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

**Ans:** Covariate shift refers to the situation where the distribution of input features changes between the training and testing datasets. This can lead to degraded model performance because the model has been trained on data that does not fully represent the testing environment. Covariate shift affects the generalization ability of the model, as it may fail to adapt to new or unseen data distributions during inference.

**2. What is the process of BATCH NORMALIZATION?**

**Ans:** Batch normalization is a technique used to normalize the activations of each layer in a neural network by adjusting and scaling them to have a zero mean and unit variance. This helps stabilize and accelerate the training process by reducing internal covariate shift and ensuring that the network learns more robust features. The process involves computing the mean and standard deviation of each mini-batch during training and using these statistics to normalize the activations. Additionally, learnable scale and shift parameters are introduced to allow the network to learn the optimal mean and variance for each layer.

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

**Ans:** LeNet architecture, proposed by Yann LeCun et al., is one of the pioneering convolutional neural network (CNN) architectures designed for handwritten digit recognition. It consists of two sets of convolutional and pooling layers followed by fully connected layers.

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

**Ans:** AlexNet architecture, proposed by Alex Krizhevsky et al., is a deep convolutional neural network designed for image classification tasks. It consists of five convolutional layers followed by max-pooling layers, three fully connected layers, and a softmax output layer.

**5. Describe the vanishing gradient problem.**

**Ans:** The vanishing gradient problem occurs during the training of deep neural networks when the gradients of the loss function with respect to the parameters diminish exponentially as they propagate backward through the network. This phenomenon makes it difficult for the lower layers of the network to learn meaningful representations, leading to slow convergence and poor performance.

**6. What is NORMALIZATION OF LOCAL RESPONSE?**

**Ans:** Normalization of local response, also known as local response normalization (LRN), is a technique used in convolutional neural networks to enhance the contrast between neurons' responses and improve generalization. It involves normalizing the activation of a neuron by the responses of its neighboring neurons within the same feature map. This helps prevent overfitting by encouraging competition between neurons and promoting sparse activations.

**7. In AlexNet, what WEIGHT REGULARIZATION was used?**

**Ans:** In AlexNet, weight regularization techniques such as L2 regularization (weight decay) were used to penalize large weights and prevent overfitting. This regularization term was added to the loss function during training to encourage smaller and more robust parameter values.

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

**Ans:** VGGNet architecture, proposed by Karen Simonyan and Andrew Zisserman, is a deep convolutional neural network known for its simplicity and effectiveness. It consists of multiple sets of convolutional layers followed by max-pooling layers, with three fully connected Layers at the end.

**9. Describe VGGNET CONFIGURATIONS.**

**Ans** : VGGNet is characterized by its uniform architecture with multiple sets of convolutional layers, each followed by max-pooling layers. The configurations are denoted by the notation "VGGx," where x represents the number of convolutional layers. Common configurations include VGG16 and VGG19, which have 16 and 19 convolutional layers, respectively.

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

**Ans:** VGGNet uses dropout regularization, where a fraction of neurons are randomly deactivated during training to prevent co-adaptation and overfitting. Additionally, weight decay (L2 regularization) is applied to the network's weights to penalize large parameter values and encourage simpler models. These regularization techniques help improve the generalization ability of VGGNet and prevent overfitting on the training data.