DEEP LEARNING – WORKSHEET 3

Q1 to Q8 are MCQs with only one correct answer. Choose the correct option.

1. Which of the following is true about model capacity (where model capacity means the ability of neural network to approximate complex functions)?
   A) As dropout ratio increases, model capacity increases
   B) As number of hidden layers increase, model capacity increases
   C) As learning rate increases, model capacity increases
   D) None of the above

ANS: B) As number of hidden layers increase, model capacity increases

2. Batch Normalization is helpful because?
   A) It is a very efficient backpropagation technique
   B) It returns back the normalized mean and standard deviation of weights
   C) It normalizes (changes) all the input before sending it to the next layer
   D) None of the above

ANS: C) It normalizes (changes) all the input before sending it to the next layer

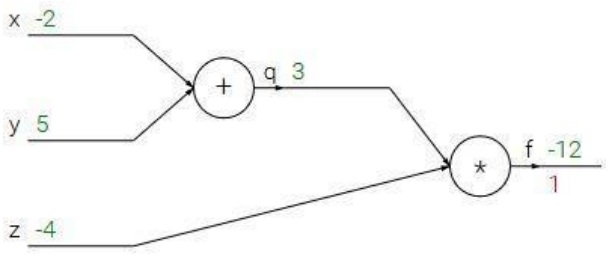
3. What if we use a learning rate that's too large?
   A) Network will not converge
   B) Network will converge
   C) either A or B
   D) None of the above

ANS: A) Network will not converge

4. What are the factors to select the depth of neural network?
   i) Type of neural network (e.g. MLP, CNN etc.)
   ii) Input data
   iii) Computation power, i.e. Hardware capabilities and software capabilities
   iv) Learning Rate
   v) The output function to map
   A) 1, 2, 4, 5
   B) 2, 3, 4, 5
   C) 1, 3, 4, 5
   D) All of these

ANS:D) All of these

5. Suppose you have inputs as x, y, and z with values -2, 5, and -4 respectively. You have a neuron 'q' and neuron 'f' with functions:
   q = x + y
   f = q * z
   Graphical representation of the functions is as follows:



What is the gradient of F with respect to x, y, and z? (use chain rule of derivatives to find the solution)

A) (3, -4, -4) B) (-3, 4, 4)
C) (-4, -4, 3) D) (4, 4, 3)

ANS: C) (-4, -4, 3)

6. Which of the following statement is the best description of early stopping?
   A) Train the network until a local minimum in the error function is reached
   B) Simulate the network on a test dataset after every epoch of training. Stop training when the generalization error starts to increase
   C) Add a momentum term to the weight update in the Generalized Delta Rule, so that training converges more quickly
   D) None of the above

ANS: B) Simulate the network on a test dataset after every epoch of training. Stop training when the generalization error starts to increase

7. Which gradient descent technique is more advantageous when the data is too big to handle in RAM simultaneously?
   A) Mini Batch Gradient Descent B) Stochastic Gradient Descent
   C) Full Batch Gradient Descent D) either A or B

ANS: B) Stochastic Gradient Descent

8. Consider the scenario. The problem you are trying to solve has a small amount of data. Fortunately, you have a pre-trained neural network that was trained on a similar problem. Which of the following methodologies would you choose to make use of this pre-trained network?
   A) Freeze all the layers except the last, re-train the last layer
   B) Assess on every layer how the model performs and only select a few of them
   C) Fine tune the last couple of layers only
   D) Re-train the model for the new dataset

ANS: A) Freeze all the layers except the last, re-train the last layer

Q9 and Q10 are MCQs with one or more correct answers. Choose all the correct options.

9. Which of the following neural network training challenge can be solved using batch normalization?
   A) Overfitting B) Training is too slow
   C) Restrict activations to become too high or low
   D) None of these

ANS: B) Training is too slow C) Restrict activations to become too high or low

10. For a binary classification problem, which of the following activations may be used in output layer?
    A) ReLU B) sigmoid
    C) softmax D) Leaky ReLU

ANS: B) sigmoid

Q11 to Q15 are subjective answer type question. Answer them briefly.

11. What will happen if we do not use activation function in artificial neural networks?

- In neural network an artificial neuron calculates the weighted sum of the input and adds a bias to it and generate the next layer. So, this value can be anything and the neuron doesn't know which value to send forward and which not to send. Here the activation function is used to determine which value to be activated and which one should be shut down
- In an Artificial Neural Network, the activation function is a non-linear transformation which is implemented on the input or the final layer to generate the output.
- So, it's clear that this process adds extra step during forward propagation. Activation functions also have a major role on determining the neural network's ability to converge and the convergence speed, sometimes, it prevents the neural networks from converging in the first place.
- So, a neural network without an activation function will simply perform liner transformation on the input with the weights and bias. This will make the network weaken and the network will not learn complex data patterns
- It will simply be a linear regression model.

12. How does forward propagation and backpropagation work in deep learning?

13. Explain briefly the following variant of Gradient Descent: Stochastic, Batch, and Mini-batch?

14. What are the main benefits of Mini-batch Gradient Descent?

- The mini batch gradient descent has both the characters of stochastic gradient descent and batch gradient descent.
- It fits in the memory easily
- If stuck in local minimums, some noisy steps can lead the way out of them
- Average of the training samples produces stable error gradients and convergence

15. What is transfer learning?

- Transfer learning may be is a machine learning technique where a pre trained model is used to train a new task.
- It is one of the most popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.
- It is a technique for optimization and time saving and better efficiency
- There are two approach for this technique they are as follows

    1. Develop Model Approach
    2. Pre-trained Model Approach