1. Why don't we start all of the weights with zeros?

**Ans : Initializing all the weights with zeros leads the neurons to learn the same features during training. In fact, any constant initialization scheme will perform very poorly. Thus, both neurons will evolve symmetrically throughout training, effectively preventing different neurons from learning different things.**

2. Why is it beneficial to start weights with a mean zero distribution?

**Ans : If you initialize all weights with zeros then every hidden unit will get zero independent of the input. So, when all the hidden neurons start with the zero weights, then all of them will follow the same gradient and for this reason "it affects only the scale of the weight vector, not the direction".**

3. What is dilated convolution, and how does it work?

**Ans : It is a technique that expands the kernel (input) by inserting holes between the its consecutive elements. In simpler terms, it is same as convolution but it involves pixel skipping, so as to cover a larger area of the input. The more important point is that the architecture is based on the fact that dilated convolutions support exponential expansion of the receptive field without loss of resolution or coverage. Allows one to have larger receptive field with same computation and memory costs while also preserving resolution.**

4. What is TRANSPOSED CONVOLUTION, and how does it work?

**Ans : Transposed convolutions are standard convolutions but with a modified input feature map. The stride and padding do not correspond to the number of zeros added around the image and the amount of shift in the kernel when sliding it across the input, as they would in a standard convolution operation.**

5.Explain Separable convolution

**Ans : The spatial separable convolution is so named because it deals primarily with the spatial dimensions of an image and kernel: the width and the height. (The other dimension, the “depth” dimension, is the number of channels of each image). A spatial separable convolution simply divides a kernel into two, smaller kernels.**

6.What is depthwise convolution, and how does it work?

**Ans : Depthwise Convolution is a type of convolution where we apply a single convolutional filter for each input channel. In the regular 2D convolution performed over multiple input channels, the filter is as deep as the input and lets us freely mix channels to generate each element in the output.**

7.What is Depthwise separable convolution, and how does it work?

**Ans : Unlike spatial separable convolutions, depthwise separable convolutions work with kernels that cannot be “factored” into two smaller kernels. The depthwise separable convolution is so named because it deals not just with the spatial dimensions, but with the depth dimension the number of channels as well.**

8.Capsule networks are what they sound like.

**Ans : A Capsule Neural Network (CapsNet) is a machine learning system that is a type of artificial neural network (ANN) that can be used to better model hierarchical relationships. The approach is an attempt to more closely mimic biological neural organization.**

9. Why is POOLING such an important operation in CNNs?

**Ans : Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. The pooling layer summarises the features present in a region of the feature map generated by a convolution layer.**

10. What are receptive fields and how do they work?

**Ans : The receptive field encompasses the sensory receptors that feed into sensory neurons and thus includes specific receptors on a neuron as well as collectives of receptors that are capable of activating a neuron via synaptic connections.**