Abelrahman Mohamed
202004789

AI Assignment

1. **Gradient Descent (GD):**
   Gradually updates model parameters by moving opposite to the gradient direction to minimize the loss function.
2. **Stochastic Gradient Descent (SGD):**
   A faster variant of GD that updates parameters using a single random data sample at each step, introducing more variability but speeding up learning.
3. **Mini-Batch Gradient Descent:**
   A hybrid between GD and SGD that processes small random batches of data, providing a good balance between efficiency and stability.
4. **Adagrad (Adaptive Gradient):**
   Adjusts the learning rate for each parameter individually based on the history of past gradients, making it effective for sparse data.
5. **RMSprop (Root Mean Square Propagation):**
   Modifies learning rates adaptively by dividing the gradient by a moving average of recent squared gradients, ideal for non-stationary problems.
6. **Adam (Adaptive Moment Estimation):**
   Combines the benefits of momentum and RMSprop by tracking both the average and variance of gradients, offering quick and reliable convergence.
7. **Adadelta:**
   An extension of Adagrad that restricts the influence of past gradients to a fixed-size window, ensuring more stable and consistent learning rates.
8. **Nadam (Nesterov-accelerated Adam):**
   Builds upon Adam by adding Nesterov momentum, allowing the optimizer to anticipate updates and converge more smoothly.
9. **Momentum:**
   Speeds up convergence by adding a fraction of the previous update to the current one, helping the model move past small local minima.
10. **FTRL (Follow-The-Regularized-Leader):**
    Commonly used in large-scale online learning, it combines adaptive learning rates with regularization to handle sparse and high-dimensional data effectively.