

# DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE

(AUTONOMOUS)

(Approved by AICTE & Affiliated to Anna University, Chennai)

Accredited with 'A' Grade by NAAC, Accredited by TCS

Accredited by NBA with BME, ECE & EEE

**PERAMBALUR - 621 212. Tamil Nadu.**

website : [www.dsengg.ac.in](http://www.dsengg.ac.in)



## DEPARTMENT OF INFORMATION TECHNOLOGY

### TWO MARK QUESTION BANK

#### UNIT I INTRODUCTON TO DEEP LEARNING

##### 1. What is Machine Learning?

Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed. A machine has the ability to learn if it can improve its performance by gaining more data.

##### 2. How does Machine Learning work?

A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it.** The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

##### 3. Features of Machine Learning?

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

##### 4. What is the Need for Machine learning?

The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly. As a human, we have some limitations as we cannot access the huge amount of data manually, so for this, we need some computer systems and here comes the machine learning to make things easy for us.

## 5. Classification of Machine Learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning

## 6. Supervised Learning

Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

Supervised learning can be grouped further in two categories of algorithms:

- **Classification**
- **Regression**

## 7. Unsupervised Learning

Unsupervised learning is a learning method in which a machine learns without any supervision.

The training is provided to the machine with the set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.

In unsupervised learning, we don't have a predetermined result. The machine tries to find useful insights from the huge amount of data. It can be further classified into two categories of algorithms:

- **Clustering**
- **Association**

## 8. Define Reinforcement Learning?

Reinforcement learning is a feedback-based learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action. The agent learns automatically with these feedbacks and improves its performance. In reinforcement learning, the agent interacts with the environment and explores it. The goal of an agent is to get the most reward points, and hence, it improves its performance.

## 9. Define Logistic Regression?

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class or not. It is a kind of statistical algorithm, which analyze the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. For example email spam or not.

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability

## 10. What is Logistic Function (Sigmoid Function)?

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1. o The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the “S” form.
- The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

## 11. List the Type of Logistic Regression?

1. Binomial Logistic Regression
2. Multinomial Logistic Regression
3. Ordinal Logistic Regression

## 12. What is Binomial Logistic Regression?

- In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
- Target variable can have only 2 possible types: “0” or “1” which may represent “win” vs “loss”, “pass” vs “fail”, “dead” vs “alive”, etc., in this case, sigmoid functions are used, which is already discussed above.

### 13. What is Multinomial Logistic Regression?

- In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as “cat”, “dogs”, or “sheep”
- Target variable can have 3 or more possible types which are not ordered (i.e. types have no quantitative significance) like “disease A” vs “disease B” vs “disease C”.

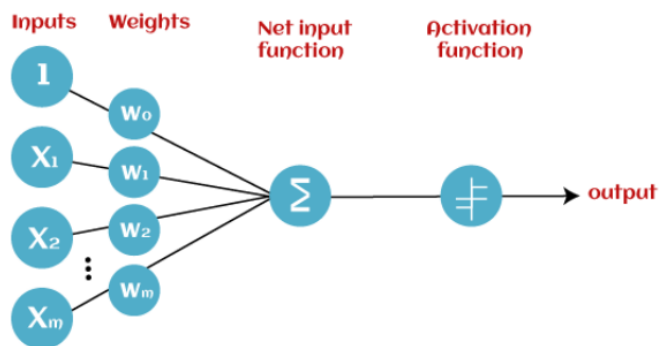
### 14. What is Ordinal Logistic Regression?

- In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as “low”, “Medium”, or “High”.
- It deals with target variables with ordered categories. For example, a test score can be categorized as: “very poor”, “poor”, “good”, or “very good”. Here, each category can be given a score like 0, 1, 2, or 3.

### 15. Define Perceptrons?

Perceptron is Machine Learning algorithm for supervised learning of various binary classification tasks. Further, Perceptron is also understood as an Artificial Neuron or neural network unit that helps to detect certain input data computations in business intelligence.

### 16. Draw the model of perceptron?



### 17. How does Perceptron work?

In Machine Learning, Perceptron is considered as a single-layer neural network that consists of four main parameters named input values (Input nodes), weights and Bias, net sum, and an activation function. The perceptron model begins with the multiplication of all input values and their weights, then adds these values together to create the weighted sum. Then this weighted sum is applied to the activation function 'f' to obtain the desired output

## 18. Define Support Vector Machines?

Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks. SVMs can be used for a variety of tasks, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection. SVMs are adaptable and efficient in a variety of applications because they can manage high-dimensional data and nonlinear relationships.

## 19. Explain the Types of SVM?

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

## 20. What is Artificial Neural Network?

The term "**Artificial Neural Network**" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes.

## 21. List the types of gradient descent algorithms?

- **Stochastic Gradient Descent** — During the training process of the model, we use only one particular example from our dataset during each iteration.
- **Batch Gradient Descent** — During the training process of the model, we optimize the loss function using the mean of all the individual losses in our whole dataset. This is computationally expensive.
- **Mini-Batch Gradient Descent** — It computes the gradients on small sets of instances called mini-batches. Mini-batch Gradient Descent is faster than Batch Gradient Descent since less number of training examples are used for every training step.

## UNIT II DEEP NETWORKS

### 1. What is Deep Learning?

Deep learning is the branch of machine learning which is based on artificial neural network architecture. An artificial neural network or ANN uses layers of interconnected nodes called neurons that work together to process and learn from the input data.

### 2. Define the Natural language processing (NLP)?

In NLP, the Deep learning model can enable machines to understand and generate human language. Some of the main applications of deep learning in NLP include:

- **Automatic Text Generation** – Deep learning model can learn the corpus of text and new text like summaries, essays can be automatically generated using these trained models.
- **Language translation:** Deep learning models can translate text from one language to another, making it possible to communicate with people from different linguistic backgrounds.
- **Sentiment analysis:** Deep learning models can analyze the sentiment of a piece of text, making it possible to determine whether the text is positive, negative, or neutral. This is used in applications such as customer service, social media monitoring, and political analysis.
- **Speech recognition:** Deep learning models can recognize and transcribe spoken words, making it possible to perform tasks such as speech-to-text conversion, voice search, and voice-controlled devices.

### 3. Define Reinforcement learning?

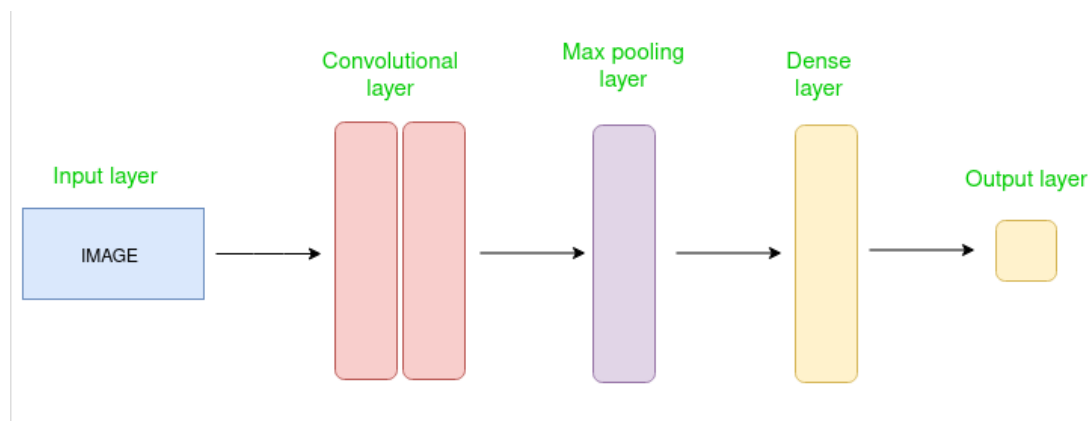
In reinforcement learning, deep learning works as training agents to take action in an environment to maximize a reward. Some of the main applications of deep learning in reinforcement learning include:

1. **Game playing:** Deep reinforcement learning models have been able to beat human experts at games such as Go, Chess, and Atari.
2. **Robotics:** Deep reinforcement learning models can be used to train robots to perform complex tasks such as grasping objects, navigation, and manipulation.
3. **Control systems:** Deep reinforcement learning models can be used to control complex systems such as power grids, traffic management, and supply chain optimization.

#### 4. Define Convolutional Neural Network?

- A Convolutional Neural Network (CNN) is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.
- Convolutional Neural Network (CNN) is the extended version of artificial neural networks (ANN) which is predominantly used to extract the feature from the grid-like matrix dataset. For example visual datasets like images or videos where data patterns play an extensive role.

#### 5. Draw the CNN architecture.



#### 6. What is activation layer?

By adding an activation function to the output of the preceding layer, activation layers add nonlinearity to the network. It will apply an element-wise activation function to the output of the convolution layer. Some common activation functions are **RELU**:  $\max(0, x)$ , **Tanh**, **Leaky RELU**, etc. The volume remains unchanged hence output volume will have dimensions  $32 \times 32 \times 12$ .

#### 7. What is Pooling layer?

This layer is periodically inserted in the convnets and its main function is to reduce the size of volume which makes the computation fast, reduces memory, and also prevents overfitting. Two common types of pooling layers are **max pooling** and **average pooling**. If we use a max pool with  $2 \times 2$  filters and stride 2, the resultant volume will be of dimension  $16 \times 16 \times 12$ .

## **8. Define Generative Adversarial Networks (GAN)**

A Generative Adversarial Network (GAN) is a deep learning architecture that consists of two neural networks competing against each other in a zero-sum game framework. The goal of GANs is to generate new, synthetic data that resembles some known data distribution.

## **9. Define Generator Model?**

The Generator is trained while the Discriminator is idle. After the Discriminator is trained by the generated fake data of the Generator, we can get its predictions and use the results for training the Generator and get better from the previous state to try and fool the Discriminator.

## **10. What is Discriminator Model?**

The Discriminator is trained while the Generator is idle. In this phase, the network is only forward propagated and no back-propagation is done. The Discriminator is trained on real data for  $n$  epochs and sees if it can correctly predict them as real. Also, in this phase, the Discriminator is also trained on the fake generated data from the Generator and see if it can correctly predict them as fake.

## **11. What is Semi-Supervised Learning?**

Semi-supervised learning is a type of machine learning that falls in between supervised and unsupervised learning. It is a method that uses a small amount of labeled data and a large amount of unlabeled data to train a model. The goal of semi-supervised learning is to learn a function that can accurately predict the output variable based on the input variables, similar to supervised learning. However, unlike supervised learning, the algorithm is trained on a dataset that contains both labeled and unlabeled data.

## **12. What is Text classification?**

In text classification, the goal is to classify a given text into one or more predefined categories. Semi-supervised learning can be used to train a text classification model using a small amount of labeled data and a large amount of unlabeled text data.

### **13. What is Image classification?**

In image classification, the goal is to classify a given image into one or more predefined categories. Semi-supervised learning can be used to train an image classification model using a small amount of labeled data and a large amount of unlabeled image data.

### **14. What is Anomaly detection?**

In anomaly detection, the goal is to detect patterns or observations that are unusual or different from the norm.

### **15. Applications of Semi-Supervised Learning**

- **Speech Analysis:** Since labeling audio files is a very intensive task, Semi-Supervised learning is a very natural approach to solve this problem.
- **Internet Content Classification:** Labeling each webpage is an impractical and unfeasible process and thus uses Semi-Supervised learning algorithms. Even the Google search algorithm uses a variant of Semi-Supervised learning to rank the relevance of a webpage for a given query.
- **Protein Sequence Classification:** Since DNA strands are typically very large in size, the rise of Semi-Supervised learning has been imminent in this field.
- **Text document classifier-** As we know, it would be very unfeasible to find a large amount of labeled text data, so semi-supervised learning is an ideal model to overcome this.

### **16. Disadvantages of Semi-Supervised Learning?**

The most basic disadvantage of any Supervised Learning algorithm is that the dataset has to be hand-labeled either by a Machine Learning Engineer or a Data Scientist. This is a very costly process, especially when dealing with large volumes of data. The most basic disadvantage of any Unsupervised Learning is that its application spectrum is limited.

### **17. Define Regularization?**

“Any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error.”

## 18. Difference between L1 & L2 Parameter Regularization?

S.No	L1 Regularization	L2 Regularization
1	Penalizes the sum of absolute value of weights.	penalizes the sum of square weights.
2	It has a sparse solution.	It has a non-sparse solution.
3	It gives multiple solutions.	It has only one solution.
4	Constructed in feature selection.	No feature selection.
5	Robust to outliers.	Not robust to outliers.
6	It generates simple and interpretable models.	It gives more accurate predictions when the output variable is the function of whole input variables.
7	Unable to learn complex data patterns.	Able to learn complex data patterns.
8	Computationally inefficient over non-sparse conditions.	Computationally efficient because of having analytical solutions.

## 19. What is Shallow Networks?

Shallow neural networks give us basic idea about deep neural network which consist of only 1 or 2 hidden layers. Understanding a shallow neural network gives us an understanding into what exactly is going on inside a deep neural network A neural network is built using various hidden layers.

## 20. Difference Between a Shallow Net & Deep Learning Net ?

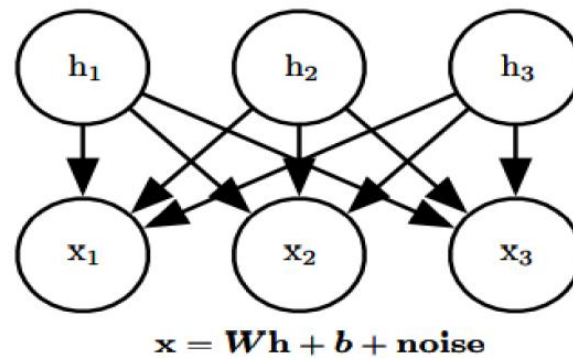
Sl.No	Shallow Net's	Deep Learning Net's
1	One Hidden layer(or very less no. of Hidden Layers)	Deep Net's has many layers of Hidden layers with more no. of neurons in each layers
2	Takes input only as VECTORS	DL can have raw data like image, text as inputs
3	Shallow net's needs more parameters to have better fit	DL can fit functions better with less parameters than a shallow network
4	Shallow networks with one Hidden layer (same no of neurons as DL) cannot place complex functions over the input space	DL can compactly express highly complex functions over input space
5	The number of units in a shallow network grows exponentially with task complexity.	DL don't need to increase it size(neurons) for complex problems
6	Shallow network is more difficult to train with our current algorithms (e.g. it has issues of local minima etc)	Training in DL is easy and no issue of local minima in DL

## UNIT III DIMENSIONALITY REDUCTION

### 1. Define Linear Factor Models?

Linear factor models are used as building blocks of mixture models of larger, deep probabilistic models. A linear factor model is defined by the use of a stochastic linear decoder function that generates  $x$  by adding noise to a linear transformation of  $h$ . It allows us to discover explanatory factors that have a simple joint distribution. A linear factor model describes the data-generation process as follows. ( we sample the explanatory factors  $h$  from a distribution)

$$h \sim p(h)$$



### 2. What is Dimensionality Reduction?

Dimensionality reduction is a technique used to reduce the number of features in a dataset while retaining as much of the important information as possible. In other words, it is a process of transforming high-dimensional data into a lower-dimensional space that still preserves the essence of the original data.

### 3. Steps involved in PCA:

- Start with data for  $n$  observations on  $p$  variables
- Form a matrix of size  $n \times p$
- Calculate the Covariance Matrix
- Calculate the Eigen vectors and Eigen Values
- Choose Principal Component from Feature Vectors
- Derive the new Data Set

#### 4. List the advantages of PCA?

- Removes Correlated Features:
- Improves Algorithm Performance:
- Reduces Over fitting
- Improves Visualization

#### 5. List the disadvantages of PCA?

**Independent variables become less interpretable:** After implementing PCA on the dataset, your original features will turn into Principal Components. Principal Components are the linear combination of your original features. Principal Components are not as readable and interpretable as original features.

**Data standardization is must before PCA:** You must standardize your data before implementing PCA, otherwise PCA will not be able to find the optimal Principal Components.

**Information Loss:** Although Principal Components try to cover maximum variance among the features in a dataset, if we don't select the number of Principal Components with care, it may miss some information as compared to the original list of features.

#### 6. Define Linear Discrimination Analysis (LDA)

Linear Discriminate Analysis as its name suggests is a linear model for classification and dimensionality reduction. Most commonly used for feature extraction in pattern classification problems.

#### 7. Need for LDA?

- Logistic Regression is perform well for binary classification but fails in the case of multiple
- Classification problems with well-separated classes. While LDA handles these quite efficiently.
- LDA can also be used in data pre-processing to reduce the number of features just as PCA which reduces the computing cost significantly.

## 8. Steps involved in LDA:

There are the three key steps.

- (i) Calculate the separability between different classes. This is also known as between-class variance and is defined as the distance between the mean of different classes.
- (ii) Calculate the within-class variance. This is the distance between the mean and the sample of every class.
- (iii) Construct the lower-dimensional space that maximizes Step1 (between-class variance) and minimizes Step 2(within-class variance).

## 9. What is Manifold Learnings?

Manifold learning for dimensionality reduction has recently gained much attention to assist image processing tasks such as segmentation, registration, tracking, recognition, and computational anatomy. The drawbacks of PCA in handling dimensionality reduction problems for non-linear weird and curved shaped surfaces necessitated development of more advanced algorithms like Manifold Learning.

## 10. List the Components of Dimensionality Reduction?

There are two components of dimensionality reduction:

1. **Feature selection:** In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It usually involves three ways:
  - Filter
  - Wrapper
  - Embedded
2. **Feature extraction:** This reduces the data in a high dimensional space to a lower dimension space, i.e. a space with lesser no. of dimensions.

## 11. What is Convolution Layers?

This is the very first layer in the CNN that is responsible for the extraction of the different features from the input images. The convolution mathematical operation is done between the input image and a filter of a specific size  $M \times M$  in this layer.

## 12. What is Fully Connected layers?

The Fully Connected (FC) layer comprises the weights and biases together with the neurons and is used to connect the neurons between two separate layers. The last several layers of a CNN Architecture are usually positioned before the output layer.

## 13. What is Pooling layer? list its types?

The Pooling layer is responsible for the reduction of the size (spatial) of the Convolved Feature. This decrease in the computing power is being required to process the data by a significant reduction in the dimensions.

There are two types of pooling

1 average pooling

2 max pooling

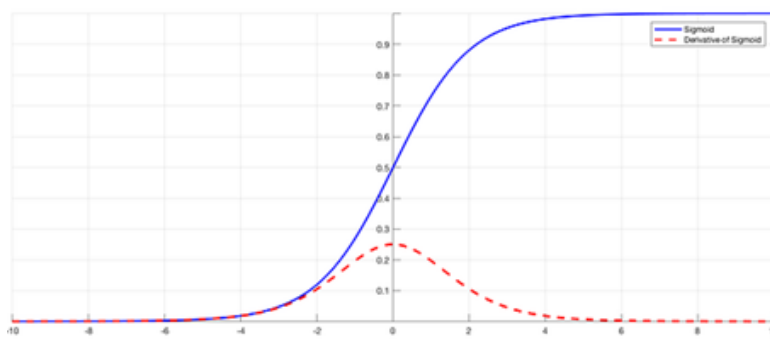
## 14. How to Training the convolution neural network

The process of adjusting the value of the weights is defined as the “training” of the neural network. Firstly, the CNN initiates with the random weights. During the training of CNN, the neural network is being fed with a large dataset of images being labelled with their corresponding class labels (cat, dog, horse, etc.). The CNN network processes each image with its values being assigned randomly and then make comparisons with the class label of the input image.

## 15. Define Zero initialization?

In general practice biases are initialized with 0 and weights are initialized with random numbers, what if weights are initialized with 0?

In order to understand let us consider we applied sigmoid activation function for the output layer.



## 16. Define Random initialization?

Assigning random values to weights is better than just 0 assignments. But there is one thing to keep in my mind is that what happens if weights are initialized high values or very low values and what is a reasonable initialization of weight values.

## 17. What is Batch Normalization?

- Before entering into Batch normalization let's understand the term "Normalization".
- **Normalization** is a data pre-processing tool used to bring the numerical data to a common scale without distorting its shape.

## 18. How does Batch Normalization work?

Since by now we have a clear idea of why we need Batch normalization, let's understand how it works. It is a two-step process. First, the input is normalized and later rescaling and offsetting is performed.

## 19. List the Advantages of Batch Normalization

- **Speed Up the Training**

By normalizing the hidden layer activation the Batch normalization speeds up the training process.

- **Handles internal covariate shift**

It solves the problem of internal covariate shift. Through this, we ensure that the input for every layer is distributed around the same mean and standard deviation. If you are unaware of what is an internal covariate shift, look at the following example.

- **Internal covariate shift**

Suppose we are training an image classification model that classifies the images into Dog or Not Dog. Let's say we have the images of white dogs only, these images will have certain distribution as well. Using these images model will update its parameters.

## 20. Why do we need batch normalization?

Batch normalization is essential because it helps address the internal covariate shift problem in deep neural networks. It normalizes the intermediate outputs of each layer within a batch during training, making the optimization process more stable and faster. By reducing internal covariate shift, batch normalization allows for higher learning rates, accelerates convergence, and improves generalization performance, leading to better and more efficient neural network training.

## 21. What is the process of batch normalization?

The process of batch normalization involves normalizing the intermediate outputs of each layer in a neural network during training. Here's the step-by-step process:

1. For each mini-batch of data during training, calculate the mean and variance of the activations across the batch for each feature in the layer.
2. Normalize the activations by subtracting the mean and dividing by the variance.
3. Scale and shift the normalized activations using learnable parameters (gamma and beta) to restore representation power. This allows the model to learn the optimal scale and shift for each feature.
4. During inference, use the population statistics (mean and variance) collected during training to normalize the activations, ensuring consistency between training and inference.
5. Batch normalization helps stabilize the optimization process, reduce internal covariate shift, and improve gradient flow, leading to faster convergence and better generalization.

## 22. Difference between RNN and Simple Neural Network

Recurrent Neural Network	Deep Neural Network
Weights are same across all the layers number of a Recurrent Neural Network	Weights are different for each layer of the network
Recurrent Neural Networks are used when the data is sequential and the number of inputs is not predefined.	A Simple Deep Neural network does not have any special method for sequential data also here the the number of inputs is fixed
The Numbers of parameter in the RNN are higher than in simple DNN	The Numbers of Parameter are lower than RNN
Exploding and vanishing gradients is the the major drawback of RNN	These problems also occur in DNN but these are not the major problem with DNN

## UNIT – IV OPTIMIZATION AND GENERALIZATION

### 1. What is Recurrent Neural Network (RNN)?

Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is its Hidden state, which remembers some information about a sequence.

### 2. How RNN works

The Recurrent Neural Network consists of multiple fixed activation function units, one for each time step. Each unit has an internal state which is called the hidden state of the unit. This hidden state signifies the past knowledge that the network currently holds at a given time step. This hidden state is updated at every time step to signify the change in the knowledge of the network about the past.

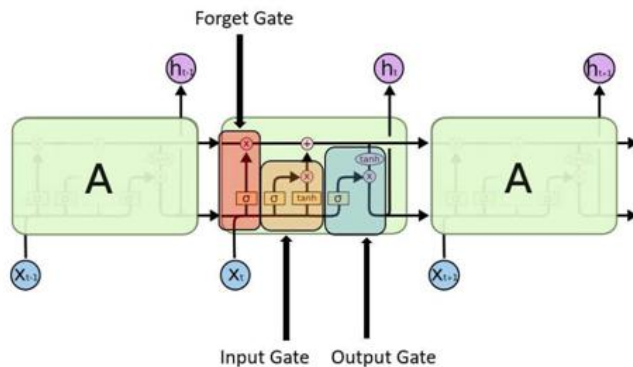
### 3. Define LSTM?

Long Short- Term Memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory

For example, LSTM is an application to tasks such as unsegmented, connected handwriting recognition, or speech recognition.

### 4. Draw the Structure of LSTM?

LSTM has a chain structure that contains four neural networks and different memory blocks called **cells**



## **5. Advantages of LSTM**

1. Long-term dependencies can be captured by LSTM networks. They have a memory cell that is capable of long-term information storage.
2. In traditional RNNs, there is a problem of vanishing and exploding gradients when models are trained over long sequences. By using a gating mechanism that selectively recalls or forgets information, LSTM networks deal with this problem.
3. LSTM enables the model to capture and remember the important context, even when there is a significant time gap between relevant events in the sequence. So where understanding context is important, LSTMs are used. eg. machine translation.

## **6. Disadvantages of LSTM**

1. Compared to simpler architectures like feed-forward neural networks LSTM networks are computationally more expensive. This can limit their scalability for large-scale datasets or constrained environments.
2. Training LSTM networks can be more time-consuming compared to simpler models due to their computational complexity. So training LSTMs often requires more data and longer training times to achieve high performance.
3. Since it is processed word by word in a sequential manner, it is hard to parallelize the work of processing the sentences.

## **7. List the Applications of LSTM?**

- Language Modeling
- Speech Recognition
- Time Series Forecasting
- Anomaly Detection
- Recommender Systems
- Video Analysis
- Robot control
- Time series prediction
- Speech recognition
- Rhythm learning
- Music composition

## 8. What Does Deep Reinforcement Learning Mean?

Deep reinforcement learning (Deep RL) is an approach to machine learning that blends reinforcement learning techniques with strategies for deep learning.

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

## 9. Difference between Reinforcement learning and Supervised learning

Reinforcement learning	Supervised learning
Reinforcement learning is all about making decisions sequentially. In simple words, we can say that the output depends on the state of the current input and the next input depends on the output of the previous input	In Supervised learning, the decision is made on the initial input or the input given at the start
In Reinforcement learning decision is dependent, So we give labels to sequences of dependent decisions	In supervised learning the decisions are independent of each other so labels are given to each decision.
Example: Chess game, text summarization	Example: Object recognition, spam detection

## 10. List and explain the Types of Reinforcement learning?

There are two types of Reinforcement:

1. **Positive:** Positive Reinforcement is defined as when an event, occurs due to a particular behavior, increases the strength and the frequency of the behavior. In other words, it has a positive effect on behavior.
2. **Negative:** Negative Reinforcement is defined as strengthening of behavior because a negative condition is stopped or avoided.

## **11. List the Elements of Reinforcement Learning**

Reinforcement learning elements are as follows:

1. Policy
2. Reward function
3. Value function
4. Model of the environment

## **12. Advantages of Reinforcement learning?**

1. Reinforcement learning can be used to solve very complex problems that cannot be solved by conventional techniques.
2. The model can correct the errors that occurred during the training process.
3. In RL, training data is obtained via the direct interaction of the agent with the environment
4. Reinforcement learning can handle environments that are non-deterministic, meaning that the outcomes of actions are not always predictable. This is useful in real-world applications where the environment may change over time or is uncertain.
5. Reinforcement learning can be used to solve a wide range of problems, including those that involve decision making, control, and optimization.
6. Reinforcement learning is a flexible approach that can be combined with other machine learning techniques, such as deep learning, to improve performance.

## **13. Disadvantages of Reinforcement learning?**

1. Reinforcement learning is not preferable to use for solving simple problems.
2. Reinforcement learning needs a lot of data and a lot of computation
3. Reinforcement learning is highly dependent on the quality of the reward function. If the reward function is poorly designed, the agent may not learn the desired behavior.
4. Reinforcement learning can be difficult to debug and interpret. It is not always clear why the agent is behaving in a certain way, which can make it difficult to diagnose and fix problems.

## **14. Optimization in Deep Learning?**

In Deep Learning, with the help of loss function, the performance of the model is estimated/ evaluated. This loss is used to train the network so that it performs better. Essentially, we try to minimize the Loss function. Lower Loss means the model performs better. The Process of minimizing any mathematical function is called Optimization.

## 15. What is the Need for Optimization?

- Presence of Local Minima reduces the model performance
- Presence of Saddle Points which creates Vanishing Gradients or Exploding Gradient Issues
- To select appropriate weight values and other associated model parameters
- To minimize the loss value (Training error)

## 16. What is Convex Optimization?

**Convex optimization** is a kind of optimization which deals with the study of problem of minimizing convex functions. Here the optimization function is convex function. All Linear functions are convex, so linear programming problems are convex problems. When we have a convex objective and a convex feasible region, then there can be only one optimal solution, which is globally optimal.

## 17. What is Non-Convex Optimization?

- The Objective function is a non-convex function
- All non-linear problems can be modeled by using non-convex functions. (Linear functions are convex)
- It has multiple feasible regions and multiple locally optimal points.
- There can't be a general algorithm to solve it efficiently in all cases
- Neural networks are universal function approximators, to do this, they need to be able to
- approximate non-convex functions.

## 18. How to solve non-convex problems?

- Stochastic gradient descent
- Mini-batching
- SVRG
- Momentum

## 19. Reasons For Non-Convexity?

- Presence of many Local Minima
- Presence of Saddle Points
- Very Flat Regions
- Varying Curvature

## **20. What is Spatial Transform Network [STN]?**

Spatial Transformer Network (SSTN) helps to crop out and scale-normalizes the appropriate region, which can simplify the subsequent classification task and lead to better classification performance. The Spatial Transformer Network contains three parts Namely, Localization, Grid Generator and Sampler. These Networks are used for performing Transformations such as Cropping, Rotation etc on the given input images.

## **UNIT V APPLICATIONS OF DEEP LEARNING**

### **1. Define Imagenet?**

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. In Machine Learning and Deep Neural Networks, machines are trained on a large dataset of various images. Machines are required to learn useful features from these training images. Once learned, they can use these features to classify images and perform many other tasks associated with computer vision. ImageNet gives researchers a common set of images to benchmark their models and algorithms.

### **2. How the images are labeled in ImageNet?**

In the early stages of the ImageNet project, a quick calculation showed that by employing a few people, they would need 19 years to label the images collected for ImageNet. But in the summer of 2008, researchers came to know about an Amazon service called Mechanical Turk. This meant that image labeling can be crowd sourced via this service. Humans all over the world would label the images for a small fee

### **3. How the images of ImageNet Licensed?**

Images for ImageNet were collected from various online sources. ImageNet doesn't own the copyright for any of the images. This has implication on how ImageNet shares the images to researchers. For public access, ImageNet provides image thumbnails and URLs from where the original images were downloaded. Researchers can use these URLs to download the original images. However, those who wish to use the images for non-commercial or educational purpose can create an account on ImageNet and request access. This will allow direct download of images from ImageNet.

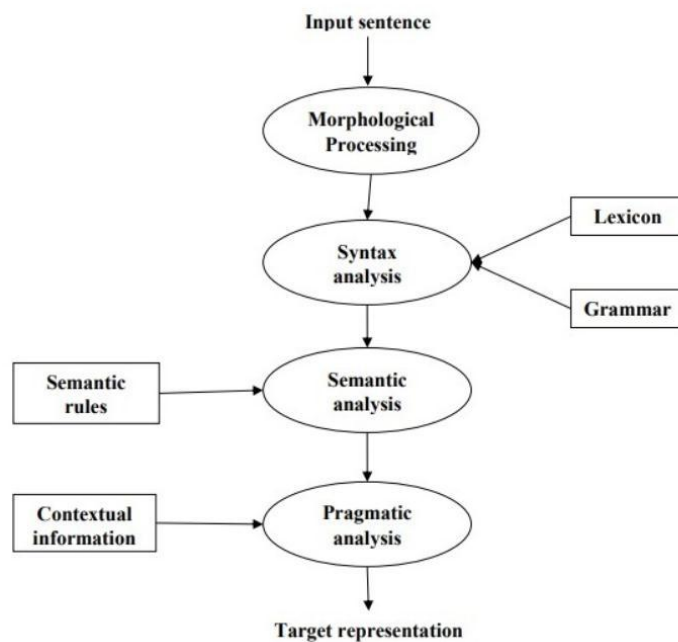
#### 4. What is WaveNet?

WaveNet is a deep generative model of raw audio waveforms. We show that WaveNets are able to generate speech which mimics any human voice and which sounds more natural than the best existing Text-to-Speech systems, reducing the gap with human performance by over 50%. Allowing people to converse with machines is a long-standing dream of human computer interaction. The ability of computers to understand natural speech has been revolutionized in the last few years by the application of deep neural networks.

#### 5. Define Natural Language Processing [NLP]?

Language is a method of communication with the help of which we can speak, read and write. For example, we think, we make decisions, plans and more in natural language; precisely, in words. However, the big question that confronts us in this AI era is that can we communicate in a similar manner with computers. In this sense, we can say that Natural Language Processing (NLP) is the sub-field of Computer Science especially Artificial Intelligence (AI) that is concerned about enabling computers to understand and process human language. Technically, the main task of NLP would be to program computers for analysing and processing huge amount of natural language data.

#### 6. List the Natural Language Processing Phases?



## **7. Define Morphological Processing?**

It is the first phase of NLP. The purpose of this phase is to break chunks of language input into sets of tokens corresponding to paragraphs, sentences and words. For example, a word like “**uneasy**” can be broken into two sub-word tokens as “**un-easy**”.

## **8. Define Syntax Analysis?**

It is the second phase of NLP. The purpose of this phase is two folds: to check that a sentence is well formed or not and to break it up into a structure that shows the syntactic relationships between the different words. For example, the sentence like “**The school goes to the boy**” would be rejected by syntax analyser or parser.

## **9. Define Semantic Analysis?**

It is the third phase of NLP. The purpose of this phase is to draw exact meaning, or you can say dictionary meaning from the text. The text is checked for meaningfulness. For example, semantic analyser would reject a sentence like “Hot ice-cream”.

## **10. What is Pragmatic Analysis?**

It is the fourth phase of NLP. Pragmatic analysis simply fits the actual objects/events, which exist in a given context with object references obtained during the last phase (semantic analysis).

For example, the sentence “Put the banana in the basket on the shelf” can have two semantic.

## **11. Applications of NLP:**

- Spam Filters
- Algorithmic Trading
- Answering Questions
- Summarizing Information's .

## 12. Define Word2Vec?

Word embedding is one of the most popular representations of document vocabulary. It is capable of capturing context of a word in a document, semantic and syntactic similarity, relation with other words, etc. What are word embeddings exactly? Loosely speaking, they are vector representations of a particular word. Having said this, what follows is how do we generate them? More importantly, how do they capture the context? Word2Vec is one of the most popular technique to learn word embeddings using shallow neural network. It was developed by Tomas Mikolov in 2013 at Google.

## 13. List the types of Pooling?

- Max Pooling
- Average Pooling
- Sum Pooling

## 14. Steps involved in face recognition?

1. **Face detection:** This is the basic step, where the face is detected and bounding boxes are drawn around it.
2. **Face alignment:** The detected face is normalized to speed up training. Experiments show that alignment alone increases face recognition accuracy by almost 1%.
3. **Feature extraction:** Local features are extracted from the image with the help of algorithms.
4. **Face recognition:** This is the last stage and involves matching the input face with images present in the dataset to identify who it belongs to.

## 15. Challenges in face recognition systems?

The accuracy of facial recognition systems depends upon certain factors. And these factors themselves present a challenge.

- **Pose:** How a person poses is very important for successful recognition. If the face is not clearly visible, the facial recognition system will fail.

- **Facial expression:** Sometimes, different facial expressions can result in different predictions for the same person's image.

- **Illumination:** Images or video streams with poor lighting can result in bad predictions.

- **Resolution:** Low-resolution pictures contain less information and, hence, are not considered suitable for recognition purposes

## **16. What is Scene Understanding?**

Scene understanding is the process, often real time, of perceiving, analyzing and elaborating an interpretation of a 3D dynamic scene observed through a network of sensors. This process consists mainly in matching signal information coming from sensors observing the scene with models which humans are using to understand the scene. Based on that, scene understanding is both adding and extracting semantic from the sensor data characterizing a scene. This scene can contain a number of physical objects of various types (e.g. people, vehicle) interacting with each other's or with their environment (e.g. equipment) more or less structured.

## **17. Define Image Captions?**

Image Captioning is the process of generating textual description of an image. It uses both Natural Language Processing and Computer Vision to generate the captions.