**Topic: Gradient Descent and Its Limitations**

1. **Recap of GD and SGD**
   * **Gradient Descent (GD)**: Updates weights after evaluating the **entire dataset** (batch).
     + Pros: Moves steadily toward a minimum.
     + Cons: Slow for large datasets.
   * **Stochastic Gradient Descent (SGD)**: Updates weights **after each data point** or small batch.
     + Pros: Faster computation, explores more points.
     + Cons: Converges approximately, may not reach the exact minimum.
2. **Illustration with Loss Function**
   * Imagine the graph of a **loss function**.
   * GD starts at a point and slowly descends toward a minimum.
   * SGD moves more dynamically, covering more points quickly but may be “bouncy” around the minimum.
3. **The Problem of Local Minima**
   * Real-world loss functions are **irregular**.
   * GD can get stuck in a **local minimum**, which is a suboptimal solution.
   * The **global minimum** is the point we truly want, but GD might miss it depending on the **starting point**.
4. **Role of Learning Rate**
   * A **higher learning rate** can help skip local minima and reach a global minimum.
   * Risk: Too high → oscillations, never settling.
5. **Key Insight**
   * GD is not “almighty”; it can fail in complex landscapes.
   * Solutions exist (to be discussed in the next lecture) to help GD escape local minima.

**Visual Concept (Simplified)**

****

 \* marks the global minimum.

 GD might stop at the first “dip” it encounters.

 SGD may bounce around, exploring multiple minima.