Practice Quiz Assessment : Deep Learning

What is the main difference between a Shallow and a Deep Neural Network?

The number of layers in a Shallow Network are lesser than the number of layers in a Deep Neural Network

You Selected

The activation functions used in these networks are different

The training algorithm used in Shallow Networks are different because they are used in scenarios where not much information extraction from the Input Data is necessary

Shallow Neural Networks are used for Structured data while Deep Neural Networks are used with unstructured data

The only defining feature of a Neural Network that makes it shallow or deep is its depth or the number of layers. Different tasks require different architectures of Neural Nets, but broadly classifying a Neural Network can either be regarded as deep or it can be regarded as shallow.

**Question 2**

Incorrect Answer

Marks: 0/1

How does Dropout layer prevent overfitting in Neural Networks?

It adds noise to the input data so that the model doesn’t closely fit to the training data

It removes random neurons during training, effectively bringing down the number of trainable weights

Correct Option

It increases the learning rate so that each backpropagation adjusts the weights with bigger step sizes

It applies L1 or L2 regularization to the weights so that it can account for a bias in the error term that leaves room to prevent overfitting

You Selected

Dropout process reduces the number of neurons by switching them off during training. This ensures that we are not overfitting with too many weights. The switched off neurons have all their weights set to zero, so effectively they all produce a zero output to the next layer.

**Question 3**

Incorrect Answer

Marks: 0/1

What is the purpose of Batch Normalization in Deep Learning?

To reduce the number of trainable parameters in the network by taking an average of the weights across neurons within the same layer

To speed up the training process by optimizing the batch size hyperparameter.

You Selected

To reduce the effects of variance within the same batch by normalizing the data fields across the batch

Correct Option

To add noise to the input data for preventing overfitting on input data

Batch-Normalization standardizes every data field for the batch of data fed to the Neural Network so that the data lies within a sensible range.

**Question 4**

Correct Answer

Marks: 1/1

In a Convolutional Neural Network, what is the purpose of the stride parameter?

To control the size of the convolutional filter

To control the number of filters in a given layer

To control the size of pooling window

To control by how much does the convolutional filter moves per step

You Selected

As the name suggests, the stride parameter determines the step size for Convolution in that particular layer.

**Question 5**

Incorrect Answer

Marks: 0/1

In the context of transformers, what do the terms 'query,' 'key,' and 'value' refer to?

Components of positional encoding

You Selected

Elements of the input sequence

Key concepts in self-attention

Correct Option

Parameters in the decoder network

In transformers, the terms 'query,' 'key,' and 'value' are fundamental components of the self-attention mechanism. They are used to calculate attention scores, determining the importance of different elements in the input sequence. The 'query' represents the element that seeks information, the 'key' represents the element being compared to the query, and the 'value' represents the associated information. The attention mechanism, driven by these key concepts, allows the model to focus on relevant parts of the input sequence, contributing to its ability to capture long-range dependencies and improve performance on various tasks.

**Question 6**

Correct Answer

Marks: 1/1

State whether the below statement is True or False:

The self-attention mechanism in transformers enables the model to consider different parts of the input sequence simultaneously.

True

You Selected

False

The self-attention mechanism in transformers indeed allows the model to consider different parts of the input sequence simultaneously. This capability is crucial for capturing long-range dependencies and relationships within the sequence, enabling transformers to effectively process and understand complex patterns in data. The parallelized processing of multiple parts of the sequence contributes to the model's ability to handle information in a more holistic and context-aware manner, making it a powerful architecture for various tasks.

**Question 7**

Correct Answer

Marks: 1/1

How does the activation function affect the performance of a Neural Network?

It determines the number of neurons in the hidden layers

It determines the learning rate of the model

It affects the speed of convergence during training

It determines the nonlinearity of the model

You Selected

Neurons are linear combinations of weights and values from previous layers. So in order to account for non linearity, activation functions are used after some layers.

**Question 8**

Correct Answer

Marks: 1/1

How does transfer learning work in deep learning?

By using pre-trained models as a starting point for a new task

You Selected

By combining multiple models to improve performance

By training a model on multiple GPUs or across several computational resources

By automatically generating new data and fine tuning on the generated data

The point of Transfer Learning is taking advantage of the fact that layers are pre-trained on different task but we can trust its learned representations to perform equally well in our new task with some fine-tuning.

**Question 9**

Correct Answer

Marks: 1/1

What is the main difference between Supervised Learning and Contrastive Learning in Convolutional Neural Networks?

Supervised Learning maximizes the similarity between positive pairs, while Contrastive Learning maximizes the dissimilarity between positive and negative pairs

Supervised Learning requires labeled data, while Contrastive Learning does not

You Selected

Supervised Learning only works with fully connected layers, while Contrastive Learning works with Convolutional Layers

Supervised Learning requires a larger number of training samples than Contrastive Learning

In Contrastive Learning, positive and negative pairs are formed from existing data. If the pair is formed from the same image, then they are used as Positive pairs and vice versa. Therefore, Contrastive Learning can do well with unlabeled data while Supervised Learning, by definition, requires labeled data.

**Question 10**

Correct Answer

Marks: 1/1

In a Convolutional Neural Network, which layer is used to convert the output of Convolutional layers into a format suitable for input to Fully Connected Layers?

Batch-Normalization layer

Dropout layer

Flatten layer

You Selected

Pooling layer

The purpose of Flatten layer is to reshape the multi dimensional vectors into one dimensional vectors so that they can be fed as input into the Dense or Fully-Connected layers.