ML Optimizers: 10 Essential Types

Abdelrahman Hamdy Mustafa Ali El-Hamshary | 202002570

This list details ten types of optimizers used in Machine Learning, each with a brief functional explanation:

1. **Gradient Descent (GD):** A core optimization algorithm that adjusts model parameters repeatedly to minimize the loss function by moving along the negative gradient direction[2].
2. **Stochastic Gradient Descent (SGD):** An iteration of GD that speeds up convergence by updating parameters based on a randomly selected subset (mini-batch) of the training data[3].
3. **Mini-batch Gradient Descent:** This method balances speed and stability by using small batches of data for parameter updates, combining elements of both GD and SGD[4].
4. **Adagrad:** Adapts the learning rate individually for each parameter based on its past gradients, making it particularly useful for sparse data applications[5].
5. **RMSprop:** An adaptive learning rate method that divides the gradient by a running average of the magnitudes of recent gradients. It's well-suited for non-stationary objectives[6].
6. **Adam (Adaptive Moment Estimation):** A highly effective optimizer that integrates both adaptive learning rates and momentum for fast convergence and efficient training[7].
7. **AdaDelta:** Extends Adagrad by restricting the accumulation of past squared gradients, thereby preventing the learning rate from decaying too aggressively over time[8].
8. **Nadam:** A variant of the Adam optimizer that incorporates Nesterov momentum, which can lead to potentially faster model convergence[9].
9. **Momentum:** Accelerates the SGD process by adding a portion of the prior update to the current parameter update, significantly boosting convergence speed[10].
10. **Newton's Method:** Finds optimal parameters by approximating the curvature of the loss surface using second-order derivatives (the Hessian matrix). This approach is generally more computationally costly[11].