1. B

2. C

3. A

4. B

5. C

6. B

7. A

8. A

9. A

10. B

11.

If we do not use activation functions in any of the hidden layers in a neural network, it’ll start behaving like a linear network with a lot of terms. The learning that will happen will be equivalent to that in the case of Linear regression.

12.

Forward and backward propagation both happen layer wise. The direction of one is completely opposite to the other. In forward propagation, we pass input through the network and compute the output using simple function calculations.

Back propagation is carried out using chain rule. We calculate the derivative of error with respect to each and every weight present in the network till the first layer.

13.

In Batch gradient descent, the entire data is passed through the network in one go.

In stochastic gradient descent, we only pass a single data point through the network and back propagate it to update weights, whether in mini batch gradient descent, we pass a fix number of points (greater than one, generally in exponents of 2) through our forward connection and calculate the error to update weights of the network through back propagation.

14.

If the size of our training data is so big that it can not be accomodated in RAM, we break the data into chunks or batches to continue with the training. In such cases, we use the concept of mini batch gradient descent to move forward with our training process smoothly.

15.

Transfer learning is a concept where we can re-use the learnings of a network again and again on similar kind of data.

We always need not train a neural network from scratch for each and every task. Rather, we can use the weights learnt by some different network for similar kind of tasks. This is called Transfer learning.