Gradient Descent Algorithm

Stochastic Gradient Descent (SGD) and Batch Gradient Descent (BGD) are both optimization algorithms used to minimize a loss function in machine learning.

Stochastic Gradient Descent (SGD):

- **Updates Weights**: Uses one training example at a time to update the model parameters.
- **Convergence Speed**: Generally faster but can have high variance in the updates.
- **Memory Usage**: Low, as it processes one example at a time.
- **Convergence Behavior**: Can oscillate and may not converge as smoothly as BGD, but often finds good solutions faster.

Batch Gradient Descent (BGD):

- **Updates Weights**: Uses the entire training dataset to compute the gradient and update the model parameters.
- **Convergence Speed**: Can be slow because it processes the entire dataset before each update.
- **Memory Usage**: High, as it requires the entire dataset to be in memory.
- **Convergence Behavior**: More stable and smoother convergence but can be computationally expensive for large datasets.