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

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

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

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

5. What should your initial learning rate be?

6. What do you do with the activation function?

7. What is NORMALIZATION OF DATA?

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

9. What is DECLINE IN LEARNING RATE?

10. What does EARLY STOPPING CRITERIA mean?

Answer:

1. Trainable parameters are the parameters in a neural network that are updated during training to minimize the loss function. They include weights and biases in the network layers. Non-trainable parameters are those that are not updated during training, such as hyperparameters and fixed layer parameters like batch normalization mean and variance.
2. The Dropout layer can be added after the fully connected layers or convolutional layers in a CNN architecture.
3. The optimal number of hidden layers in a neural network depends on the complexity of the problem being solved, the amount of data available, and other factors. There is no one-size-fits-all answer, but in general, adding more layers can help to capture complex features and patterns in the data.
4. The number of units or filters in each layer depends on the size of the input, the complexity of the problem, and other factors. There is no fixed number, but in general, deeper and wider networks tend to perform better than shallow and narrow ones.
5. The initial learning rate should be chosen carefully based on the problem being solved, the optimizer being used, and other factors. A good starting point is often a value between 0.001 and 0.1.
6. The choice of activation function can have a significant impact on the performance of a neural network. Common choices include ReLU, sigmoid, and tanh. It's important to choose an activation function that is appropriate for the problem being solved and the architecture of the network.
7. Normalization of data refers to the process of scaling the input data to have zero mean and unit variance. This can help to improve the performance of a neural network by making the optimization process more stable and efficient.
8. Image augmentation is a technique used to artificially increase the size of a training dataset by generating new versions of the original images through transformations like rotation, translation, and flipping. This can help to prevent overfitting and improve the generalization performance of the network.
9. The learning rate decay refers to the strategy of gradually reducing the learning rate over time during training. This can help to fine-tune the network and improve its performance.
10. Early stopping criteria refers to the strategy of stopping the training process early based on some pre-defined criteria, such as when the validation loss stops decreasing or when the accuracy on the validation set stops improving. This can help to prevent overfitting and improve the generalization performance of the network.