Non-programming Assignment

1.What is learning rate decay and why is it needed?

Learning rate decay is a technique where the learning rate decreases gradually as training progresses. It is needed because initially, a higher learning rate helps the model converge faster, but as the model approaches an optimal solution, a lower learning rate allows for finer adjustments and avoids overshooting. This improves convergence and ensures the model finds a better minimum.

2.What are saddle and plateau problems?

Saddle points are points in the loss surface where the gradient is zero, but it's not a local minimum or maximum. They can trap the optimization process and slow down learning.

Plateau problems occur when the loss function flattens, meaning the gradients are very small or nearly zero, causing the model to make little progress during training. Both saddle points and plateaus can cause the training process to get stuck or slow down.

3.Why should we avoid the grid approach in hyperparameter choice?

Grid search involves exhaustively searching through a fixed set of hyperparameters, which can be very inefficient and time-consuming, especially for large neural networks with multiple hyperparameters. Instead, approaches like random search or Bayesian optimization can sample hyperparameters more efficiently, reducing the number of trials and exploring more promising regions of the search space.

4.What is a mini batch and how is it used?

Mini batch is a subset of the training data used to update model weights in each iteration of training. Instead of updating weights after every single training sample (stochastic gradient descent) or after processing the entire dataset (batch gradient descent), mini-batch gradient descent updates weights after processing a small random batch of data (e.g., 32 or 64 samples). This approach strikes a balance between efficiency and convergence speed, and helps smooth out the noise introduced by stochastic updates.