**🎓 Lecture: Visualizing the Learning Rate**

**1. Why Use a Graph?**

* Just talking about learning rate isn’t enough — we understand it better by **seeing it on a graph**.
* Graph = **Loss vs. Epochs** (error on the y-axis, training steps on the x-axis).
* A picture shows clearly how different learning rates behave.

**2. Case 1: Small Learning Rate**

* **Behavior:** Loss goes down steadily, but **very slowly**.
* Eventually reaches the goal, but it takes many epochs.
* Good side: stable and safe.
* Bad side: inefficient (slow progress).

**3. Case 2: High Learning Rate**

* **Behavior:**
  + At first, loss drops quickly.
  + But then it **oscillates** (jumps up and down around the minimum).
  + Sometimes, the loss even **explodes upward** instead of going down (diverges).
* Conclusion: Too high learning rate is unstable and may fail completely.

**4. Case 3: Well-Selected Learning Rate (e.g., Exponential Schedule)**

* **Behavior:**
  + Starts fast (like high learning rate).
  + Then gradually slows down (like low learning rate).
  + Minimizes the loss **quickly and accurately**.
* This is what we aim for in practice.

**5. The Real Problem**

* We **don’t know the best learning rate in advance**.
* The optimal rate depends on the dataset and the model.

**6. Practical Solution**

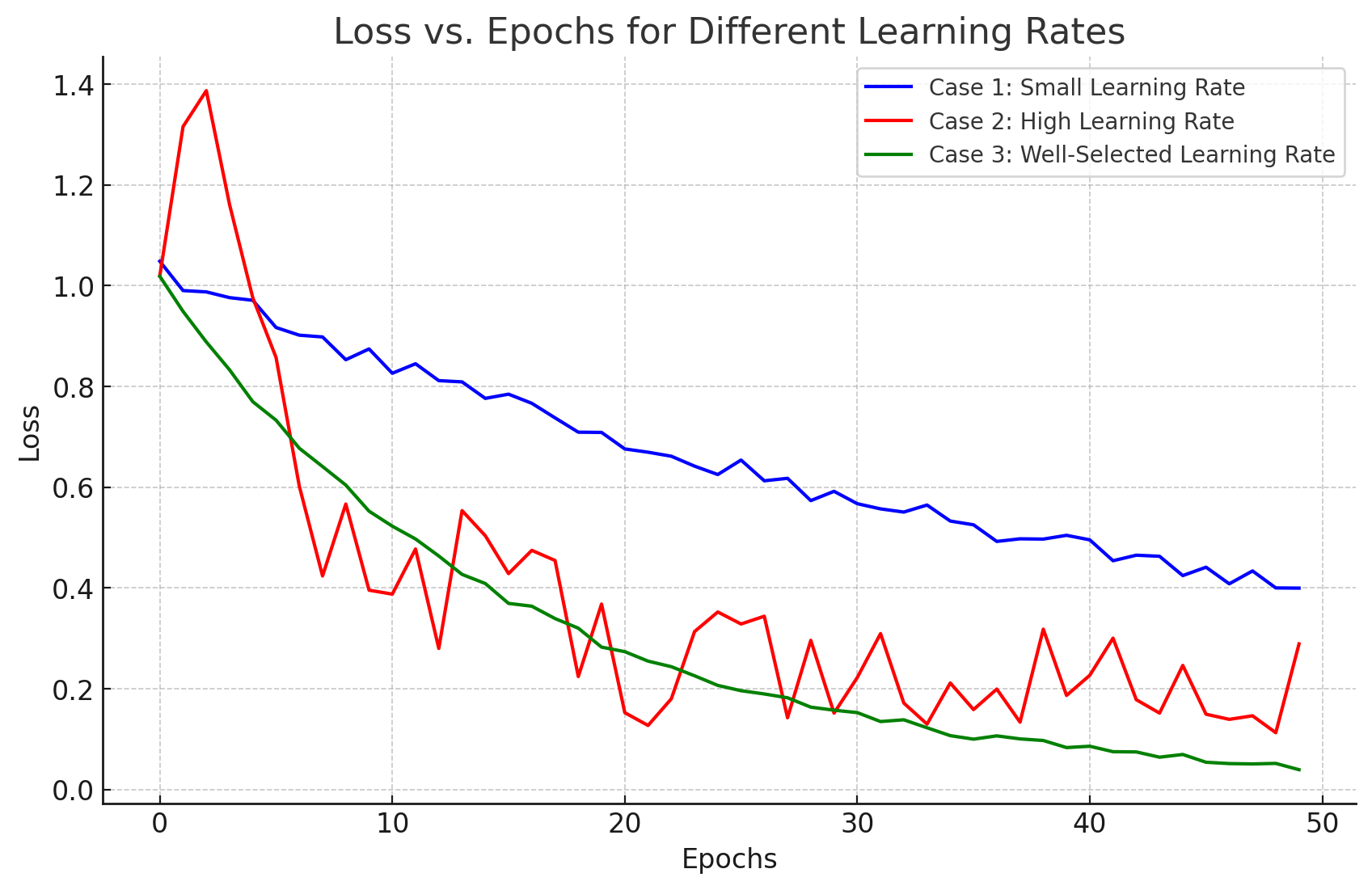
* Test multiple learning rates.
* Plot **Loss vs. Epochs** for each.
* Choose the one that looks most like the “good curve”:
  + Fast descent.
  + Smooth convergence.
  + No oscillation or divergence.

**✅ Summary**

* **Low learning rate** → safe, but slow.
* **High learning rate** → fast at first, but unstable (oscillations, divergence).
* **Good learning rate (with schedule)** → fast **and** accurate.
* To find the right one → **experiment and plot curves**.

👉 Simple analogy:  
It’s like adjusting the **water flow in a tap** 💧.

* If too little → water drips slowly (low LR).
* If too much → splashes everywhere (high LR).
* If adjusted well → steady stream, just right (good LR).



Here’s the graph of **Loss vs. Epochs** for the three cases:

* 🔵 **Small Learning Rate** → slow but steady decrease.
* 🔴 **High Learning Rate** → oscillates and can diverge.
* 🟢 **Well-Selected Learning Rate (Schedule)** → fast drop and smooth convergence.