. The equation for a simple linear regression model is:

$$y = mx + b$$

where:

- y is the dependent variable (exam score in this case)
- x is the independent variable (hours studied in this case)
- m is the slope (weight parameter)
- b is the y-intercept (bias parameter)

We will use gradient descent to update the weight parameters m and b iteratively. The cost function we want to minimize is the Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - (mx_i + b)\right)^2$$

where N is the number of data points, and y_i and x_i are the exam score and hours studied of the i-th data point, respectively.

Let's start the gradient descent process with initial values m=0 and b=0, and a learning rate α=0.01.

Step 1: Calculate the predicted values for each data point using the current weight parameters:

$$y_{pred} = mx + b$$

Step 2: Calculate the partial derivatives of the cost function with respect to m and b:

$$\frac{\partial MSE}{\partial m} = \frac{-2}{N}\sum_{i=1}^{N} x_i(y_i - y_{pred_i})$$

$$\frac{\partial MSE}{\partial b} = \frac{-2}{N}\sum_{i=1}^{N} (y_i - y_{pred_i})$$

Step 3: Update the weight parameters using the gradients and the learning rate:

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

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

Step 4: Repeat steps 1 to 3 for the desired number of iterations.

Let's perform 3 iterations of gradient descent:

Iteration 1:

$m_{old} = 0$

$b_{old} = 0$

Step 1: Calculate the predicted values for each data point:

$y_{pred} = 0 \cdot x + 0 = 0$

Step 2: Calculate the partial derivatives of the cost function with respect to m and b:

$$\frac{\partial MSE}{\partial m} = \frac{-2}{10}\sum_{i=1}^{10} x_i(y_i - y_{pred_i})$$

$$\frac{\partial MSE}{\partial b} = \frac{-2}{10}\sum_{i=1}^{10} (y_i - y_{pred_i})$$

Now, plug in the values for x_i and y_i from the dataset:

$$\frac{\partial MSE}{\partial m} = -941.0$$

$$\frac{\partial MSE}{\partial b} = -156.2$$

Step 3: Update the weight parameters using the gradients and the learning rate:

$$m_{new} = 0 - 0.01 \cdot (-941.0) = 9.41$$

$$b_{new} = 0 - 0.01 \cdot (-156.2) = 1.562$$

Iteration 2:

$$m_{old} = 9.41$$

$$b_{old} = 1.562$$

Step 1: Calculate the predicted values for each data point:

$$y_{pred} = 9.41 \cdot x + 1.562$$

Step 2: Calculate the partial derivatives of the cost function with respect to m and b:

$$\frac{\partial MSE}{\partial m} = -199.248$$

$$\frac{\partial MSE}{\partial b} = -49.566$$

Step 3: Update the weight parameters using the gradients and the learning rate:

$$m_{new} = 9.41 - 0.01 \cdot (-199.248) = 11.402$$

$$b_{new} = 1.562 - 0.01 \cdot (-49.566) = 2.058$$

Iteration 3:

$$m_{old} = 11.402$$

$$b_{old} = 2.058$$

Step 1: Calculate the predicted values for each data point:

$$y_{pred} = 11.402 \cdot x + 2.058$$

Step 2: Calculate the partial derivatives of the cost function with respect to m and b:

$$\frac{\partial MSE}{\partial m} = -40.408$$

$$\frac{\partial MSE}{\partial b} = -26.662$$

Step 3: Update the weight parameters using the gradients and the learning rate:

$m_{new} = 11.402 - 0.01 \cdot (-40.408) = 11.806$

$b_{new} = 2.058 - 0.01 \cdot (-26.662) = 2.325$

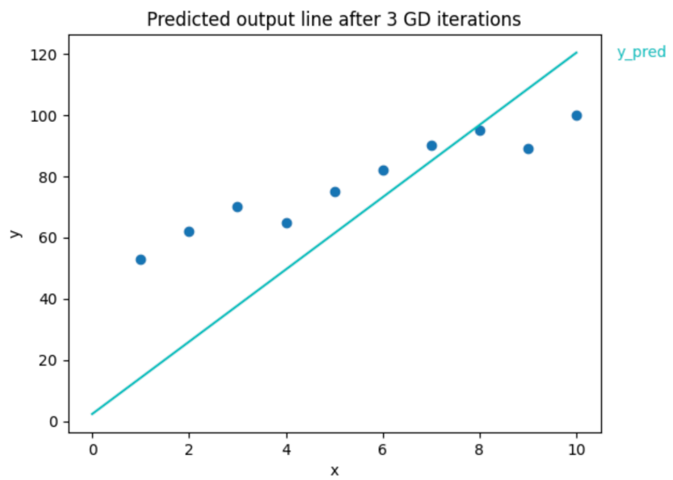
After three iterations of gradient descent, the updated weight parameters are approximately:

$m \approx 11.806$

$b \approx 2.325$

These values represent the best fit line for the dataset, and you can use them to predict exam scores based on the number of hours studied using the equation:

$$y_{pred} = 11.806x + 2.325$$



Keep in mind that in practice, you would usually perform many more iterations until the parameters converge to a stable solution, but in this example, we used a small number of iterations to illustrate the process.