**1. What is the difference between TRAINABLE and NON-TRAINABLE PARAMETERS?**

**Ans:** Trainable parameters are the weights and biases in a neural network that are updated during the training process to minimize the loss function. These parameters are learned from the data and directly affect the model's predictions.

Non-trainable parameters are fixed values or transformations applied to the data that do not change during training. Examples include hyperparameters like activation functions, dropout rates, and regularization parameters, as well as pre-trained weights in transfer learning.

**2. In the CNN architecture, where does the DROPOUT LAYER go?**

**Ans:** Dropout layers are typically placed after the fully connected layers in a CNN architecture. They randomly deactivate a fraction of neurons during training to prevent overfitting by introducing noise and encouraging the network to learn more robust features.

**3. What is the optimal number of hidden layers to stack?**

**Ans:** There is no fixed rule for the optimal number of hidden layers to stack in a neural network architecture. It depends on factors such as the complexity of the task, the amount of available data, and the computational resources. Experimentation and empirical testing are often necessary to determine the most effective architecture for a specific problem.

**4. In each layer, how many secret units or filters should there be?**

**Ans:** The number of units or filters in each layer depends on the complexity of the problem, the size of the input data, and the desired model capacity. Generally, deeper layers may require more filters to capture increasingly abstract features, while shallower layers may have fewer filters to preserve spatial information.

**5. What should your initial learning rate be?**

**Ans:** The initial learning rate should be chosen based on empirical testing and validation performance. It should be set at a level that allows for stable training progress without causing the loss function to diverge or oscillate excessively. Common initial learning rates range from 0.001 to 0.1.

**6. What do you do with the activation function?**

**Ans:** The activation function introduces non-linearity into the neural network, allowing it to learn complex relationships and patterns in the data. Common activation functions include ReLU (Rectified Linear Unit), sigmoid, and tanh functions, each with different properties suited to different tasks and architectures.

**7. What is NORMALIZATION OF DATA?**

**Ans:** Normalization of data involves scaling and centering the input features to ensure they have a consistent scale and distribution. This helps stabilize the training process, prevent vanishing or exploding gradients, and improve the convergence of optimization algorithms.

**8. What is IMAGE AUGMENTATION and how does it work?**

**Ans:** Image augmentation is a technique used to artificially increase the size and diversity of a training dataset by applying various transformations to the input images, such as rotation, flipping, scaling, and cropping. This helps improve the model's generalization ability and robustness to variations in the input data.

**9. What is DECLINE IN LEARNING RATE?**

**Ans:** Decline in learning rate refers to the process of gradually reducing the learning rate during training to fine-tune the model's parameters towards convergence. This helps prevent overshooting or oscillation around the optimal solution and improves the stability of the optimization process.

**10.What does EARLY STOPPING CRITERIA mean?**

**Ans:** Early stopping criteria is a technique used to prevent overfitting by monitoring the performance of the model on a validation dataset during training. Training is halted when the validation performance starts to degrade or plateau, indicating that further training may lead to overfitting on the training data.