|  |
| --- |
| Ten Types of Optimizers in Machine Learning |
| | **No.** | **Optimizer** | **One-line Explanation** | | --- | --- | --- | | 1 | Gradient Descent (GD) | Updates parameters by moving in the direction of the negative gradient of the loss. | | 2 | Stochastic Gradient Descent (SGD) | Updates parameters using one sample (or mini-batch) at a time for faster but noisier convergence. | | 3 | Mini-Batch Gradient Descent | Combines advantages of GD and SGD by updating using small random batches of data. | | 4 | Momentum Optimizer | Accelerates gradient descent by adding a fraction of the previous update to the current one. | | 5 | Nesterov Accelerated Gradient (NAG) | Similar to Momentum but looks ahead before computing the gradient for better accuracy. | | 6 | Adagrad | Adapts learning rate for each parameter based on how frequently it’s updated. | | 7 | RMSProp | Uses a moving average of squared gradients to normalize the learning rate dynamically. | | 8 | Adam (Adaptive Moment Estimation) | Combines Momentum and RMSProp for adaptive learning with first and second moment estimates. | | 9 | AdaDelta | An improved Adagrad that limits the accumulated past gradients to avoid vanishing learning rates. | | 10 | AdamW | A variant of Adam that decouples weight decay from gradient updates for better generalization. | |