# 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)?
   1. As dropout ratio increases, model capacity increases
   2. As number of hidden layers increase, model capacity increases
   3. As learning rate increases, model capacity increases
   4. None of the above

**Ans. B) As number of hidden layers increase, model capacity increases**

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

**Ans. C) It normalizes (changes) all the input before sending it to the next layer**

1. What if we use a learning rate that’s too large?
   1. Network will not converge B) Network will converge

C) either A or B D) None of the above

**Ans. A) Network will not converge**

1. What are the factors to select the depth of neural network?
2. Type of neural network (e.g. MLP, CNN etc.)
3. Input data
4. Computation power, i.e. Hardware capabilities and software capabilities
5. Learning Rate
6. The output function to map

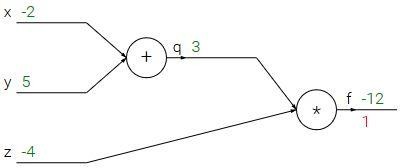
A) 1, 2, 4, 5 B) 2, 3, 4, 5

C) 1, 3, 4, 5 D) All of these

**Ans. D) All of these**

1. 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)**

1. Which of the following statement is the best description of early stopping?
   1. Train the network until a local minimum in the error function is reached
   2. Simulate the network on a test dataset after every epoch of training. Stop training when the generalization error starts to increase
   3. Add a momentum term to the weight update in the Generalized Delta Rule, so that training converges more quickly
   4. 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**

1. Which gradient descent technique is more advantageous when the data is too big to handle in RAM simultaneously?
   1. Mini Batch Gradient Descent B) Stochastic Gradient Descent

C) Full Batch Gradient Descent D) either A or B

**Ans. B) Stochastic Gradient Descent**

1. 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?
   1. Freeze all the layers except the last, re-train the last layer
   2. Assess on every layer how the model performs and only select a few of them
   3. Fine tune the last couple of layers only
   4. 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.

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

**Ans. A) Overfitting, B) Training is too slow, and C) Restrict activations to become too high or low**

1. For a binary classification problem, which of the following activations may be used in output layer?
   1. ReLU B) sigmoid

C) softmax D) Leaky ReLU

**Ans. B) sigmoid**

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

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

Ans. A neural network without an activation function is essentially just a linear regression model.

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

## Ans. Forward propagation: In the forward propagation, we check what the neural network predicts for the first training example with initial weights and bias. First, we initialize the weights and bias randomly.

## ****Backward propagation:**** We can define a cost function that measures how good our neural network performs. For an input, x, and desired output, y, we can calculate the cost of a specific training example as the square of the difference between the network’s output and the desired output, that is,

Image for post

Where k stands for the training example and the output is assumed to be the activation of the output neuron, and y is the actual desired output.

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

Ans. **Stochastic Gradient Descent:** Uses only single training example to calculate the gradient and update parameters.

**Batch Gradient Descent:** Calculate the gradients for the whole dataset and perform just one update at each iteration.

**Mini-batch Gradient Descent:** Mini-batch gradient is a variation of stochastic gradient descent where instead of single training example, mini-batch of samples is used. It’s one of the most popular optimization algorithms.

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

Ans. **Benefits of Mini-batch Gradient Descent -**

* **Computational Efficiency**: In terms of computational efficiency, this technique lies between the two previously introduced techniques.
* **Stable Convergence:** Another advantage is the more stable converge towards the global minimum since we calculate an average gradient over n samples that results in less noise.
* **Faster Learning**: As we perform weight updates more often than with stochastic gradient descent, in this case, we achieve a much faster learning process.

1. What is transfer learning?

Ans. Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task.

For example, knowledge gained while learning to recognize cars can be used to some extent to recognize trucks.