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?

What does EARLY STOPPING CRITERIA mean?