Understanding Gradient Descent Convergence:

When we use gradient descent, it's like we're hiking down a hill to find the lowest point. The slope of the hill is like the derivative of our cost function, telling us which way is downhill.

* We start with a point on the hill and calculate the slope (derivative).
* We take steps in the direction of the slope, with the step size determined by the learning rate.
* As we move downhill, the slope changes. It gets smaller as we get closer to the bottom.
* Eventually, we reach the bottom, where the slope becomes almost flat (close to 0).
* Our steps become smaller as we approach the bottom, helping us fine-tune our parameter.

**Role of Learning Rate (α):**

Picking the right learning rate is important:

* Too big: We might overshoot the bottom and keep going back and forth.
* Too small: Steps become tiny, taking a long time to reach the bottom.

**Bottom Line:**

Gradient descent adjusts steps based on the slope. As we approach the optimal parameter value, steps become smaller, helping us land at the best parameter value for our problem.