What are the two primary aspects controlling the capacity of a neural network?

A. Number of epochs and batch size

B. Learning rate and loss function

C. Number of nodes and number of layers

D. Weight initialization and activation function

Answer: C

Which gradient descent approach uses a batch size of one?

A. Batch Gradient Descent

B. Minibatch Gradient Descent

C. Stochastic Gradient Descent

D. Momentum Gradient Descent

Answer: C

What is the purpose of a learning rate schedule?

A. To maintain a constant learning rate throughout training.

B. To increase the learning rate exponentially over time.

C. To vary the learning rate over the training process.

D. To randomly change the learning rate during training.

Answer: C

In a regression problem, what is the commonly used loss function?

A. Cross-Entropy

B. Hinge Loss

C. Mean Squared Error (MSE)

D. Binary Cross-Entropy

Answer: C

What technique is used to address exploding gradients during training?

A. ReLU function

B. Gradient clipping or gradient normalization

C. Dropout

D. Early Stopping

Answer: B