1. What do you mean by bias and variance?

Ans.

Variance is the amount that the estimate of the target function will change given different training data.

2. What do you mean by overfitting and underfitting?

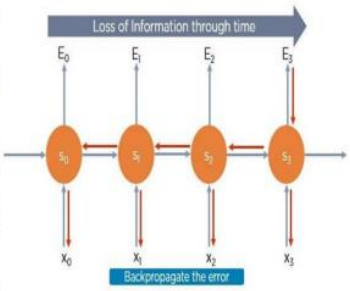
Underfitting is a situation when your model is **too simple** for your data. More formally, your hypothesis about data distribution is wrong and too simple — for example, your data is quadratic and your model is linear. This situation is also called **high bias**. This means that your algorithm can do accurate predictions, but the initial assumption about the data is incorrect.

Opposite, **overfitting** is a situation when your model is **too complex** for your data. More formally, your hypothesis about data distribution is wrong and too complex — for example, your data is linear and your model is high-degree polynomial. This situation is also called **high variance**. This means that your algorithm can't do accurate predictions — changing the input data only a little, the model output changes very much.

3. What do you mean by vanishing and exploding gradient?

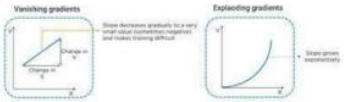
Vanishing Gradient Problem:

- RNN is hard to train because of the gradient problem.
- RNNs suffer from the problem of vanishing gradients.
- The gradients carry information used in the RNN, and when the gradient becomes too small, the parameter updates become insignificant.
- This makes the learning of long data sequences difficult.



Exploding Gradient Problem:

- While training a neural network, if the slope tends to grow exponentially instead of decaying, this is called an Exploding Gradient.
- This problem arises when large error gradients accumulate, resulting in very large updates to the neural network model weights during the training process.
- Long training time, poor performance, and bad accuracy are the major issues in gradient problems.



4. What are the different activation functions? Give the significance of the tanh activation function in LSTM.

Activation Function helps the neural network to use important information while suppressing irrelevant data points.

Binary Step Function Binary step function depends on a threshold value that decides whether a neuron should be activated or not.

Linear Activation Function The linear activation function, also known as "no activation," or "identity function" (multiplied x1.0), is where the activation is proportional to the input.

Non-Linear Activation Functions

Types of Non-Linear activation functions:

1. Sigmoid
2. Tanh
3. ReLU
4. Leaky ReLU
5. Parameterised ReLU
6. Exponential Linear Unit
7. Swish
8. Softmax

In Tanh, the larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to -1.0.In LSTM network, tanh activation function is used to determine candidate cell state (internal state) values (\tilde{C}_{t}) and update the hidden state (h_{t}).

5. What is the importance of max pool layer in CNN?

Ans. Pooling layers are used **to reduce the dimensions of the feature maps**. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. The pooling layer summarises the features present in a region of the feature map generated by a convolution layer.

6. What do you mean by stride in CNN

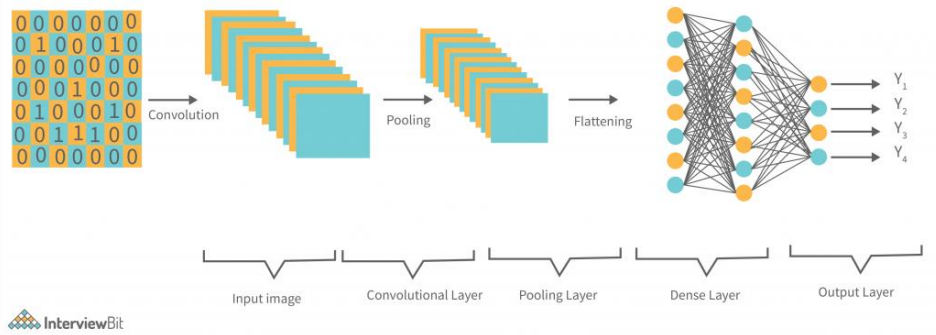
Ans

Stride is a component of convolutional neural networks, or neural networks tuned for the compression of images and video data. Stride is a parameter of the neural network's filter that modifies the amount of movement over the image or video. For example, if a neural network's stride is set to 1, the filter will move one pixel, or unit, at a time. The size of the filter affects the encoded output volume, so stride is often set to a whole integer, rather than a fraction or decimal.

7. Explain in detail the basic architecture of CNN and give an example of any CNN architecture (case study).

Ans.

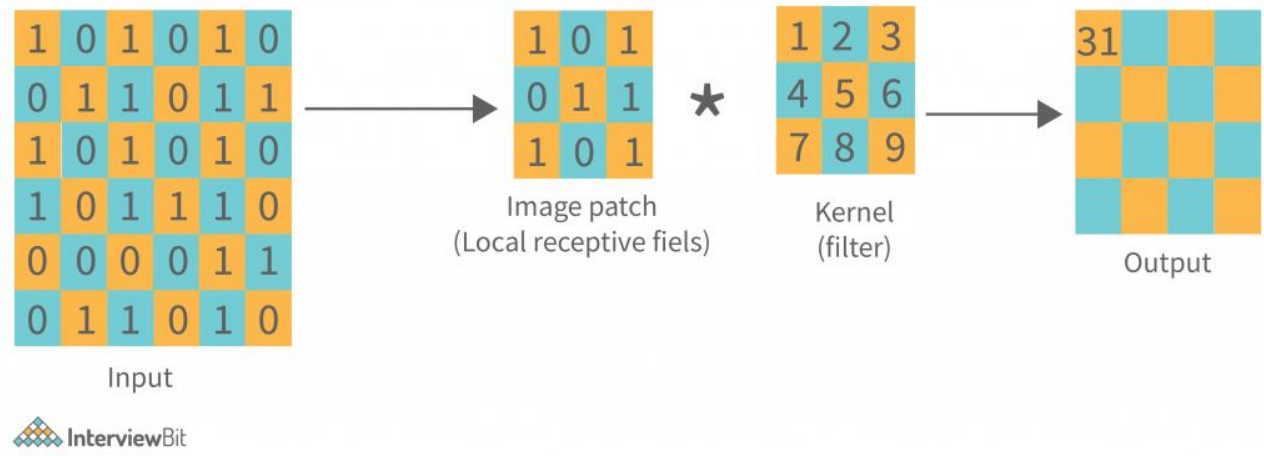
Convolutional Neural Networks (CNN, or ConvNet) are a type of multi-layer neural network that is meant to discern visual patterns from pixel images. In CNN, ‘convolution’ is referred to as the mathematical function. It’s a type of linear operation in which you can multiply two functions to create a third function that expresses how one function’s shape can be changed by the other. In simple terms, two images that are represented in the form of two matrices, are multiplied to provide an output that is used to extract information from the image. CNN is similar to other neural networks, but because they use a sequence of convolutional layers, they add a layer of complexity to the equation. CNN cannot function without convolutional layers.



The ConvNet's job is to compress the images into a format that is easier to process while preserving elements that are important for obtaining a decent prediction. This is critical for designing an architecture that is capable of learning features while also being scalable to large datasets.

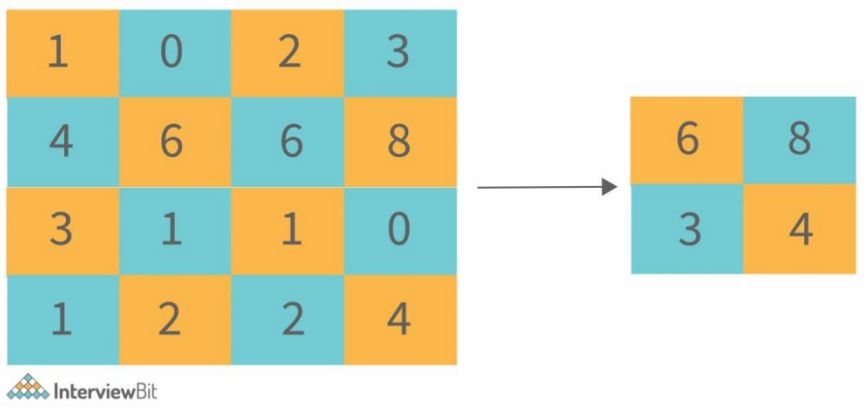
A convolutional neural network, ConvNets in short has three layers which are its building blocks, let's have a look:

Convolutional Layer (CONV): They are the foundation of CNN, and they are in charge of executing convolution operations. The Kernel/Filter is the component in this layer that performs the convolution operation (matrix). Until the complete image is scanned, the kernel makes horizontal and vertical adjustments dependent on the stride rate. The kernel is less in size than a picture, but it has more depth. This means that if the image has three (RGB) channels, the kernel height and width will be modest spatially, but the depth will span all three.

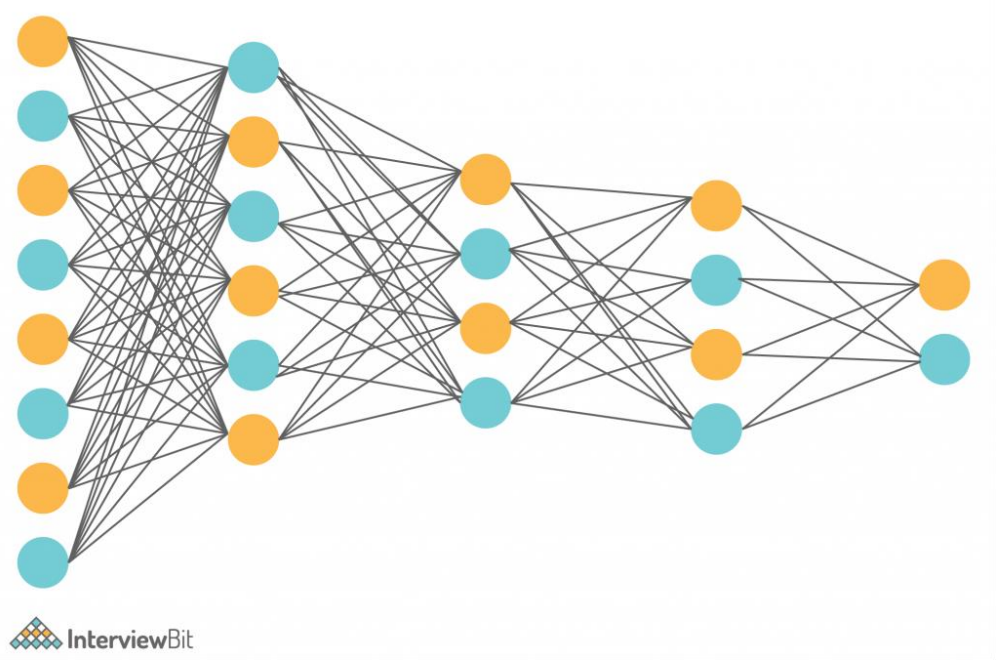


Other than convolution, there is another important part of convolutional layers, known as the Non-linear activation function. The outputs of the linear operations like convolution are passed through a non-linear activation function. Although smooth nonlinear functions such as the sigmoid or hyperbolic tangent (tanh) function were formerly utilized because they are mathematical representations of biological neuron actions. The rectified linear unit (ReLU) is now the most commonly used non-linear activation function. $f(x) = \max(0, x)$

Pooling Layer (POOL): This layer is in charge of reducing dimensionality. It aids in reducing the amount of computing power required to process the data. Pooling can be divided into two types: maximum pooling and average pooling. The maximum value from the area covered by the kernel on the image is returned by max pooling. The average of all the values in the part of the image covered by the kernel is returned by average pooling.



Fully Connected Layer (FC): The fully connected layer (FC) works with a flattened input, which means that each input is coupled to every neuron. After that, the flattened vector is sent via a few additional FC layers, where the mathematical functional operations are normally performed. The classification procedure gets started at this point. FC layers are frequently found near the end of CNN architectures if they are present.



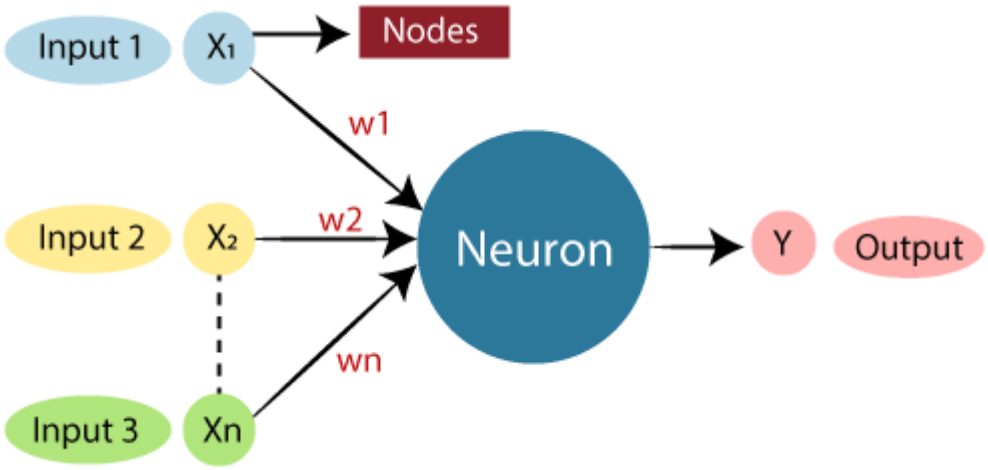
8. What do you mean by LSTM? Explain each and every component of LSTM in detail.

9. What do you mean by ANN? Explain it with disadvantages of ANN.

Ans.

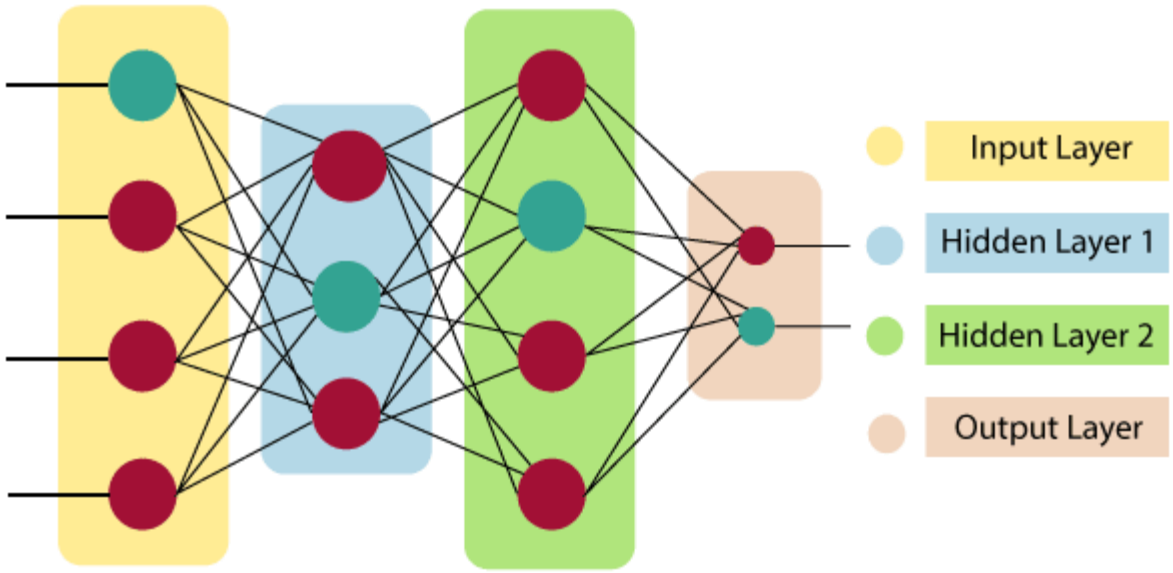
The term "**Artificial Neural Network**" is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers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

The typical Artificial Neural Network looks something like the given figure.



An **Artificial Neural Network** in the field of **Artificial intelligence** where it attempts to mimic the network of neurons makes up a human brain so that computers will have an option to understand things and make decisions in a human-like manner. The artificial neural network is designed by programming computers to behave simply like interconnected brain cells.
To understand the concept of the architecture of an artificial neural network, we have to understand what a neural network consists of. In order to define a neural network that consists of a large number of artificial neurons, which are termed units arranged in a sequence of layers. Lets us look at various types of layers available in an artificial neural network.

Artificial Neural Network primarily consists of three layers:



Input Layer:

As the name suggests, it accepts inputs in several different formats provided by the programmer.

Hidden Layer:

The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer:

The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.

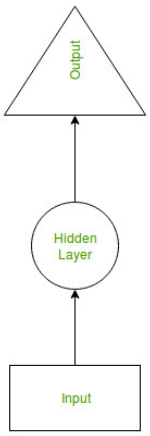
$$\sum_{i=1}^{n} Wi * Xi + b$$

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

Disadvantages of Artificial Neural Network:

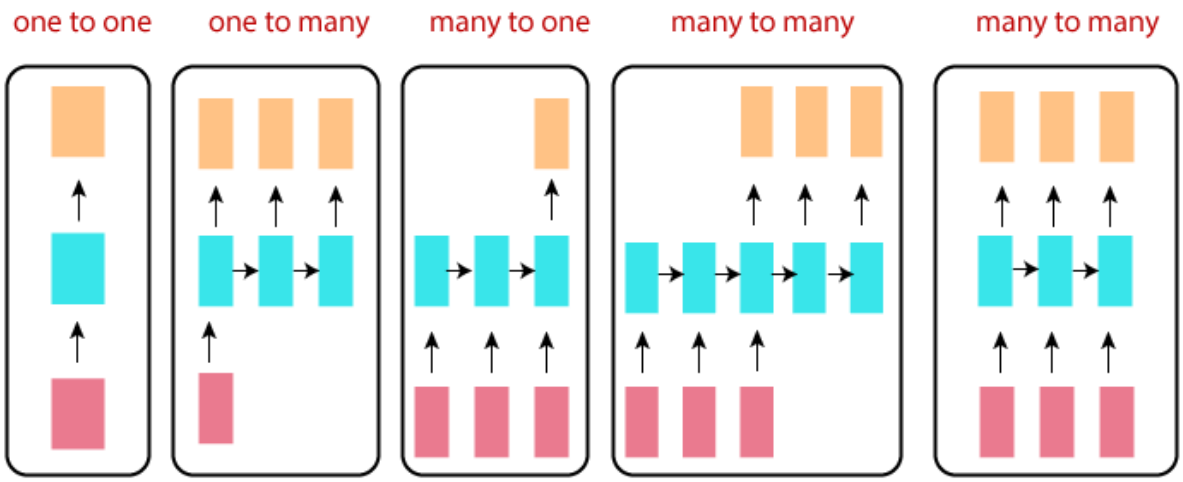
Assurance of proper network structure:
There is no particular guideline for determining the structure of artificial neural networks. The appropriate network structure is accomplished through experience, trial, and error.
Unrecognized behavior of the network:
It is the most significant issue of ANN. When ANN produces a testing solution, it does not provide insight concerning why and how. It decreases trust in the network.
Hardware dependence:
Artificial neural networks need processors with parallel processing power, as per their structure. Therefore, the realization of the equipment is dependent.
Difficulty of showing the issue to the network:
ANNs can work with numerical data. Problems must be converted into numerical values before being introduced to ANN. The presentation mechanism to be resolved here will directly impact the performance of the network. It relies on the user's abilities.
The duration of the network is unknown:
The network is reduced to a specific value of the error, and this value does not give us optimum results.

10. What do you mean by RNN? Give the different types of RNN with their applications
Ans.
Recurrent Neural Network(RNN) are a type of Neural Network where the **output from previous step are fed as input to the current step**. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like 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 **Hidden state**, which remembers some information about a sequence.



RNN have a **"memory"** which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

The main reason that the recurrent nets are more exciting is that they allow us to operate over sequences of vectors: Sequence in the input, the output, or in the most general case, both. A few examples may this more concrete:



One-to-one:

This is also called **Plain Neural networks**. It deals with a fixed size of the input to the fixed size of output, where they are independent of previous information/output.

Example: Image classification.

One-to-Many:

It deals with a fixed size of information as input that gives a sequence of data as output.

Example: Image Captioning takes the image as input and outputs a sentence of words.

Many-to-One:

It takes a sequence of information as input and outputs a fixed size of the output.

Example: sentiment analysis where any sentence is classified as expressing the positive or negative sentiment.

Many-to-Many:

It takes a Sequence of information as input and processes the recurrently outputs as a Sequence of data.

Example: Machine Translation, where the RNN reads any sentence in English and then outputs the sentence in French.

11. ANN vs CNN

Ans.

ANN	*CNN*	*RNN*	
Type of Data	Tabular Data, Text Data	Image Data	Sequence data
Parameter Sharing	No	Yes	Yes
Fixed Length input	Yes	Yes	No
Recurrent Connections	No	No	Yes
Vanishing and Exploding Gradient	Yes	Yes	Yes
Spatial Relationship	No	Yes	No
Performance	ANN is considered to be less powerful than CNN, RNN.	CNN is considered to be more powerful than ANN, RNN.	RNN includes less feature compatibility when compared to CNN.
Application	Facial recognition and Computer vision.	Facial recognition, text digitization and Natural language processing.	Text-to-speech conversions.
Main advantages	Having fault tolerance, Ability to work with incomplete knowledge.	High accuracy in image recognition problems, Weight sharing.	Remembers each and every information, Time series prediction.
Disadvantages	Hardware dependence, Unexplained behavior of the network.	Large training data needed, don't encode the position and orientation of object.	Gradient vanishing, exploding gradient.

12. RNN vs LSTM

13. What are the different performance measures of classification algorithms?(pdf)

14. What is ensemble learning?

15. Difference between Random forest and Adaboost

Ans.

S. No.	Categories	Random Forest	AdaBoost
1.	Tree Size	Random forest uses a full-sized Decision tree with no predetermined depth size.	AdaBoost combines a lot of “weak learners” to make classifications. The weak learners are almost always stumps.
2.	Distribution of say for each tree.	Each tree has an equal vote on the final classification	Some stumps get more say in the classification than others.
3.	Classification Accuracy	Random Forest often gets outperformed by AdaBoost in terms of classification accuracy.	AdaBoost is often much better at making accurate classifications.
4.	Tree Construction Order	Each decision tree is made independently of the others	Each stump is made by taking the previous stump’s mistakes into account.
5.	Overfitting Tolerance	Random Forest is less sensitive to overfitting as compared to AdaBoost	Adaboost is also less tolerant to overfitting than Random Forest.
6.	Data Sampling Technique	In Random forest, the training data is sampled based on the bagging technique.	Adaboost is based on boosting technique.
7.	Estimate Calculation	Random Forest aims to decrease variance not bias.	Adaboost aims to decrease bias, not variance.

S. No.	Categories	Random Forest	AdaBoost
8.	Ensembling Operation	Random Forest employs parallel assembly. Forest processes trees in parallel, allowing jobs to be parallelized on a multiprocessor machine.	Adaboost makes use of sequential ensembling. It takes a step-by-step method.

16. What do you mean by bootstrapping?

17. What do you mean by bagging and boosting?

18. What are the different classification algorithms?

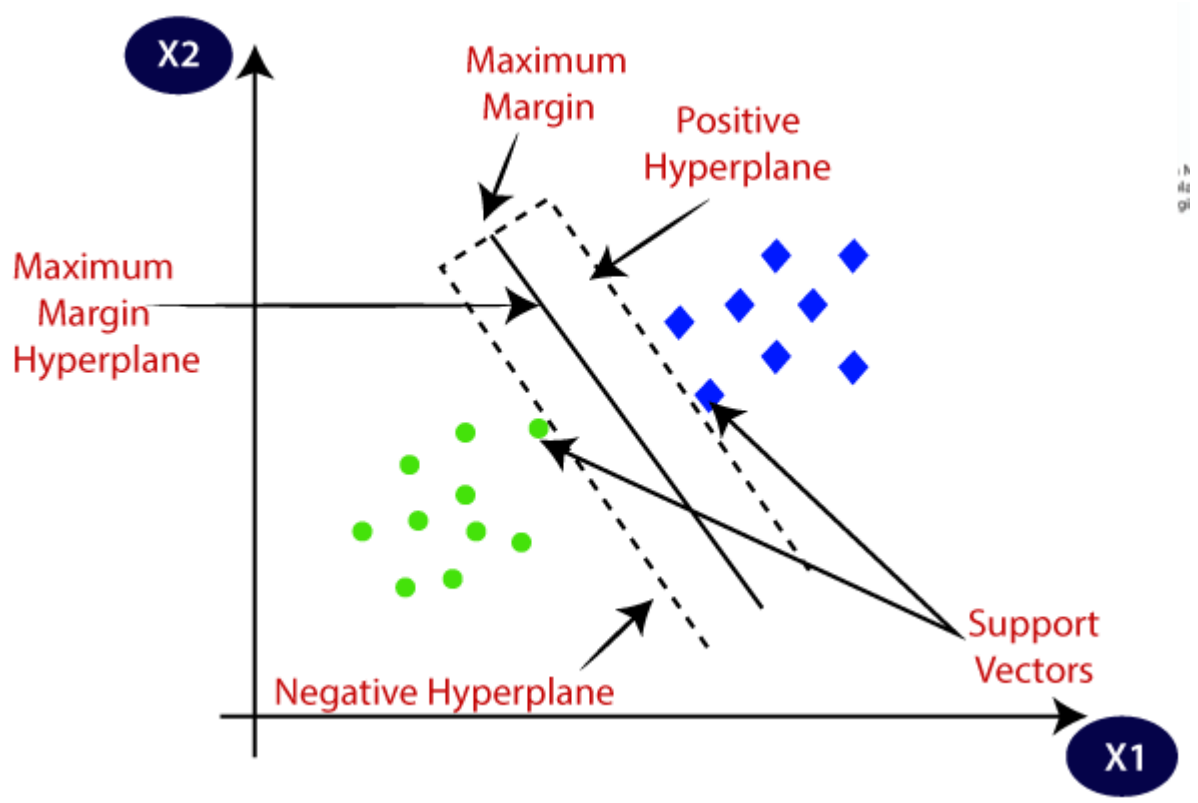
19. Explain SVM in detail

Ans

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:

SVM algorithm can be used for **Face detection, image classification, text categorization,** etc.

Types of SVM

SVM can be of two types:

- **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.

Hyperplane and Support Vectors in the SVM algorithm:

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors:

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

20. How the decision algorithm works and explain with diagrams and examples.

21. Differentiate between biological neural networks and artificial neural networks.

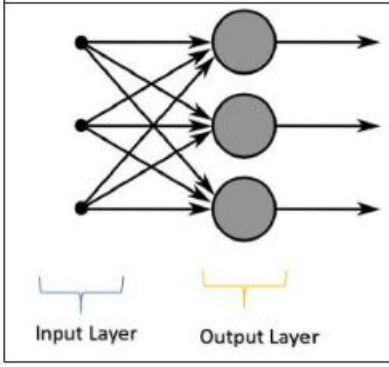
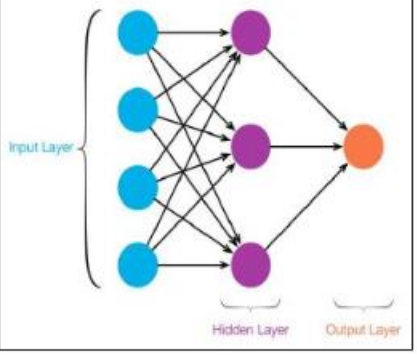
Ans.

S.No.	ANN	BNN
1.	It is short for Artificial Neural Network.	It is short for Biological Neural Network.
2.	Processing speed is fast as compared to Biological Neural Network.	They are slow in processing information.
3.	Allocation for Storage to a new process is strictly irreplaceable as the old location is saved for the previous process.	Allocation for storage to a new process is easy as it is added just by adjusting the interconnection strengths.
4.	Processes operate in sequential mode.	The process can operate in massive parallel operations.
5.	If any information gets corrupted in the memory it cannot be retrieved.	Information is distributed into the network throughout into sub-nodes, even if it gets corrupted it can be retrieved.
6.	The activities are continuously monitored by a control unit.	There is no control unit to monitor the information being processed into the network.

22. Draw and explain McCulloch Pitts neuron architecture. (NB)

23. Differentiate between Multi-layer and Single layer feed forward networks.

Ans.

Single Layer Feed-Forward Neural Network	Multi Layer Feed-Forward Neural Network
Layer is formed by taking processing element & combining it with other processing element.	It is formed by interconnection of several layers.
Input & output are linked with each other.	There are multiple layers between input & output layers which are known as hidden layers.
Inputs are connected to the processing nodes with various weights resulting series of output one per node.	Input layers receives input & buffers input signal, output layer generates output.
Zero hidden layers are present.	Zero to several hidden layers are in a network.
Not efficient in certain areas.	More the hidden layers, more the complexity of networks, but efficient output is produced.
	

24. Implement XOR function using Mc-Culloch Pitts neuron. (NB)
25. Implement OR function using Mc-Culloch Pitts neuron. (NB)
26. Write a short note on Autoencoder. (PPt)
27. List types of Autoencoders and write advantages and disadvantages of each. (PPt)s
28. What is Reinforcement Learning? Explain the significance of Award and Action in RL.
29. What is the class imbalance problem ? Explain it with an example.
30. Write a short note on the Holdout method.
31. Explain how we can compare classifiers based on cost benefit and ROC curve.