Deep Learning Interview Questions

1) **What is Deep learning?**
   Deep learning is nothing but a paradigm of machine learning which has shown incredible promise in recent years. This is because of the fact that Deep learning shows a great analogy with the functioning of the human brain.
2) **What is the Difference between machine learning and Deep learning?**
   Machine learning is a field of computer science that gives computers that ability to learn without being explicitly programmed. Machine Learning can be categorized in the following
   • supervised machine learning
   • Unsupervised machine learning
   • Reinforcement machine learning
   Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural network.
3) **What, in your opinion is the reason for the popularity of Deep learning in recent times?**
   This is because of two main reasons.
   • The increase in the amount of data generated through various sources.
   • The growth in hardware resources required to run these models.
   For example GPU's are multiple times faster and they help us build bigger and deeper deep learning models in comparatively less time than we required previously.
4) **What is Reinforcement Learning?**
   Reinforcement learning is learning what to do and how to map situations to actions. The end result to maximize the numerical reward signal. The learned is not told which action to take but instead must discover which action will yield the maximum reward. Reinforcement learning is inspired by the learning of human beings. It is based on the reward / penalty mechanism.
5) **What are Artificial Neural Networks?**
   Artificial neural networks are a specific set of algorithms that have revolutionized machine learning. They are inspired by biological neural networks. Neural networks can adapt to changing the inputs so the network generates the best possible results without needing to redesign the output criteria.
6) **Describe the structure of ANN?**
   ANN works on the same principle as biological neural network. It consists of inputs which get processed with weight and bias, with the help of activation functions.
7) **How weights initialized in neural network?**
   • We can either initialize the weights to zero or assign them randomly.
   • **Initializing all weights to zero:** This makes your model similar to a linear model. All neurons and every layer performs the same operation, giving the same output and making the deep net useless.
   • **Initializing all weights randomly:** Here weights are assigned randomly by initializing them very close to "0". It gives better accuracy to the model since every neuron performs different computations. This is the most commonly used method.

8) **What is cost function?**

It is also referred as “loss” or “error” cost function is a measure to evaluate how good your models performance is. It is used to compute the error of the output layer during backpropagation. We push that error backwards through the neural network and use that during the different training functions.

9) **What are Hyper-parameters?**

A hyper-parameter is a parameter whose value is set before the learning process begins. It determines how a network is trained and the structure of the network (such as the numbers of hidden units, the learning rate, epochs, etc.)

10) **What will happen if the learning rate is set inaccurately (too low or too high)?**

When learning rate is too low, training of the model will progress very slowly as we are making minimal updates to the weights. It will take many updates before reaching the minimum point.

If the learning rate is set too high, this causes undesirable divergent behavior to the loss function.

11) **What is the difference between epoch, Batch and iteration in Deep Learning?**

• Epoch represents one iterations over the entire dataset (everything put into the training model)

• Batch – refers to when we cannot pass the entire dataset into the neural network at once, so we divide the dataset into several batches.

• Iteration – if we have 10,000 images as data and a batch size of 200, then epoch should run 50 iterations (10,000 divided by 50)

12) **What are the different layers in CNN?**

• **Convolutional layer:** The layer that performs a convolutional operations creating several smaller picture windows to go over the data.

• **Relu Layer:** It brings non-linearity to the network and converts all the negative pixels to zero the output is a rectified feature map.

• **Pooling layer:** Pooling is a down sampling operation that reduces the dimensionality of the feature map.

• **Fully Connected layer:** This layer recognizes and classifies the objects in the image.

13) **What is pooling on CNN, and how does it works?**

Pooling is used to reduce the spatial dimension of a CNN, It performs down-sampling operations to reduce the dimensionality and creates a pooled feature map by sliding a filter matrix over the input matrix.

14) **What are recurrent neural network?**

RRN is type of network designed for capturing information from sequence/time series data. They can take variable as input & gives variable size outputs.

15) **How does an LSTM network works?**

LSTM is a special kind of neural network capable of learning long-term dependencies remembering information for long periods as its default behavior. There are three steps in an LSTM network.

• The network decides what to forget and what to remember

• It selectively updates cell state values

• The network decide what part of the current state makes it to the output.

16) **What is multi-layer perceptron?**

In neural networks, multi-layer perceptron's having input layer, hidden layer and output layer. It has the same structure as a single layer perceptron with one or more hidden layers. A single layer perceptron can classify only linear separable classes with binary output (0, 1) but MLP can classify non-linear classes except for the input layer, each node in the other layers uses a non-linear activation function. This means the input layers the data coming in and the activation function is based upon all nodes and weights being added together producing the output. MLP uses a supervised learning method called backpropagation.

In the backpropagation, the neural network calculates the error with the help of cost function. It propagates this error backwards from where it came it adjust the weights to train the model more accurately.

17) **Explain Gradient Descent?**

A Gradient measures how much the output of a function changes, if you change the inputs a little bit it simply measures the changes in all weights with regard to the change in error. You can also think of a gradient as the slope of a function.

Gradient Descent can be a thought of climbing down to the bottom of a valley, instead of climbing up a hill. This is because it is minimization algorithm that minimizes a given function.

18) **What is Stochastic Gradient Descent?**

Fitting the right slope that minimize the error or cost function for large dataset can be tricky. However, this can be achieved through a stochastic gradient descent optimization algorithm.

19) **What is Dropouts?**

Dropout is a technique to overcome overfitting problem. It has functionality to activate and deactivate selected subset of features and subset of outputs from hidden layer similar it goes on & on. In case of test data all neurons will be connected there is no activation and deactivation functionality.

20) **What is Adagrad optimization?**

Adagrad is an optimizer with parameter specific learning rates, which are adapted relative to how frequently a parameter gets updated during training. The more updates a parameter receives, the smaller the smaller learning rate.

21) **What is adadelta optimizer?**

Adadelta is more robust extension of Adagrad that adapts learning rates based on a moving window of gradient updates, instead of accumulating all past gradients.

22) **What is RMSprop Optimizer?**

RMSprop is an optimizer that utilizes the magnitude of recent gradients to normalize the gradients. We always keep moving average over the root mean squared hence (RMS) Gradients, by which we divide the current Gradient.

23) **What is exploding Gradient?**

While training an Deep Neural network, if you see exponentially growing (very large) error gradient which accumulate and result in very large updates weights to neural network model during training, they are known as exploding gradients. At an extreme, the values of weights can become so large as to overflow and results in NaN values. This is the effect of your model is unstable to learn from your training data.

24) What is vanishing Gradient?

While training a Deep Neural Network, your slop can become either two small. This makes the training difficult when the slope is too small, the problem is known as a vanishing Gradient. It leads to long training time's poor performance & low accuracy.

25) What is backpropagation? Explain how its working?

Backpropagation is a training algorithm used for multi-layer neural network. In this method, we move the error from an end of the network to all weights inside the network & thus allowing efficient computation of the gradient. It has the following steps

- Forward propagation of a training data
- Derivatives are computed using output & target
- Back propagate for computing derivatives of error with respect to output activation.
- using previously calculated derivatives for output.
- updates the weights.

26) What are the variants of Back propagation?

- **Stochastic Gradient Descent:** We use only a single training example for calculation of the gradient and update parameters.
- **Batch Gradient Descent:** We calculate the Gradient for the whole dataset & performing the update at each iteration.
- **Mini-Batch Gradient Descent:** it is one of the most popular optimization algorithm it's variants of stochastic Gradient Descent and here instead of single training example, mini batch samples is used.

27) What is the role of Activation function?

The activation function is used to introduce non-linearity into the neural network helping it to learn more complex function, which is a linear combination of its input data.

An activation function is a function in an artificial neural network that delivers an output based on inputs.

28) What is an Auto encoder?

Auto encoder are simple learning networks that aim to transform input into outputs with the minimum possible error. This means that we want the output with the minimum possible error. This means that we want the output to be close to input as possible. We add a couple of layers between the input and the output, and the sizes of these layers are smaller than the input layer. The auto-encoder receives unlabeled input which is then encoded to reconstruction the input.

29) What is Restricted Boltzman machine?

Boltzman machine have a simple learning algorithm allows them to discover interesting features that represent complex regularities in the training data. The Restricted Boltzman machine is basically used to optimize the weights and the quantity for the given problem. The learning algorithm is very slow in networks with many layers of feature detectors. Restricted Boltzman Machine algorithm has a single layer of feature detectors which makes it faster than the rest.

30) What is Dropout?

Dropout is a technique of dropping out hidden and visible units of a networks randomly to prevent overfitting of data (typically dropping 20 percent of the nodes). It doubles the number of iterations needed to converge the network.

31) What is Batch-Normalization?

Batch-normalization is the technique to improve the performance and stability of neural networks by normalizing the inputs in every layer so that they have mean output activation of zero and standard deviation of one.

32) Why Tensorflow is the most preferred library in Deep learning?

Tensorflow provides both c++ and python API's making it easier to work on and has a faster compliment time compared to deep learning libraries like Keras and Torch. Tensorflow supports both CPU and GPU computing Devices.

33) What do you mean by Tensor in Tensorflow?

A tensor is a mathematical object represented an array of higher dimensions. There arrays of data with different dimensions and ranks fed on input to the neural network are called "Tensor".

34) What is the computational graph?

Everything in a Tensorflow is based on creating a computational graph. It has network of nodes where each node operates, nodes represent mathematical operations and edges represent tensor since data flows in the form of a graph it is also called a "Dataflow graph".

35) What is convolutional Neural Network?

CNN's are similar to ordinary neural networks, expect that it explicitly assumes that the inputs are images, which allows us to encode certain properties into the architecture. These then make forward function efficient to implement and reduces the parameters in the network.

36) What is Recurrent Neural Network?

The feed forward network is not known to do well on sequential events model such as the probabilistic languages model of predicting the next word based on the previous word at every given point.

RNN address this issue. It is similar to MLP except that they have a feedback loop, which means they feed previous time steps into the current step. This type of architecture generates sequences to simulate situation and create synthetic data, making them the ideal modeling choice to work on sequence data such as speech, text mining, image captioning , time series prediction robot control language modelling.

The previous steps hidden layer and final outputs are fed back into the network and will resemble the past and it will repeatedly predicts what will happen next.

The drawback in the general RNN architecture is that it can be memory heavy, and hard to train for long term temporal dependency (i.e. context of long text should be known as any given stage)

37) What is long term Short term Memory neural network?

LSTM is an implementation of improved RNN architecture to address the issue of general RNN, and it enables long range dependencies. It is designed to have better memory through linear memory cells surrounded by a set of gate units used to control the flow of information should enter the memory, when to forget, and when to output. It uses no activation function within its recurrent components, thus the gradient term does not vanish with back backpropagation.

38) What is Gated Recurrent Unit?

GRU is a variant of the RNN architecture, and uses gating mechanism to control and manage the flow of information between cells in the neural network.