1. Batch Gradient Descent

- Computes the gradient using the entire dataset before updating the weights.
- **Pros**: More stable convergence, precise updates.
- **Cons**: Slow with large datasets, requires more memory.

2. Stochastic Gradient Descent (SGD)

- Updates the weights after computing the gradient for each individual data sample.
- **Pros**: Faster updates, works well for large datasets, can escape local minima.
- **Cons**: High variance in updates, can be noisy, may not converge smoothly.

3. Mini-Batch Gradient Descent

- A mix between **Batch GD and SGD**, where the dataset is split into **small batches**, and the gradient is computed for each batch.
- **Pros**: Balances stability and speed, reduces noise compared to SGD.
- **Cons**: Requires tuning batch size, may still have some variance.

4. Momentum-based Gradient Descent

- Uses the idea of **momentum** to accelerate learning by considering past gradients.
- **Pros**: Faster convergence, helps avoid oscillations.
- **Cons**: Needs careful tuning of momentum factor.

5. Nesterov Accelerated Gradient (NAG)

- A variation of **Momentum GD** that **predicts the next step** before computing the gradient.
- **Pros**: More stable updates, reduces overshooting.
- **Cons**: Slightly more complex calculations.