1) Machine learning

It is a superset of Deep learning

uses structured data

Consists of thousands of data points

Training can be performed used the CPU

Models take less time in training

Deep learning

It is a subset of machine learning

uses neural networks

big data

GPU is required for training

a huge amount of time is taken for training

2) A perceptron is the most basic type of artificial neural network, functioning as a single-layer binary classifier that makes decisions based on a linear combination of input features.

3) The activation functions help the network use the imp information and suppress the irrevelant data pts.
It will decide whether the neuron's i/p to the n/wk is relevant or not.
are used to map the i/p b/w (0,1) or (-1, 1)

$$y = \text{Activation function} \sum (\text{weight} * i/p) + \text{bi'y.}$$

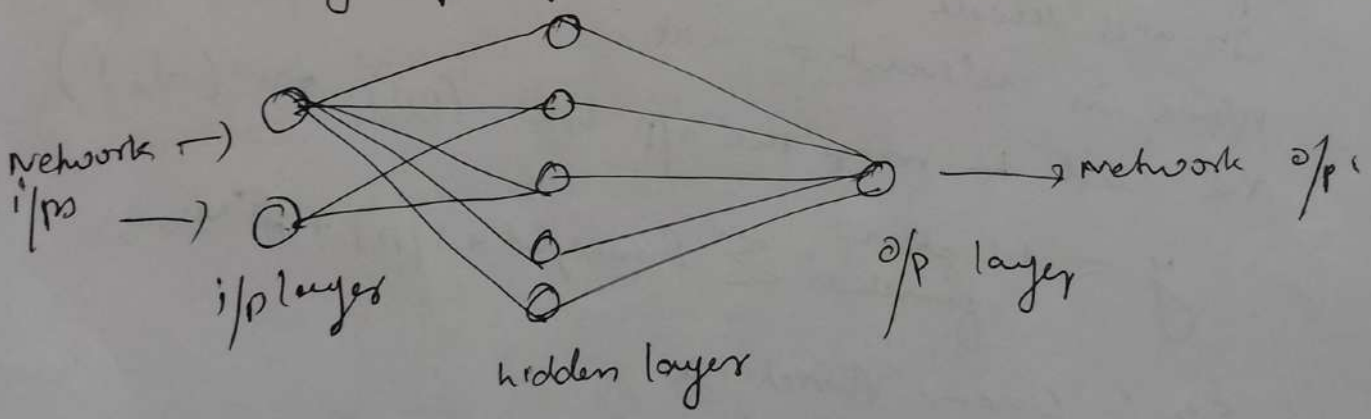
Eg: Linear function
sigmoid function

4) In deep learning, an epoch refers to one complete pass through the entire training dataset- where every data sample is passed through the model and its parameters are updated based on the calculated error. The training process requires multiple epochs allowing the model to improve

on the calculated errors.

5) Overfitting in a neural network occurs when the network learns the training data too well, including its noise and irrelevant details, leading to excellent performance on the training data but poor generalization on new, unseen data. The model becomes overly complex and specialized for training set, failing to capture underlying patterns that would allow it to make accurate predictions on data it hasn't encountered before.

8) There are 3 layers in the network architecture. The i/p layer, hidden layer and the o/p layer. A typical feedforward network processes information in one direction from i/p to o/p. Because of the numerous layers are sometimes referred as the multi-layer perceptron.



Hidden layer as distillation layer, which extracts some of the most relevant patterns from the i/ps and sends them on to the next layer for further analysis. It accelerates and improves the efficiency of the network by identifying the imp information from the i/ps and removing the redundant information.

The activation function is also imp for 2 reasons.

first it allows you to turn on your computer and it also contributes the conversion of i/p to a more usable final o/p. It also captures the presence of non linear relationships btw the i/p's.

9)

Supervised	Unsupervised
uses labelled data	uses unlabelled data
predicts outcomes based on known labels	discovers hidden patterns, structures in data
less complex	more complex
classification & regression	clustering & association

11) ReLU function is a piecewise linear function that o/ps the i/p directly if it is true, otherwise it o/ps zero. ReLU allows positive values to pass through unchanged while setting all negative values to zero. This helps the neural network maintain the necessary complexity to learn patterns while avoiding some of the pitfalls associated with other activation functions

$$f(x) = \max(0, x)$$

$x \rightarrow$ i/p to the neuron

function returns x, if $x > 0$

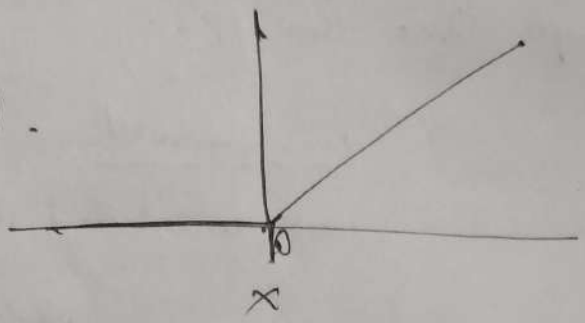
if $x \leq 0$, the function returns 0

also written as

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

It helps to maintain non-linearity without complicated transformations

ReLU(x).



(3) A loss function is a mathematical way to measure how good or bad a model's predictions are compared to the actual results. Loss functions are used to train models. It gives a single number that tells us how far off the predictions are. Algorithms use the loss function to adjust the model's parameters and to reduce the error and improve prediction.
It helps to improve the model's performance by finding the difference b/w predicted and actual value

Eg: for Regression tasks.

a) Mean Squared Error → It calculates the avg of squared difference b/w the predicted values and actual values.

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

for classification tasks

a) Categorical Cross-Entropy loss

for multiclass classification problems. It measures the performance of a classification model whose o/p is a probability distribution over multiple classes.

14) The CNNs are used for

a) Image Recognition and classification → used to identify and classify objects, faces

b) computer vision - are core component of computer vision systems, enabling machines to interpret and understand visual information

c) Medical Imaging → detect conditions like diabetic retinopathy

d) Autonomous driving

e) Natural Language Processing → identifying keywords in audio data

f) Handwritten Digit Recognition.

15) Back Propagation is a method used to train ~~model~~ neural networks. It's goal is to reduce the difference b/w the model's predicted o/p and the actual o/p by adjusting the weights and biases in the network.

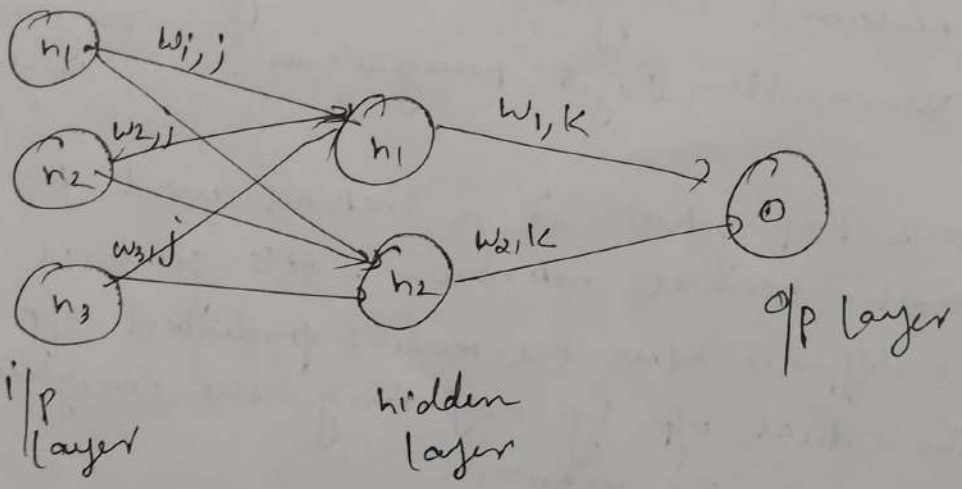
It works iteratively to adjust weights and bias to minimize the cost function. In each epoch the model adapts these parameters by reducing loss by following the error gradient. The algorithm computes the gradient using the chain rule from calculus allowing it to effectively navigate complex layers in the neural network to minimize the cost function.

2 steps forward Pass and Backward Pass

(1) In this the i/p data is fed into the i/p layer. These i/ps are combined with their respective weights are passed to hidden layers. for eg.

In a network with 2 hidden layers, h1 and h2, the o/p from h1 serves as the i/p to h2. Before applying an activation function, a bias is added to the weighted i/ps.

Each hidden layer computes the weighted sum ('a') of the i/ps then applies an activation function like ReLU to obtain the o/p ('o'). The o/p is passed to the next layer where an activation function converts the weighted o/ps into probabilities for classifications



Backward Pass

In this the error is propagated back through the network to adjust the weights and biases. One common method is mean Squared Error

$$MSE = \left(\text{Predicted o/p} - \text{Actual o/p}\right)^2$$

Once the error is calculated the network adjusts weights using gradients which are computed with the chain rule. These gradients indicate how much each weight and bias should be adjusted to minimize the error in the next iteration. The backward pass continues layer by layer ensuring that the network learns and improves its performance.

16) Convolutional Neural Networks are deep learning models designed to process data with a grid like topology such as images. They are the foundation for most modern computer vision applications to detect features within visual data.

CNN consists of

Convolutional layers:-
These layers apply convolutional operations to inputs images using filters or kernels to detect features such as edges, textures. Convolutional operations help preserve the spatial relationships between pixels.

Pooling layers:-
They downsample the spatial dimensions of the input reducing the computational complexity and the no: of parameters in the network. Max pooling is a common pooling operation where we select a maximum value from a group of neighbouring pixels.

Activation functions:-
They introduce non-linearity to the model by allowing it to learn more complex relationships in the data.

fully connected layers
These layers are responsible for making predictions based on the high level features learned by the previous layers. They connect every neuron in one layer to every neuron in the next layer.

CNN works

i/p image :- CNN receives an i/p image which is preprocessed to ensure uniformity in size and format.

convolutional layers.
filters are applied to the i/p image to extract features like shapes, edges.

Pooling layers : The feature maps generated by the convolutional layers are downsampled to reduce dimensionality.

fully connected layers : The downsampled feature maps are passed through fully connected layers to produce the final o/p.

o/p : CNN o/ps a prediction.