PreLab:-

1.**What is the fundamental concept behind Stochastic Gradient Descent, and how does it differ from batch gradient descent?**
Stochastic Gradient Descent (SGD) updates the model's parameters using the gradient from a single data point, leading to faster but noisier updates compared to batch gradient descent.

2.**How is the learning rate utilized in the context of Stochastic Gradient Descent, and what considerations should be made in its selection?**
The learning rate controls the step size of parameter updates; it must be carefully selected to balance convergence speed and stability, as too high a rate can cause divergence.

3.**What is mini-batch gradient descent, and how does it combine aspects of both batch and stochastic gradient descent? What are the benefits of using mini batches?**
Mini-batch gradient descent combines aspects of both batch and stochastic methods by updating the model using a small subset (mini-batch) of the dataset, offering a trade-off between computational efficiency and update stability.

4.**Compare the convergence behavior of Stochastic Gradient Descent with batch gradient descent. In what scenarios might one converge faster than the other?**
SGD often converges faster in large datasets due to more frequent updates but may oscillate around the minimum, whereas batch gradient descent is more stable but slower.

VIVA:-

1.**Define Stochastic Gradient Descent:**
Stochastic Gradient Descent (SGD) updates model parameters iteratively using the gradient from a single data point, allowing for faster updates but introducing variability in the optimization process.

2.**How might a small or large mini-batch size impact the convergence and generalization of Stochastic Gradient Descent?**
Small mini-batch sizes provide noisy but frequent updates, potentially leading to better generalization, while large sizes offer more stable updates but can slow convergence and risk overfitting.

3.**Are there strategies or algorithms that adaptively adjust the learning rate during the training process in Stochastic Gradient Descent?**
Yes, algorithms like AdaGrad, RMSprop, and Adam adjust the learning rate adaptively during training to improve convergence and stability in SGD.

4.**SGD can provide noisy estimates of the gradient due to the use of individual or small subsets of examples. How does this noise impact the optimization process, and how might it be addressed?**
The noise can help escape local minima but may cause oscillations; this can be addressed by techniques like learning rate decay, momentum, or using mini-batches.

5.**In what scenarios would you prefer using Stochastic Gradient Descent over batch gradient descent?**

SGD is preferred in large datasets where faster convergence and reduced memory usage are critical, especially when computational efficiency outweighs the need for precise updates.