1. Difference between a neural network with all the neurons in a single layer vs neurons stacked in multiple layers?
2. In a neural network with multiple neurons in a single layer or the dense neural network, the network is able to learn the intricate representations of the input data taken from the previous layer.
3. The model complexity is limited by the number of neurons available in the single layer. It is simple to design; it is computationally effective and it is easily interpretable.
4. In a neural network with distributed neurons which is also called a deep neural network; the layers are arranged in a way to process the data sequentially; it is able to learn abstract and the complex representations of the data.
5. The model is useful in hierarchical feature extraction; we can increase the model capacity by stacking multiple layers over the previous layers giving improved performance; it helps in generalization tasks.
6. Difference between Generative and Discriminative neural networks?
7. A generative model aims to understand and model the underlying distribution of the input data in order to generate new samples that resemble the original distribution.
8. These models learn from the joint probability distribution of the input data and labels. Given this learned distribution it can generate new samples that exhibit similar characteristics to the training data.
9. It can be used for data generation, unconditional and conditional generation of data.
10. Discriminative models focus on learning the decision boundary between different classes or categories of the input data.
11. The model learns the conditional probability of the labels given the input data.
12. These models aim to directly lean the conditional between classes rather than generating the new samples.
13. It can be used for making class-specific decision making, object detection and segmentation.
14. These models have generally simple architectures and are computationally efficient.
15. What is cross entropy loss?
    1. It is a commonly used loss function in machine learning, particularly in classification tasks.
    2. It quantifies the dissimilarity between the predicted probability distribution and the true label distribution of the training data.
    3. It measures the average number if bits needed to encode the true labels based on the predicted probabilities.
    4. By minimizing the cross-entropy loss, the model is encouraged to assign higher probabilities to correct classes and lower probabilities to the incorrect classes.
16. What is KL divergence?
17. KL divergence short for Kullaback-Leibler divergence, is a measure of how one probability distribution differs from another.
18. It is often used in information theory and machine learning to quantify the dissimilarity between two probability distribution.
19. KL divergence is not a symmetric measure, meaning the KL divergence from distribution A to distribution B is not the same as the KL divergence from B to A.
20. KL divergence is always non-negative, it is not symmetric, it not a true distance metric.
21. What is the relation between KL divergence and entropy?
    1. While KL divergence quantifies the difference between two distributions, entropy measures the uncertainty or average amount of information in a single distribution.
    2. KL divergence can be decomposed into the difference between the entropy of P and the cross entropy of P and Q.
    3. Can be represented mathematically as:

KL(P||Q)=H(P)-H(P,Q)