1)what is the difference between trainable & non trainable parameters ?

Ans :

trainable parameters' are those which value is modified according to their gradient (the derivative of the error/loss/cost relative to the parameter), whereas 'non-trainable parameters' are those which value is not optimized according to their gradient.

2)In the CNN architecture what does the dropout layer go ?

Ans : In CNN architecture dropout layer does go,Dropout is just a regularization technique for preventing overfitting in the network. It sets a node's weight to zero with a given probability during training, reducing the number of weights required for training at each iteration. It can be applied for each layer of the network (regardless if it is fully connected or convolutional), or after selected layers. To which layers dropout is applied is really just a design decision for what results in best performance.

3)what is the optimal number of hidden layers to stack.?

Ans :

The number of hidden neurons should be between the size of the input layer and the size of the output layer. The number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer. The number of hidden neurons should be less than twice the size of the input layer.

4)In each layer,how many secret units or filters should there be ?

Ans : The optimal number of hidden units could easily be smaller than the number of inputs, there is no rule like multiply the number of inputs with N... If you have a lot of training examples, you can use multiple hidden units, but sometimes just 2 hidden units works best with little data

Filters should we use it is common for a convolutional layer to learn from 32 to 512 filters in parallel for a given input.

5)what should you initial learning rate be ?

Ans : This means that a learning rate of 0.1, a traditionally common default value, would mean that weights in the network are updated 0.1 \* (estimated weight error) or 10% of the estimated weight error each time the weights are updated.

Specifically, the learning rate is a configurable hyperparameter used in the training of neural networks that has a small positive value, often in the range between 0.0 and 1.0.

6)what do you do with the activation function ?

Ans : Activation function decides, whether a neuron should be activated or not by calculating weighted sum and further adding bias with it. The purpose of the activation function is to introduce non-linearity into the output of a neuron.

neural network has neurons that work in correspondence of weight, bias and their respective activation function. In a neural network, we would update the weights and biases of the neurons on the basis of the error at the output. This process is known as back-propagation. Activation functions make the back-propagation possible since the gradients are supplied along with the error to update the weights and biases.

Why do we need Non-linear activation functions :- A neural network without an activation function is essentially just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks.

Types of activation function :

1). Linear Function :-

Equation : Linear function has the equation similar to as of a straight line i.e. y = x

2). Sigmoid Function :-

It is a function which is plotted as ‘S’ shaped graph.

Equation : A = 1/(1 + e-x)

3). Tanh Function :- The activation that works almost always better than sigmoid function is Tanh function also knows as Tangent Hyperbolic function. It’s actually a mathematically shifted version of the sigmoid function. Both are similar and can be derived from each other.

Equation :-

f(x) = tanh(x) = 2/(1 + e-2x) - 1

OR

tanh(x) = 2 \* sigmoid(2x) - 1

4). Softmax Function :- The softmax function is also a type of sigmoid function but is handy when we are trying to handle multi- class classification problems

7)what is normalization of data ?

Ans : normalization is a process that changes the range of pixel intensity values. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching or histogram stretching.

Normalization is the process of organizing data in a database. This includes creating tables and establishing relationships between those tables according to rules designed both to protect the data and to make the database more flexible by eliminating redundancy and inconsistent dependency.

8)what is image augmentation and how does it work ?

Ans : Image augmentation is a technique of altering the existing data to create some more data for the model training process. In other words, it is the process of artificially expanding the available dataset for training a deep learning model.

Image augmentation artificially creates training images through different ways of processing or combination of multiple processing, such as random rotation, shifts, shear and flips, etc

9)what is decline in learning rate ?

Ans : Learning rate (λ) is one such hyper-parameter that defines the adjustment in the weights of our network with respect to the loss gradient descent. It determines how fast or slow we will move towards the optimal weights.the simplest learning rate schedule is to decrease the learning rate linearly from a large initial value to a small value. This allows large weight changes in the beginning of the learning process and small changes or fine-tuning towards the end of the learning process

10)what does early stopping Criteria mean.?

Ans : Early stopping is a method that allows you to specify an arbitrary large number of training epochs and stop training once the model performance stops improving on a hold out validation dataset.

The use of early stopping requires the selection of a performance measure to monitor, a trigger to stop training, and a selection of the model weights to use.

There are three elements to using early stopping; they are:

1.Monitoring model performance.

2.Trigger to stop training.

3.The choice of model to use.

Early stopping is an optimization technique used to reduce overfitting without compromising on model accuracy. The main idea behind early stopping is to stop training before a model starts to overfit.