假设有以下三个坐标点：

- $(x_1, y_1) = (1, 2)$
- $(x_2, y_2) = (2, 4)$
- $(x_3, y_3) = (3, 6)$

1. 定义损失函数

选择均方误差（MSE）作为损失函数：

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - wx_i)^2$$

代入数据点，有：

$$L(w) = \frac{1}{3} \left[(2 - w \cdot 1)^2 + (4 - w \cdot 2)^2 + (6 - w \cdot 3)^2 \right]$$

2. 计算梯度

对 w 求损失函数的导数（梯度）：

$$\frac{\partial L(w)}{\partial w} = -\frac{2}{n} \sum_{i=1}^{n} x_i \cdot (y_i - wx_i)$$

代入数据点，有：

$$\frac{\partial L(w)}{\partial w} = -\frac{2}{3} \left[1 \cdot (2 - w \cdot 1) + 2 \cdot (4 - w \cdot 2) + 3 \cdot (6 - w \cdot 3) \right]$$

展开并整理：

$$\frac{\partial L(w)}{\partial w} = -\frac{2}{3} \left[2 - w + 8 - 4w + 18 - 9w \right]$$

$$\frac{\partial L(w)}{\partial w} = -\frac{2}{3} (28 - 14w)$$

$$\frac{\partial L(w)}{\partial w} = -\frac{56}{3} + \frac{28w}{3}$$

3. 梯度下降更新公式

使用梯度下降法更新 w：

$$w_{\text{new}} = w_{\text{old}} - \eta \cdot \frac{\partial L(w)}{\partial w}$$

其中 η 是学习率。

假设初始 $w = 0$，学习率 $\eta = 0.1$，计算第一步：

$$\frac{\partial L(0)}{\partial w} = -\frac{56}{3} + \frac{28 \cdot 0}{3} = -\frac{56}{3}$$

$$w_{\text{new}} = 0 - 0.1 \cdot \left(-\frac{56}{3} \right) = 0 + \frac{5.6}{3} \approx 1.87$$

继续迭代，重复上述步骤，最终 w 会逐步收敛到真实值 $w = 2$。

$$y = w_1 x_1 + w_2 x_2 + b$$

我们使用三个数据点来说明计算过程。假设数据点如下：

- $(x_1, x_2, y) = (1, 2, 5)$
- $(x_1, x_2, y) = (2, 3, 8)$
- $(x_1, x_2, y) = (3, 4, 11)$

1. 定义损失函数

我们使用均方误差 (MSE) 作为损失函数。对于每个数据点 (x_1, x_2, y)，损失函数为：

$$L(w_1, w_2, b) = \frac{1}{3} \sum_{i=1}^{3} \left(y_i - (w_1 x_{1i} + w_2 x_{2i} + b) \right)^2$$

带入这三个数据点，损失函数变成：

$$L(w_1, w_2, b) = \frac{1}{3} \left[(5 - (w_1 \cdot 1 + w_2 \cdot 2 + b))^2 + (8 - (w_1 \cdot 2 + w_2 \cdot 3 + b))^2 + (11 - (w_1 \cdot 3 + w_2 \cdot 4 + b))^2 \right]$$

2. 计算梯度

接下来我们对损失函数分别对 w_1, w_2, 和 b 求梯度。

对 w_1 求导:

$$\frac{\partial L(w_1, w_2, b)}{\partial w_1} = \frac{1}{3} \sum_{i=1}^{3} 2 \left(y_i - (w_1 x_{1i} + w_2 x_{2i} + b) \right) \cdot (-x_{1i})$$

代入数据点:

$$\frac{\partial L(w_1, w_2, b)}{\partial w_1} = \frac{1}{3} \left[2 \cdot (5 - (w_1 \cdot 1 + w_2 \cdot 2 + b)) \cdot (-1) + 2 \cdot (8 - (w_1 \cdot 2 + w_2 \cdot 3 + b)) \cdot (-2) + 2 \cdot (11 - (w_1 \cdot 3 + w_2 \cdot 4 + b)) \cdot (-3) \right]$$

对 w_2 求导:

$$\frac{\partial L(w_1, w_2, b)}{\partial w_2} = \frac{1}{3} \sum_{i=1}^{3} 2 \left(y_i - (w_1 x_{1i} + w_2 x_{2i} + b) \right) \cdot (-x_{2i})$$

代入数据点:

$$\frac{\partial L(w_1, w_2, b)}{\partial w_2} = \frac{1}{3} \left[2 \cdot (5 - (w_1 \cdot 1 + w_2 \cdot 2 + b)) \cdot (-2) + 2 \cdot (8 - (w_1 \cdot 2 + w_2 \cdot 3 + b)) \cdot (-3) + 2 \cdot (11 - (w_1 \cdot 3 + w_2 \cdot 4 + b)) \cdot (-4) \right]$$

对 b 求导:

$$\frac{\partial L(w_1, w_2, b)}{\partial b} = \frac{1}{3} \sum_{i=1}^{3} 2 \left(y_i - (w_1 x_{1i} + w_2 x_{2i} + b) \right) \cdot (-1)$$

代入数据点:

$$\frac{\partial L(w_1, w_2, b)}{\partial b} = \frac{1}{3} \left[2 \cdot (5 - (w_1 \cdot 1 + w_2 \cdot 2 + b)) \cdot (-1) + 2 \cdot (8 - (w_1 \cdot 2 + w_2 \cdot 3 + b)) \cdot (-1) + 2 \cdot (11 - (w_1 \cdot 3 + w_2 \cdot 4 + b)) \cdot (-1) \right]$$

3. 梯度下降更新

假设我们使用学习率 $\eta = 0.1$，并假设初始值 $w_1 = 0, w_2 = 0, b = 0$。我们可以通过梯度下降法更新参数。

更新 w_1:

$$w_1^{\text{new}} = w_1^{\text{old}} - \eta \cdot \frac{\partial L(w_1, w_2, b)}{\partial w_1}$$

更新 w_2:

$$w_2^{\text{new}} = w_2^{\text{old}} - \eta \cdot \frac{\partial L(w_1, w_2, b)}{\partial w_2}$$

更新 b:

$$b^{\text{new}} = b^{\text{old}} - \eta \cdot \frac{\partial L(w_1, w_2, b)}{\partial b}$$

4. 迭代

每次通过计算梯度并更新参数，w_1, w_2, 和 b 会逐渐收敛到一个最优值。通过多次迭代，模型会拟合数据点。

通过这个过程，梯度下降法将会逐步调整 w_1, w_2 和 b，最终使得损失函数最小化，从而获得最佳的线性模型。