Deep Learning

Introduction to deep learning (DL), MLP - Multilayer Perceptron, Backpropagation, Loss Functions, Hyperparameter tunning, Overview of RNN, CNN and LSTM

Overview of Data science Models; Applications to text, images, videos, recommender systems, Image classification, social network Graphs.

① Introduction to Deep Learning:

Deep Learning is a branch of ML, which is completely based on artificial neural networks. Deep learning is also a kind of mimic of human brain.

Deep Learning is a particular kind of ML that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts.

In human brain approximately 100 billion neurons all together. Each neuron is connected through thousand of their neighbours.

The question is how to recreate these neurons in a computer. So, we create an artificial structure called an artificial neural network. Where we have nodes. or neurons.

we have some neurons for input value and some for output value and in b/w there are lots of neurons interconnected in the hidden layer.

Architectures:-

(i). Deep Neural Network:-

It is a neural network with a certain level of complexity. (Having multiple hidden layers).

They are capable of modeling and processing non-linear relationships.

(ii) Deep Belief Network (DBN):-

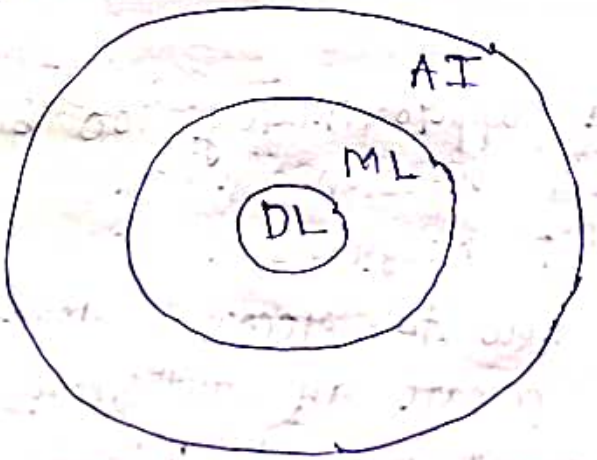
It is class of Deep Neural Network. It is multi layer belief network.

(iii) Recurrent Neural Network:-

RNN performs same task for every element of a sequence.

It allows for parallel and sequential computations, similar to the human brain.

They are able to remember important things about the input.



Difference b/w ML and DL:-

Machine learning	Deep Learning.
(i) Works on small amount of datasets for accuracy.	(i) Works on large amount of datasets.

(i) Lightly dependent on low-end Machine.	(i) Heaving dependent on High-end machine.
(ii) Divides the tasks into sub-tasks, solves them individually and finally combine the results.	(ii) Solves problem end to end.
(iii) Takes less time to train.	(iii) Takes longer time to train.
(iv) Testing time may increase.	(iv) Less time to test the data.

Working:-

(i) Understands the problem and check feasibility for Deep learning.

(ii) Identify the relevant data and prepare it.

(iii) Choose deep learning algorithm.

(iv) Training the algorithm.

(v) Test the model's performance.

Tools Used:- Anaconda, Jupyter, Pycharm etc.

Languages Used:- R, Python, Matlab, CPP, Java, Java Script etc.

Eg:- Recognizing an animal (It is cat (or) dog).

How to recognize square from other shapes.

Limitations:-

(i) Learning through observations only.

(ii) The issue of biases.

Advantages:-

(i) Best in class performance on problems.

(ii) Reduces need for feature engineering.

(iii) Eliminates Unnecessary costs.

(iv) Identifies defects easily that are difficult to detect.

Disadvantages:-

(i) Large amount of data required.
(ii) Computationally expensive to train.
(iii) No strong theoretical foundation.

Applications:-

(i) Automatic Text Generation.
(ii) Healthcare - Helps in diagnosing various diseases and treating it.
(iii) Automatic Machine Translation - DL is achieving top results in the areas of text, images.
(iv) Image Recognition - Recognizes and identifies peoples and objects in images.
(v) Predicting Earthquakes - Teaches a computer to perform viscoelastic computations which are used in predicting earthquakes.

---

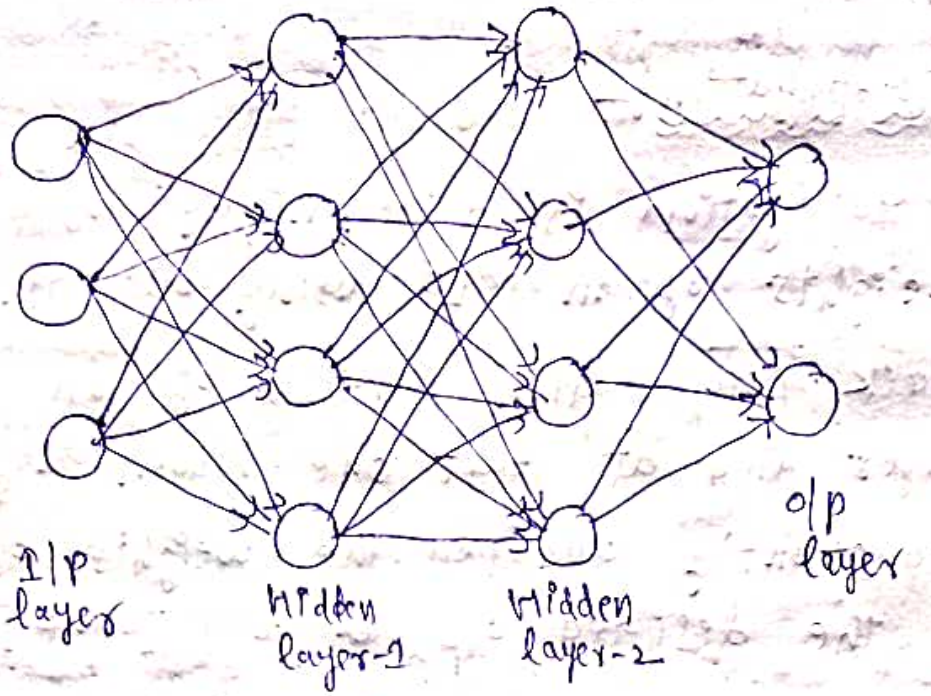
(2) Multilayer Perceptron:-

A MLP is one of the most common neural network used in the field of DL. It is also referred to as "Vanilla" neural network.

The MLP is used for a variety of tasks, such as stock analysis, image identification, spam detection and election voting predictions.

A MLP consists of interconnected neurons transferring information to each other like the human brain.

The network can be divided into 3 main layers.



Input layer:-

This is the first layer of the network which takes an input.

Hidden Layer:-

The network can any network have at least one hidden layer. The hidden layer performs computations and operations on the input data.

Output layer:-

This is the last layer of the network, The nodes or neurons in this layer produce meaningful output.

Connections:-

The MLP is a feed forward neural network. Which means the data or information is transferred from the i/p layer to the output layer in the forward direction.

The connections b/w the layers are referred as as weights. The weight specifies the importance.

This concept is the backbone of an MLP's learning process

Back Propagation:-

Back propagation is a technique used to optimize the weights of an MLP using outputs as inputs.

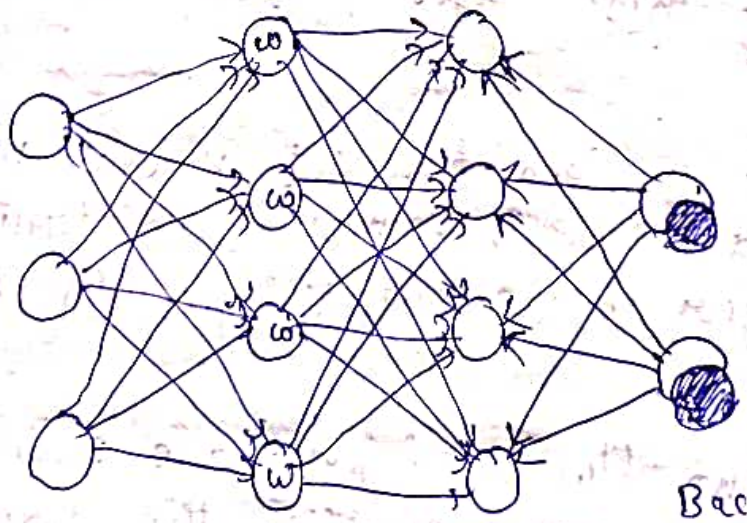
The back propagation consists of an i/p layer, an o/p layer and at least one hidden layer.

It constantly updates the weights of the network, until the desired o/p will come.

It includes the following factors that are responsible for the training and performance of the network.

(i) Random values of weights
(ii) Number of training cycles.
(iii) Number of hidden neurons.
(iv) The training set
(v) Teaching parameter values such as learning rate and momentum.

Working of Back Propagation:-



i/p layer Hidden layers Back propagation o/p layer.

(i) It is fast as well as simple, it is easy to implement.
(ii) It is a standard method that generalizes well.

Deep learning:-

Deep learning is the branch of ML, which is the subset of AI.
Deep learning algorithms are used, especially we have huge number of i/ps and o/ps.
The idea of the deep learning is to build the algorithms that can mimic the brain.
DL is implemented with the help of Neural Networks.

Deep learning is a collection of statistical techniques of ML for learning feature hierarchies that are actually based on artificial neural networks.
When the hidden layers are increased, we are able to solve complex problems.

Types of DL networks:-

(i) CNN - Convolutional Neural Networks.
(ii) RNN - Recurrent Neural Network
(iii) LSTM - Long short - Term Memory Networks
(iv) Stacked Auto Encoders.
(v) DBM - Deep Boltzmann Machine.
(vi) DBN - Deep Belief Networks.

DL Applications:-

(i) Self Driving Cars. (iv) Automatic Image Caption Generation
(ii) Voice controlled Assistance. (v) Automatic Machine Translation
Voice)

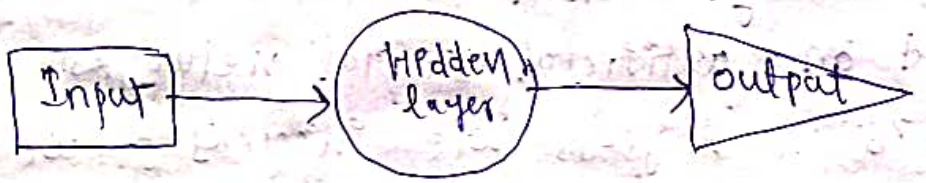
(4) RNN- Recurrent Neural Networks:-

Recurrent Neural Network is a type of Neural Network, where the o/p 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.

The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

In some cases, 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 comes into existance.



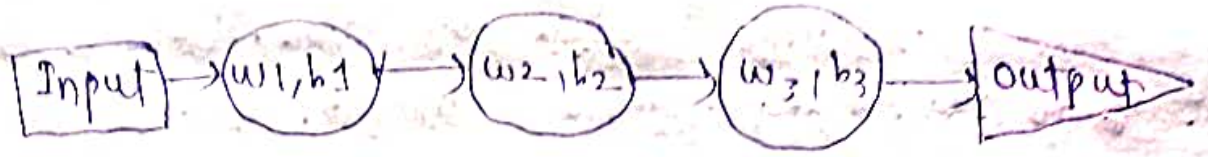
RNN have a memory which remembers all information about what has been calculated.

Working of RNN:-

The working of a RNN can be understood with the help of below example:

ex:- Suppose, a network with one i/p layer, three hidden layers and one o/p layer.

Each hidden layer will have its own set of weights and biases. For hidden layer-1 weights and biases are (w_1, b_1), (w_2, b_2) for 2nd and hidden layer and (w_3, b_3) for 3rd hidden layer. That means each of these layers are independent of each other. They do not memorize the previous outputs



(i) RNN converts the independent activations into dependent activations by providing the same weights and biases to all the layers.

(ii) Thus reducing the complexity and increasing parameters and memorizing each previous outputs by giving each o/p as i/p to the next hidden layer.

⇒ h_t = current state,

h_{t-1} = previous state

x_t = input state

Formula for calculating current state,

$$h_t = f(h_{t-1}, x_t)$$

⇒ w_{hh} = weight at recurrent neuron.

w_{xh} = weight at input neuron.

Formula for applying activation function

$$h_t = \tanh(w_{hh} h_{t-1} + w_{xh} x_t)$$

⇒ Formula for calculating output

$$y_t = w_{hy} h_t$$

y_t = output

w_{hy} = weight at output layer.

Advantages of RNN:-

(i) RNN remembers each and every information through time.

(ii) It is useful in time series prediction. Because of the feature to remember previous inputs. This is called LSTM - Long Short Term Memory.

(iii) RNN are even used with convolutional layers.

Disadvantages:-

(i) Gradient vanishing and exploding problems.

(ii) Training an RNN is a very difficult task.

(iii) It cannot process very long sequences.

- Today different ML techniques are used to handle different types of data. One of the most difficult types of data to handle and the forecast is sequential data.

To handle such type of data, the concept of Recurrent Neural Networks was conceived.

While the other artificial neural networks travel in a linear direction during the feed forward process (or back propagation process.

The RNN follows a recurrence relation instead of a feed-forward pass and uses Back-Propagation through time to learn.

---

5) Convolutional Neural Network (CNN):-

Artificial Neural Networks are used in various classification tasks like image, audio, words.

Different types of neural networks are used for different purposes. For example, for predicting the sequence of words we use RNN.

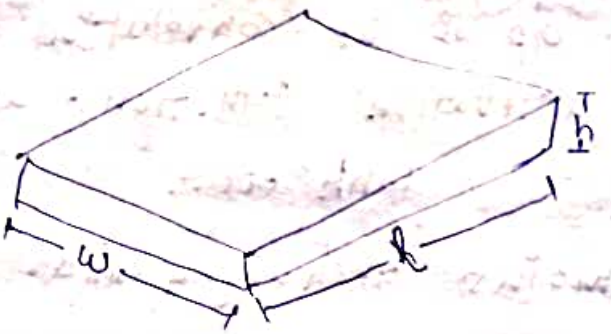
Similarly, for image classification we use CNN.

The data is fed into the model and o/p from each layer is obtained this step is called feed forward.

Then, we calculate the error using error function. Some common error functions are cross-entropy, square loss error etc.

After, we backpropagate into the model by calculating the derivatives. This step is called backpropagation. It is used to minimize the loss.

CNN are neural networks that share their parameters. Imagine you have an image, having length, width and height as shown in the figure.



Now taking a small patch of this image and running a small neural network on it say k outputs and represent them vertically.

Now slide that neural network across the whole image, as a result, we will get another image with different width, height and depth. This operation is called Convolution.

If the patch size is the same as that of the image, it will be regular neural network.

Layers used to build CNN (ConvNets):-

A CNN is a sequence of layers, and every layer transforms one volume to another through a differentiable function.

Types of layers:-

Take an example of image dimension 32x32x3.

Input layer:- This layer holds the raw input of image with width 32, height 32, and depth 3.

Convolution Layer:-

This layer computes the output volume by computing the dot product b/w all filters and image patches. Suppose we have total 12 filters then we will get o/p volume of dimension 32×32×12.

Activation Function Layer:-

This layer will apply an element-wise activation function to the o/p of the convolution layer.

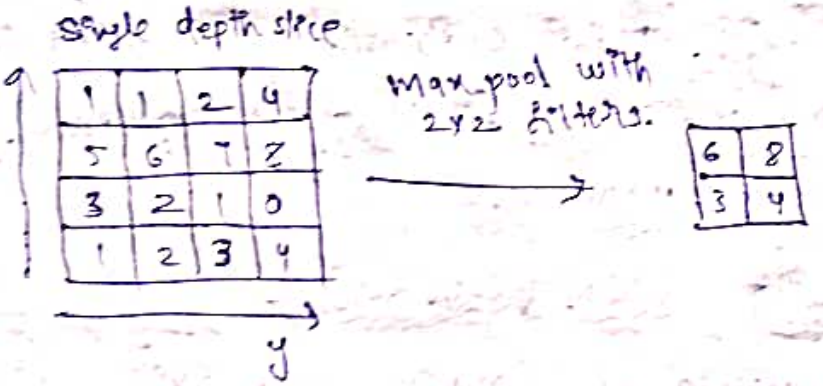
Some common activation functions are RELU : $\max(0, x)$, sigmoid : $\frac{1}{1+e^{-x}}$, Tanh etc..

The volume remains unchanged and have dimension 32×32×12.

Pool Layer:-

Two common types of pooling layers are max pooling and average pooling. It reduce the size of the volume and also prevents overfitting.

The resultant volume will be of dimension 16×16×12.



Fully Connected Layer:-

This layer is a regular neural network that takes input from the previous layer.

LSTM - Long Short Term Memory :-

LSTM is a kind of RNN.

In RNN output from the last step is fed as i/p in the current step.

LSTM was designed by Hochreiter & schmidhuber.

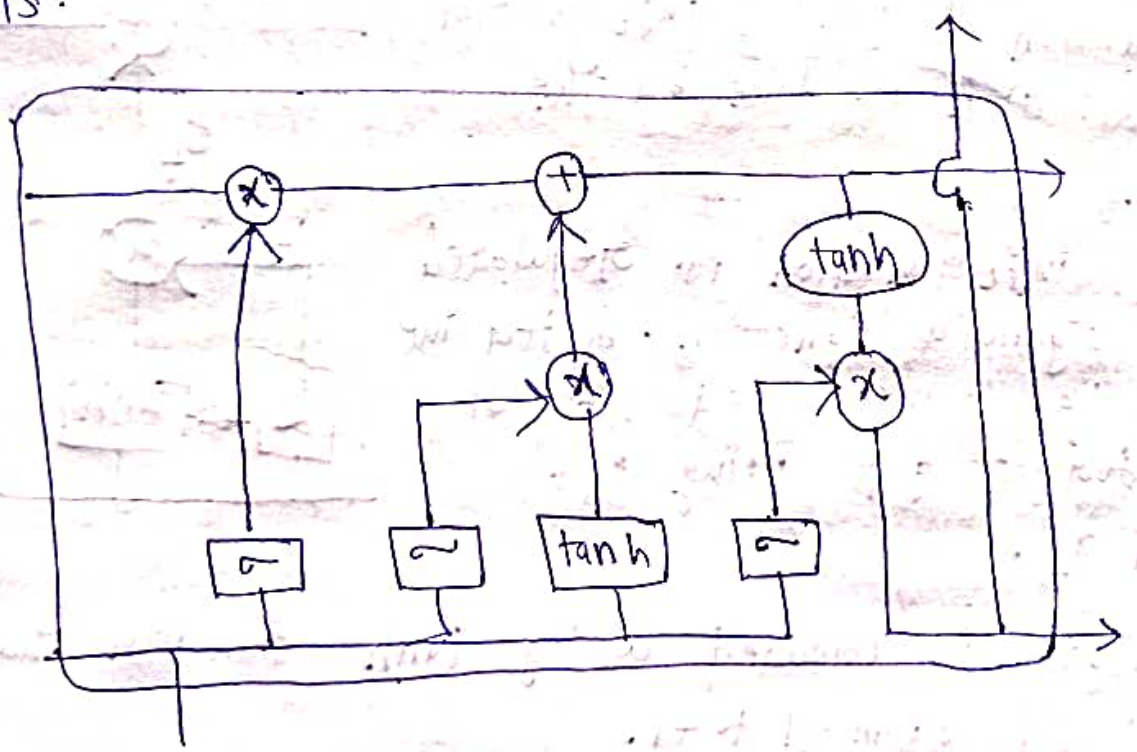
As the gap length increases RNN does not give an efficient performance.

LSTM can be default retain the information for a long period of time.

It is used for processing, predicting and classifying on the basis of time-series data.

Structure of LSTM :-

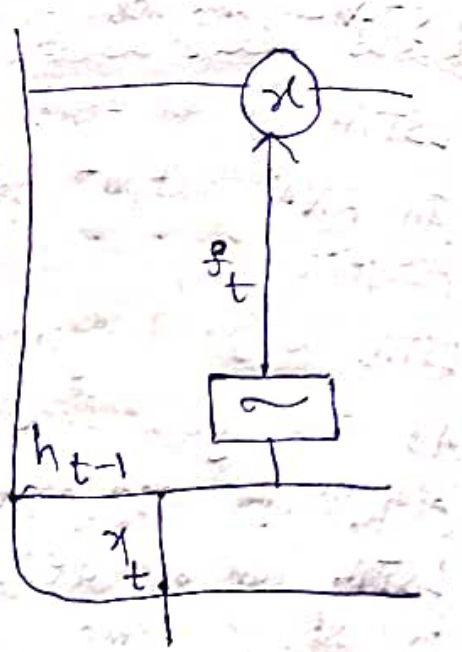
LSTM has a chain structure that contains four neural networks and different memory blocks called cells.



Information is retained by the cells and the memory manipulations are done by the gates.

There are three gates.

(i) Forget Gate:-



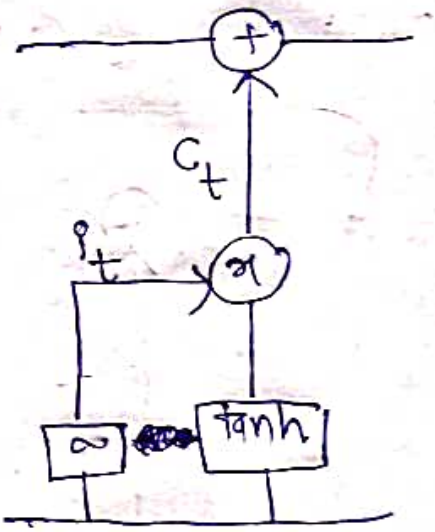
The information that is no longer useful in the cell state is removed with the forget gate.

Two inputs x_t and h_{t-1} are fed to the gate and multiplied with weight matrices followed by the addition of bias.

The resultant is passed through an activation function which gives the binary output.

If the o/p is '0', the information is forgotten and the o/p is '1', the information is retained for future use. x_t = i/p at particular time h_{t-1} = previous cell o/p

(ii) Input Gate:-



The addition of useful information to the cell state is done by the i/p gate.

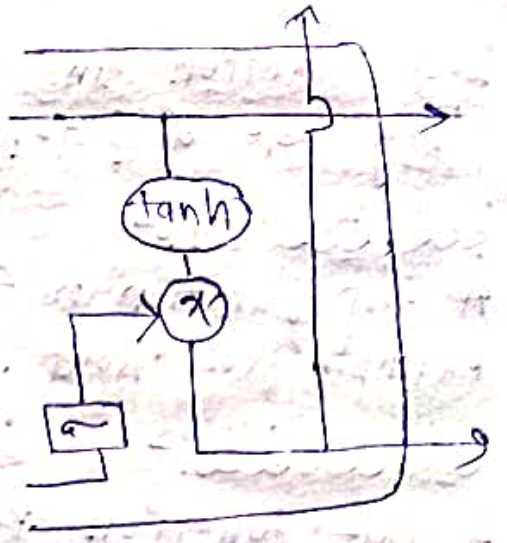
First the information is regulated using the sigmoid function, filter the values to be remembered similar to the forget gate using i/ps h_{t-1} and x_t.

Then a vector is created using tanh function that gives an o/p from −1 to +1.

It contains all possible values from h_{t-1} and x_t.

At last, the vector values and the regulated values are multiplied to obtain the useful information.

(iii) Output Gate:-



The task of the o/p Gate is extracting useful information from the current cell state as output.

First, a vector is generated by applying tanh function on the cell.

Then the information is regulated by the sigmoid function and filter the values using inputs h_{t-1} and x_t.

The vector values and the regulated values are multiplied to be sent as an o/p and i/p to the next cell.

Applications of LSTM:-

(i) Language Modelling.
(ii) Machine Translation
(iii) Image Captioning.
(iv) Hand writing generation.
(v) Question Answering Chatbots.

---

(7) Loss Functions:-

The loss function is very important in machine learning or deep learning.

The loss function is a method of evaluating, how well your algorithm is modeling your dataset.

It is a mathematical function of the parameters of the ML algorithm. It is also called error function or loss function.

If the value of the loss function is lower then it is a good model otherwise we have to change the parameters of the model and minimize the loss.

Loss Function and Cost functions are synonyms but they are different.

Loss Function:-
A loss function (or Error function) is for a single training example (or input).

Cost Function:-
A cost function is the average loss over the entire training dataset.

Loss Function in Deep Learning:-

(i) Regression
- → MSE (Mean Squared Error)
- → MAE (Mean Absolute Error)
- → Huber Loss

(ii) Classification
- → Binary cross-entropy
- → Categorical cross-entropy

(iii) Auto Encoders
- → KL. Divergence.

(iv) GAN
- → Discriminator loss
- → Minmax GAN loss.

(v) Object Detection
- → Focal Loss

(vi) Word embeddings
- → Triplet loss.

Regression Loss:-

(i) MSE / Squared Loss / L2 Loss:-

The MSE is the simplest and most common loss function. To calculate the MSE, you take the difference b/w the actual value and model prediction, square it and average it across the whole data set.

In regression at the last neuron use linear activation function.

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

(ii) MAE / L1 Loss:-

It is also a simplest loss function. To calculate the MAE, you take the difference b/w actual value and model prediction and avg it across the whole data set.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}|$$

(iii) Huber loss:-

In statistics, the Huber loss is a loss function used in robust regression, It is less sensitive to outliers in data than the squared error loss.

$$Huber = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(y_i - \bar{y})^2 \text{ , for } |y_i - \bar{y}| \le \delta$$

$$Huber = \frac{1}{n}\sum_{i=1}^{n}\delta\left(|y_i - \bar{y}| - \frac{1}{2}\delta\right) \text{ , for } |y_i - \bar{y}| > \delta$$

n = The number of data points.

y = The actual value (or True value.

$\bar{y}$ = The predicted value. $= \hat{y}_i$

δ = where the huber loss function transitions from a quadratic to linear.

Classification Loss:-

(i) Binary Cross Entropy / Log-loss:-

It is used in binary classification problems.

ex:- A person has covid or not.

$$\text{Log loss} = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \bar{y}_i + (1-y_i)\log(1-\bar{y})$$

y_i = Actual values

$\bar{y} = \hat{y}_i$ = Neural Network Prediction.

In classification at last neuron use sigmoid activation function.

(ii) Categorical Cross Entropy:-

It is used for multiclass classification.

It is also used in softmax regression.

$$\text{Loss} = -\sum_{j=1}^{k} y_j \log(\hat{y}_j)$$

where, k = the number of classes in the data.

$$\text{Cost} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[y_{ij}\log(\hat{y}_{ij})\right]$$

In multiclass classification at the last neuron use the softmax activation function.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

$$\text{Softmax activation} - f(z) = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}}$$

Hyper Parameter Tuning:-

Hyper parameters are the variables which determines the network structure and the variables determining how the network is trained.

Hyper parameters are set before training.

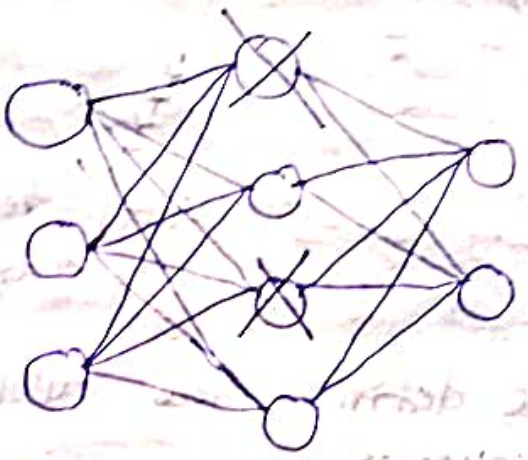
Hyper parameters are related to network structure.

In the network structure, the hidden layers are the layers b/w i/p and o/p layers.

Many hidden layers b/w the i/p layer and the o/p layer can increase the accuracy.

Small number of units may cause underfitting.

Dropout:-



Random neurons are cancelled.

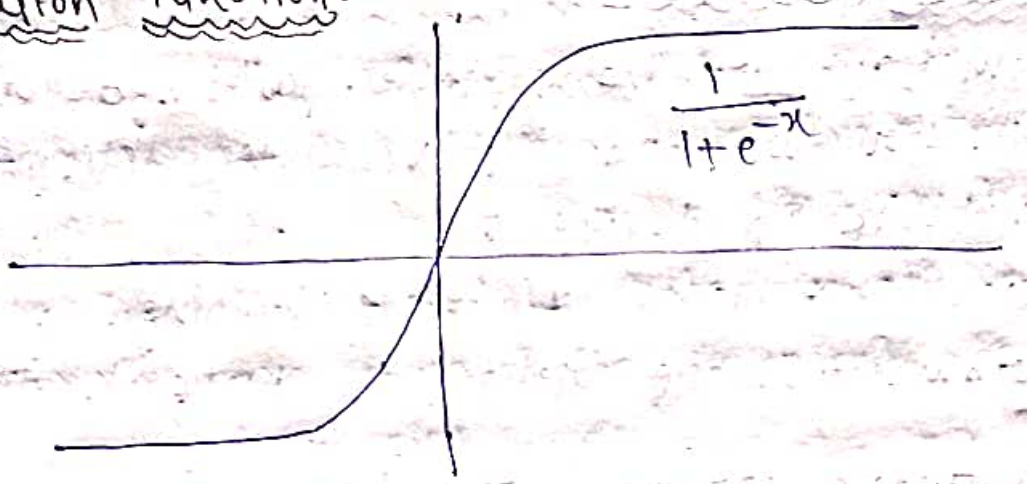
Dropout is a technique to avoid overfitting thus increasing the generalizing power and validation accuracy.

Network Weight Initialization:-

Mostly uniform distribution is used.

It may be better to use different weight initialization schemes a/c to the activation function used on each layer.

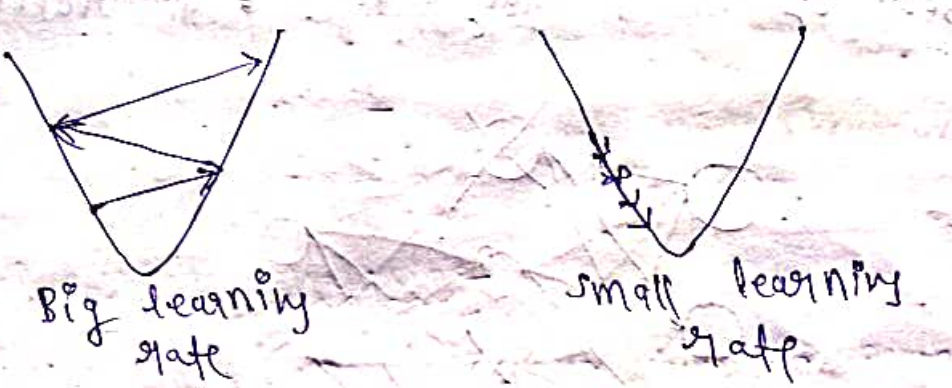
Activation Function:-



Hyper parameters are related to training algorithm.

Learning Rate:-

Gradient Descent



The learning rate defines how quickly a network updates its parameters.
Low learning rate slows down the learning process but converges smoothly.
Large learning rate speed up the learning process but may not converge.
Usually decaying or medium learning rate is preferred.

Momentum:-

Momentum helps to know the direction of the next step. It helps to prevent the oscillations.
The momentum is b/w 0.5 to 0.9 is better choice.

Number of epochs:-

Number of epochs is the number of times the whole training data is shown to the network while training. Increase the number of epochs until the validation accuracy starts decreasing even training accuracy is increasing (over fitting).

Batch size:-

Batch size is the number of sub samples given to the network.

A good for batch size might be 32, 64, 128, 256, ...

(9) Data Science Applications in Real Life:-

Data science often known as data-driven science, combining several aspects of statistics and computation to transform data into actionable information.

Data Mining, Statistics, ML, Data Analytics, and some programming are some of the technical disciplines that make up the data science field.

(i) Healthcare:-

Data Science is used in a variety of sectors in healthcare

(i) Image Analysis in Medicine.
(ii) Genetics and Genomics
(iii) Drug Development
(iv) Virtual Assistants and Health Bots.

⇒ Medical Image analysis:-

Procedures such as detecting tumors, organ delineation employ various different methods and frameworks like MapReduce to find optimal parameters for tasks like lung texture classification.

It applies ML methods like support vector Machine, content based medical image indexing, wavelet analysis for solid texture classification.

⇒ Genetics and Genomics:-

The goal is to understand the impact of the DNA on our health and find individual biological connections b/w genetics, diseases and drug response.

Data Science application enables an advanced level of treatment personalization through research in genetics and genomics.

⇒ Drug Development :-

Instead of Lab tests, ML algorithms can predict how the chemical will behave in the body using extensive mathematical modelling and simulations.

The goal of computational drug discovery is to construct computer model simulations in the form of a physiologically appropriate network.

⇒ Virtual Assistants and Health Bots :-

Basic healthcare help may be provided via AI powered smartphone apps, which are often chatbots.

You just explain your symptoms or ask questions and you'll get vital information about your health condition.

(ii) Targeted Advertising :-

If you thought, search is the most important data science use, consider this the whole digital marketing spectrum.

Data science algorithms are used to determine virtually ~~anywhere~~ display banners on various websites

This is the reason why digital ads have been able to get a lot higher CTR-Call Through Rate than the traditional advertisements. They can be targeted on a users past behaviour.

(iii) Website Recommendations :-

The recommendations are made based on previous search results for a user.

Many business companies have used this engine to promote their products in a/c with users interest and relevance of information.

This method is used by Internet companies such as Amazon, Twitter, Google Play, Netflix, LinkedIn, IMDb etc.

(iv) E-Commerce:-

The e-commerce sector greatly benefits from DS techniques and ML ideas such as NLP - Natural Language processing and recommendation systems.

These are used by e-commerce platforms to analyze consumer purchases and comments in order to gain valuable information for their company development.

(v) Transport:-

Data science establishes a foothold in transportation like study of fuel usage trends, driver behaviour and vehicle tracking. DS and ML algorithms are used in self-driving automobiles.

⇒ Airline Route Planning:-

Airlines may use data science to make strategic changes such as anticipating flight delays. Selecting which aircraft to buy, planning routes and developing marketing tactics.

Whether to directly land at the destination (or take a halt in b/w..

Southwest Airlines, Alaska Airlines are among top companies who've embraced data science to bring changes in their way of working.

(vi) Text and Advanced Image Recognition:-

Speech and picture recognition are ruled by data science algorithms.

In speech recognition technology, to comprehend and evaluate your words and delivering useful results from your use.

Image recognition may be found on Facebook, Instagram, and Twitter and other social media platforms.

You upload your image with friends on Facebook and you start getting suggestions to tag your friends. This automatic tag suggestion feature uses face recognition algorithm.

Google provides you with the option to search for images by uploading them. It uses image recognition and provides related search results.

(vii) Gaming:-

Games are now designed using ML algorithms that upgrade themselves as the player moves up to a higher level.

In motion gaming, your opponent (computer) also studies your past actions and adjusts its game.

EA sports, Zynga, Sony, Nintendo, and Activision-Blizzard have all used data science to take gaming to the next level.

(viii) Speech Recognition:-

Some of the best examples of speech recognition products are Google voice Assistant, Siri, Alexa, Cortana etc.

Using the speech recognition feature, if you aren't in a position to type a message, simply speak out the message it will be converted into text.

(ix) Security:-

Data science may be utilized to improve your company's security and secure critical data.
Banks uses these sophisticated ML algorithms to detect fraud detection.
Because of the massive amount of data created every day, these algorithms can detect fraud faster and more accurately than people.
Learning about data privacy may help you to avoid sharing sensitive information from consumers, such as credit card numbers, medical records, social security numbers and contact information.

(x) Customer Insights:-

Data on your clients may offer a lot of information about their behaviours, demographics, interests, aspirations and more.
Matching a customer's email address to their credit card information, social media handles and transaction identifications is one example of this.

(xi) Augmented Reality:-

Augmented Reality is a term that refers to one of the most exciting uses of technology.
Data science and Virtual Reality do have a relationship, considering a VR headset contains computing knowledge,

algorithms and data to provide you with the best viewing experience.

eg:- Pokemon Go.

Data science is used in marketing, finance and human resources, healthcare, government programmes and any other industry that generates data.