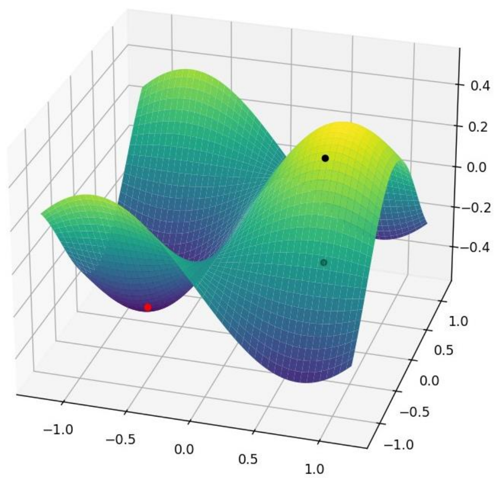
Gradient Descent Algorithm

Introduction

The gradient descent algorithm is a technique for optimizing various parameters and minimizing the cost function. It has two important parts: learning rate (α) and derivative.



Mathematics for this algorithm

Gradient Descent equation

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

θ_j is a parameter.

α is the learning rate.

$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ is the derivative of the cost function with respect to each parameter.

Learning rate

The learning rate is a parameter that controls how quickly the value changes. If the learning rate is large, the values will change quickly, and there may be large errors. If the learning rate is small, the values change slowly and there may be few errors.

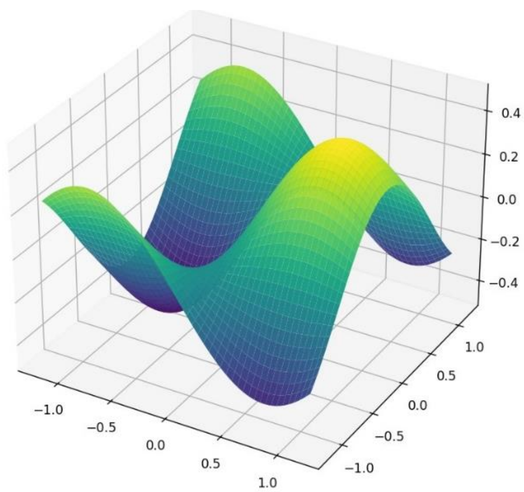
Partial derivative

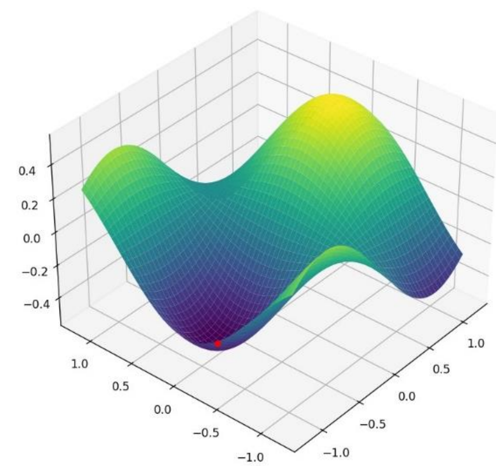
Each partial derivative is important for improving the direction leading to a smaller cost function.

Example

$$f(x, y) = \frac{(\sin 2x)(\cos 2y)}{2}$$

When $x = [-1.2,\ 1.2]$ and $y = [-1.2, 1.2]$





Using the Gradient Descent equation

$$x := x - \alpha \frac{\partial}{\partial f_x} f(x, y)$$

Find the partial derivative of this function with respect to x.

$$x := x - \alpha(\cos 2x \cos 2y)$$

Using the Gradient Descent equation

$$y := y - \alpha \frac{\partial}{\partial f_y} f(x, y)$$

Find the partial derivative of this function with respect to y.

$$y := y - \alpha(-\sin 2x \sin 2y)$$

By: Rati Kaewkam