Question 1: Learning Rate Scheduling in Gradient Descent In the lecture on Gradient Descent, we learned about the importance of the learning rate in optimizing neural networks. However, using a fixed learning rate can sometimes lead to suboptimal results. Research and describe two learning rate scheduling techniques (e.g., Learning Rate Decay, Cyclical Learning Rates, or Step Decay). Explain how these techniques adjust the learning rate during training, their benefits, and in what scenarios they are most effective

Learning Rate Scheduling in Gradient Descent

A fixed learning rate in gradient descent can lead to issues such as **slow convergence, overshooting the optimal point, or getting stuck in local minima**. Learning rate scheduling techniques dynamically adjust the learning rate during training to improve performance. Below are two widely used techniques:

1. Learning Rate Decay (Exponential Decay)

How It Works:

- The learning rate **gradually decreases** over time according to an **exponential function**:

Benefits:
✓ Prevents **overshooting** the optimal solution.
✓ Helps fine-tune the model as training progresses.
✓ Reduces oscillations in later stages of training.

Best Used When:

- Training deep networks where a **high initial learning rate** helps in faster convergence, but **smaller rates** improve final accuracy.

- Tasks where training is **long** and requires **gradual fine-tuning** (e.g., NLP, image classification).

---

2. Cyclical Learning Rate (CLR)

How It Works:

- Instead of continuously decreasing, the learning rate **oscillates** between a lower and upper bound.

- It follows a **triangular** or **cosine** wave pattern.

Benefits:
✓ Helps escape **local minima** by **increasing** the learning rate periodically.
✓ Improves **generalization** by preventing overfitting.
✓ Often eliminates the need for fine-tuning the learning rate manually.

Best Used When:

- Training **complex models** where fixed learning rates fail to reach an optimal solution.
- Cases where the loss function has **many local minima**, such as **reinforcement learning** or **deep vision models**.

Conclusion

Both techniques aim to optimize learning rates dynamically, preventing inefficient training**.** **Exponential Decay** is great for long training processes that require fine-tuning, while **Cyclical Learning Rates** help models escape local minima and improve generalization. The choice depends on the model and problem being solved.