1. What is the COVARIATE SHIFT Issue, and how does it affect you?

Ans. Covariate shift occurs when the distribution of variables in the training data is different to real-world or testing data. This means that the model may make the wrong predictions once it is deployed, and its accuracy will be significantly lower.

2. What is the process of BATCH NORMALIZATION?

Ans. it is a process to make neural networks faster and more stable through adding extra layers in a deep neural network. The new layer performs the standardizing and normalizing operations on the input of a layer coming from a previous layer.

3. Using our own terms and diagrams, explain LENET ARCHITECTURE.

Ans. In general, LeNet refers to LeNet-5 and is a simple convolutional neural network. Convolutional neural networks are a kind of feed-forward neural network whose artificial neurons can respond to a part of the surrounding cells in the coverage range and perform well in large-scale image processing.

4. Using our own terms and diagrams, explain ALEXNET ARCHITECTURE.

Ans. AlexNet was the first convolutional network which used GPU to boost performance. 1. AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer.

5. Describe the vanishing gradient problem.

Ans. The vanishing gradient problem requires us to use small learning rates with gradient descent which then needs many small steps to converge. This is a problem if you have a slow computer which takes a long time for each step. If you have a fast GPU which can perform many more steps in a day, this is less of a problem.

6. What is NORMALIZATION OF LOCAL RESPONSE?

Ans. "The local response normalization layer performs a kind of “lateral inhibition” by normalizing over local input regions. In ACROSS_CHANNELS mode, the local regions extend across nearby channels, but have no spatial extent (i.e., they have shape local_size x 1 x 1).

7. In AlexNet, what WEIGHT REGULARIZATION was used?

Ans. The regularization used in this network is L2 with a weight decay of 5e-4. It was trained on GTX580 GPU which contains 3GB of memory. It has an error rate of 16.4 in the ImageNet Large Scale Visual Recognition Challenge(ILSVRC).

8. Using our own terms and diagrams, explain VGGNET ARCHITECTURE.

Ans. VGG- Network is a convolutional neural network model proposed by K. Simonyan and A. Zisserman in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition” [1]. This architecture achieved top-5 test accuracy of 92.7% in ImageNet, which has over 14 million images belonging to 1000 classes.

9. Describe VGGNET CONFIGURATIONS.

Ans. To reduce the number of parameters, authors propose to use a small respective field to replace large one. Authors conclude:

- Incorporate multiple non-linear rectification layers instead of a single rectification layer are more discriminative.
- It helps to decrease the number of parameters while keeping performance. For example, using 2 layers of 3x3 filter is equal to 1 layer of 5x5 filter but using fewer parameters. The number of a parameter is reduced by 28% ((25–18)/25). For

Number of Parameters of 2 Layers of 3x3 Filter: 2x3x3 = 18

Number of Parameters of 1 Layer of 5x5 Filter: 1x5x5 = 25

If you want to further understand how stacking small layers perform better than a single large layer, you may check out this story.

Simonyan et al. initialized 6 different ConvNet to see the performance of stacking layers. The difference is the number stacking layer within the same blocks. For example, VGG-11 (i.e Config A) uses 2 Conv3–256 layers while VGG-19 (i.e. Config E) uses 4 Conv3–256 layers in the third layer of blocks.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Experiment Architecture of VGGNet (Simonyan et al., 2014)

10. What regularization methods are used in VGGNET to prevent overfitting?

Ans. L2 and L1 Regularization -L2 and L1 are the most common types of regularization. Regularization works on the premise that smaller weights lead to simpler models which in results helps in avoiding overfitting.