MARCH 14TH, 2025

Gradients

CE 311K - L27

Gradients

A **gradient** is a vector that represents the direction and rate of the steepest increase of a function

It generalizes the concept of derivatives to functions with multiple variables

$$\nabla f = \left[\frac{\delta f}{\delta x_1}, \frac{\delta f}{\delta x_2}, \ldots, \frac{\delta f}{\delta x_n}\right]$$

Gradients can be used for optimization to find minima/maxima i.e. the gradient is zero

Finite Difference

In **Finite Difference** we approximate the derivative by using a small step for each variable in *f(x)*

Hold the other variable values constant

Apply difference methods (forward, backward, or central) multiple times

---

$$\nabla f = \left[\frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h}, \frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h} \right]$$

Gradient Descent

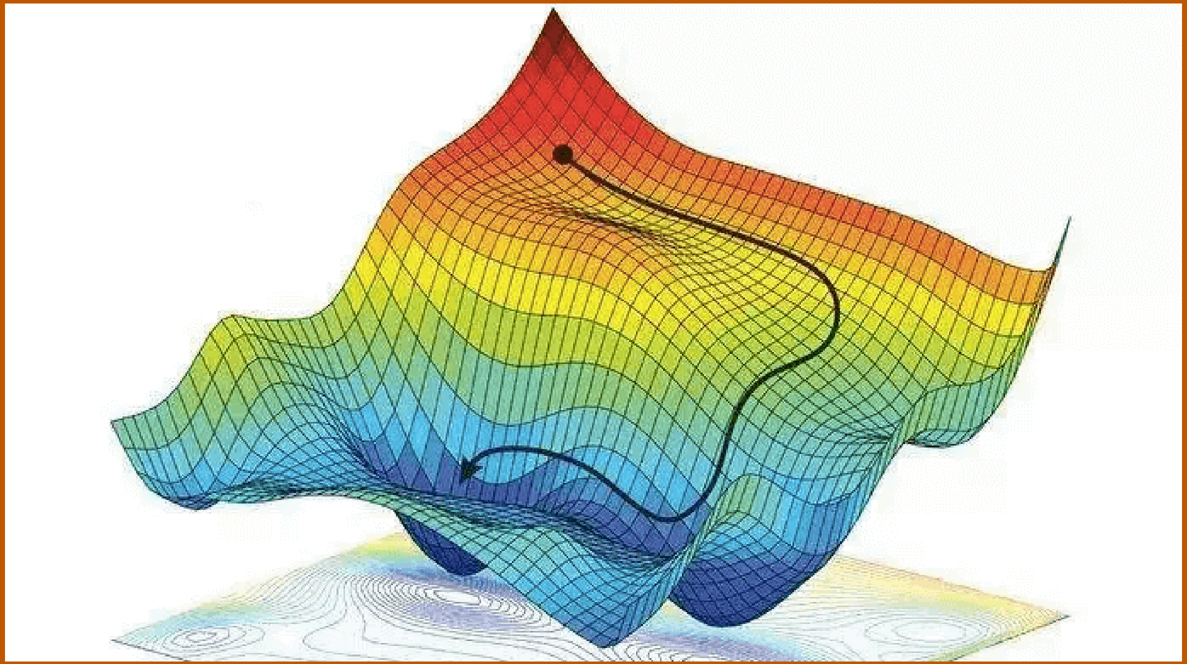
Gradient Descent finds the minimum by iteratively moving in the direction of the steepest descent

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

The **learning rate** is an important parameter the controls the step size

Too large: algorithm might diverge

Too small: convergence might be slow



Extensions of Gradient Descent

Stochastic Gradient Descent: Uses a random subset of data to compute the gradient at each step

Helps speed up convergence for larger datasets

Introduces "noise" that can help escape local minima

Momentum: Adds a fraction of the previous update to the current update to accelerate convergence

Maintain a **velocity vector** that "remembers" previous gradients

Adaptive Methods: Adjust the learning rate dynamically

Numerous methods to update learning rate