**Gradient descent** is a name for a generic class of computer algorithms which minimize a function. These algorithms achieve this end by starting with initial parameter values and iteratively moving towards a set of parameter values that minimize some cost function or metric—that's the descent part. The movement toward best-fit is achieved by taking the derivative of the variable or variables involved, towards the direction with the lowest (calculus-defined) gradient—that's the gradient part.

Ordinary linear regression is a good and simple way of demonstrating how gradient descent works. We start with some error function. We could use any metric we want, but in OLS the obvious one is the residual sum of squares.

Given a sequence of points, $y_i$, and a sequence of points predicted by our model, $\hat{y}_i$, RSS is:

$$\text{error}_{(m,b)} = \sum_{i=1}^{n} (y_i - \hat{y_i})$$

Our objective is to minimize this value. Inserting our linear regression model in for the $\hat{y}_i$ predictions, and assuming (for the sake of simplicty) that we're doing regression on only one variable, we get:

$$\text{RSS} = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

Where $b$ is the intercept and $m$ is the slope of the line of best fit.

Now we need to take the gradient. Since this is an equation of two variables ($b$ and $m$) the gradient will consist of two partial derivatives. Hence the gradient is:

$$\left\langle \frac{\partial}{\partial b}(\text{RSS}), \frac{\partial}{\partial m}(\text{RSS}) \right\rangle = \left\langle -2\sum_{i=1}^{n}(y_i - (mx_i + b)), -2\sum_{i=1}^{n} x_i(y_i - (mx_i + b)) \right\rangle$$

To solve, take a step in the negative gradient direction every iteration. Eventually we will have something that converges.

Initially let m = 0 and b = 0. Let L be our learning rate. This controls how much the value of m changes with each step. L could be a small value like 0.0001 for good accuracy.
Calculate the partial derivative of the loss function with respect to m,b.

Now we update the current value of **m** and **b** using the following equation:

$$m = m - L * \frac{\partial}{\partial m}(\text{RSS}) \qquad b = b - L * \frac{\partial}{\partial b}(\text{RSS})$$