Gradient Descent for Linear Regression

Here, we present the equations involved in applying gradient descent to linear regression, structured in a logical sequence:

1. Linear Regression Equation

For a single data point $(x^{(i)}, y^{(i)})$, the predicted value $\hat{y}^{(i)}$ is given by the linear regression equation:

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

Where:

- w is the weight (slope of the line).
- $x^{(i)}$ is the input feature.
- b is the bias (intercept).
- $\hat{y}^{(i)}$ is the predicted output.

2. Cost Function Equation

The cost function $J(w, b)$, often the Mean Squared Error (MSE), measures how well the linear regression model fits the data:

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

Where:

- $y^{(i)}$ is the actual value for the i-th data point.
- $\hat{y}^{(i)}$ is the predicted value.
- m is the number of data points.

3. Gradient Descent Algorithm (General Form)

Gradient descent is a general algorithm used to minimize a function $J(\theta)$ by iteratively updating the parameter θ in the direction of the negative gradient:

$$\theta := \theta - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta}$$

Where:

- θ represents any parameter of the model (e.g., w or b).
- α is the learning rate.
- $\frac{\partial J(\theta)}{\partial \theta}$ is the gradient of the cost function with respect to θ.

4. Gradients with Respect to w and b

The specific gradients of the cost function with respect to the parameters w and b are:

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

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

5. Update Rules for Parameters w and b

Finally, the gradient descent update rules for the parameters w and b are:

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

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

Substituting the gradients into these equations, we get:

$$w := w - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\cdot x^{(i)}$$

$$b := b - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)$$

These equations iteratively adjust the parameters w and b to minimize the cost function, ultimately leading to the optimal values for these parameters.