1 Gradient Descent Tutorial

You may be familiar with functions in 2D space.

For example, we can have $f(x) = x^2$

PLOT THIS

We could add a dimension and have $f(x, y) = x^2 + y^2$

PLOT THIS

Where $f(x, y)$ is the function that takes in a point in 2D space and returns a value.

Some people label $f(x, y)$ as z

Recall what the derivative of $f(x)$ is. It is the slope of the tangent line at a point which is the rate of change of $f(x)$ with respect to x (how does $f(x)$ change as x changes)

How do we format this in the context of $f(x, y)$?

So now we have two questions to answer, how does $f(x, y)$ change as x changes and how does $f(x, y)$ change as y changes?

We can find this by taking the partial derivative of $f(x, y)$ with respect to x and y.

The partial derivative of $f(x, y)$ with respect to x is denoted as $\frac{\partial f}{\partial x}$ and the partial derivative of $f(x, y)$ with respect to y is denoted as $\frac{\partial f}{\partial y}$.

For the previous example, $f(x, y) = x^2 + y^2$, we have $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = 2y$.

We simply take the derivative of $f(x, y)$ with respect to x and y and treat the other variable as a constant.

$\frac{\partial f}{\partial x} = 2x$ means that as x changes, $f(x, y)$ changes at a rate of $2x$.

The gradient of $f(x, y)$ is the vector (a vector is an object with magnitude (size) and direction)of the partial derivatives of $f(x, y)$ with respect to x and y. It is denoted as $\nabla f(x, y)$.

Lets do another example, $f(x, y) = -x^2 - y^2 + 10$

$\frac{\partial f}{\partial x} = -2x$ and $\frac{\partial f}{\partial y} = -2y$

So the gradient of $f(x, y)$ is $\nabla f(x, y) = \begin{bmatrix} -2x \\ -2y \end{bmatrix}$

Lets plot this function

PLOT THIS

The gradient of $f(x, y)$ at a point (x, y) is the vector corresponding to the direction that would increase $f(x, y)$ the most at that point.

Key points:

The gradient of $f(x, y)$ at a point (x, y) is the vector corresponding to the direction that would increase $f(x, y)$ the most at that point.

A gradient is simply a vector of partial derivatives.

A partial derivative is the rate of change of a function with respect to one of its variables.