**🎓 Lecture: Momentum in Gradient Descent**

**1. The Problem: Local vs Global Minimum**

* Our goal in optimization is to find the **global minimum** (best possible solution).
* But sometimes, the algorithm gets stuck in a **local minimum** (a small dip in the loss function).
* Regular Gradient Descent (GD) and Stochastic Gradient Descent (SGD) are good, but they can fall into these dips and stop improving.

**2. Physics Analogy: Rolling a Ball**

* Imagine rolling a ball down a hill.
* **Without momentum:** The ball rolls slowly and may stop in a small dip (local minimum).
* **With momentum:** If the ball is rolling fast, it won’t stop at a small dip. It keeps going until it reaches a large valley (global minimum).

👉 Local minimum = small dip in the hill  
👉 Global minimum = big valley at the bottom

**3. Gradient Descent Formula (Without Momentum)**

So far, we’ve been using the update rule:

* : weight (parameter of the model)
* : learning rate
* : gradient of the loss

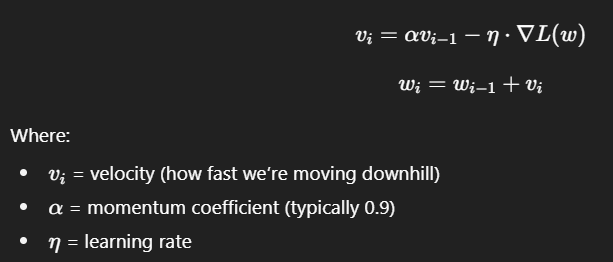
This rule updates weights based **only on the current gradient**.

**4. Adding Momentum**

To fix the local minimum problem, we introduce **momentum**.

Instead of only considering the current step, we also add information from the **previous update step**.

The new rule looks like this:



**5. How It Works**

* **If the ball rolled fast before**, momentum will push it further.
* **If it slowed down before**, momentum will be weaker.
* (usually 0.9) decides how much the **previous step matters**.

👉 With momentum, we move more smoothly and avoid getting stuck.  
👉 It’s like giving the algorithm a "push" to escape dips.

**6. Practical Notes**

* Momentum is a **hyperparameter technique**.
* We usually start with .
* It helps converge faster and more reliably than plain GD/SGD.

✅ **Summary:**

* Momentum is like rolling a ball downhill: it keeps moving past small dips.
* In gradient descent, we add a fraction of the previous update to the current update.
* Typical momentum value: **0.9**.
* This prevents the algorithm from getting stuck in local minima and helps reach the global minimum faster.