2 marks

1) How decision trees are making decision

A decision tree starts with a point (node) and starts splitting in two or more directions. Each branch offers different possible outcomes until final outcome is acheived

2) Infer of Ensemble of decision tree

Decision tree ensembles are commonly used in a wide range of applications becoming algorithms for decision tree based classifiers. Different trees in an ensemble can be processed in parallel during tree Inference

3) When do yo stop splitting decision tree.

Decision tree stops splitting when the number of records in the node to be split is less than the specified value

4) Interpret on kernel trick

kernel trick allows us to project data from a training set which isn't linearly seperable into higher dimensional space where it becomes linearly seperable

5) Entropy and Information gain

↳ Entropy measures impurity in data

↳ Information gain measures reduction in impurity in data

6) Why support vectors important

The use of support vector ensures that only a subset of data points determine the decision boundary

7) Leaf node

Leaf node are nodes of the tree that have no additional nodes coming off them.

8) Applications of clustering techniques

* Medical imaging
* Image segmentation
* Anomaly detection
* market segmentation
* Google photos

9) Demerits of k means algorithm

* Time complexity
* Sensitivity
* Cannot handle noisy data & outliers
* Defining number of clusters

10) Methods to merge clusters in Agglomerative clustering

* Single link
* Complete link
* Average link

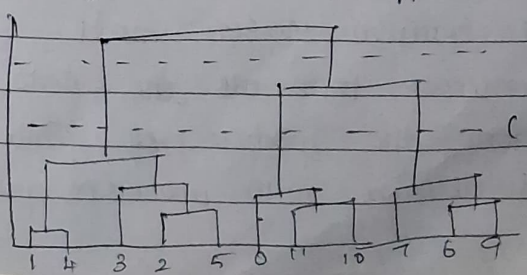
11) Parameters required to do clustering in DBSCAN algorithm

* Epsilon
* Midpoint

12)

| Collaborative filtering | Content based filtering |
| --- | --- |
| * Uses past interactions to recommend new items | * Uses ML algorithms to predict & recommand new items |
| * Item features are not required | * Are required |

13) Dendrogram for bottom-up approach



16) List the parameters used in DBSCAN algorithm

Qs: 11

15) Different scales of K-means clustering algorithm
- Standard Scale
- Robust scale
- Min Max Scale
- Normalizer

16) Challenges in unsupervised learning
- Lack of labelled data
- Determining no. of clusters
- Scaling and normalization
- Noise & outliers

17) Steps involved in K-means clustering algorithm
- Assigning each data point to the closet cluster centre
- Setting each cluster center to the mean of data points assigned to it

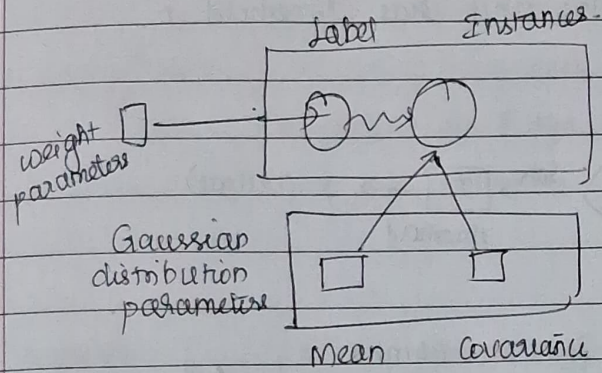
18) How choice of number of clusters affect K-mean result
- K-means is non-deterministic
- The choice of initial clusters have impact on final cluster formation

19) How DBSCAN classify datapoints into different groups
- eps: If the distance between two points are less than equal to eps, they are considered as neighbours
- minpoints: Minimum number of points to form a dense layer

20) Idea behind Gaussian Mixture model

It assumes that all the data points are generated from the mixture of finite no. of gaussian distributions with unknown parameters

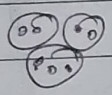
21) Architecture of Gaussian mixture model




22) Hard clustering

→ Every node belong to only one cluster

→

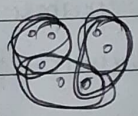


→ Eg: k means

Soft clustering

→ Every node may belong to several clusters

→

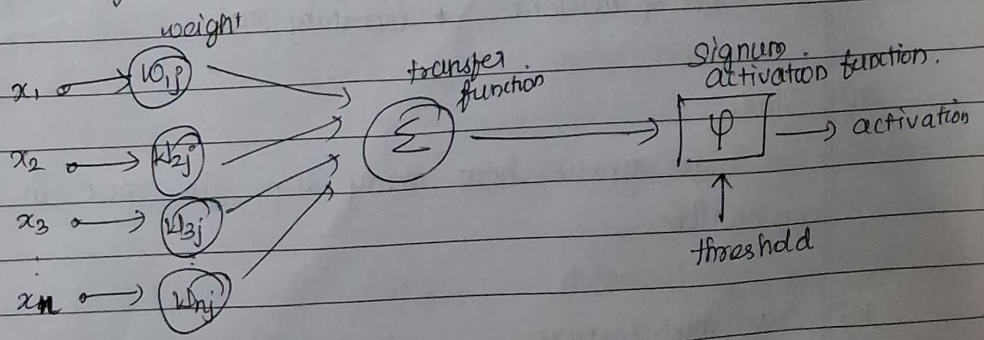


→ Eg: HMMs.

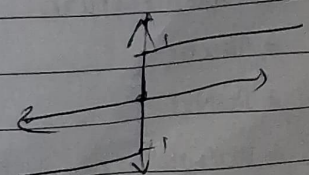
23) ANN.

ANN is a modeling technique inspired from human nervous system that allows learning by example from representative data

24) Signum activation function

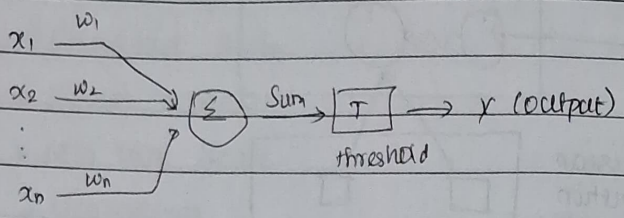



$$Sgn(x) = 1 \quad ; x > 0$$
$$-1 \quad ; x < 0$$
$$0 \quad ; x = 0.$$



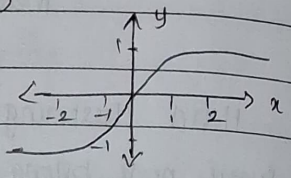
25) threshold logic unit.

It is a processing units with numbers for n inputs and one output. The unit has threshold θ.




26) Hyperbolic tangent function (tanh)

$$\tanh x = e^{x} - e^{-x} / e^{x} + e^{-x}$$




27) Significance of filters used in convolutional layer

The purpose of filters in convolutional layer is to help extract features from images

28) Zero padding

Adding zeros to the borders of input feature map

29) Softmax activation function

It scales numbers / logits into probabilities. mainly used in classification

30) Hyper parameters used for tuning neural network

* Epoch
* Batch size
* Activation function
* No. of neurons
* learning rate

31) Stride:

Stride denote how many steps we move in each step in convolution

32) CNN architectures

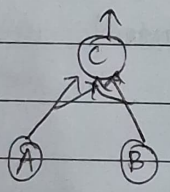
* Alexnet
* ResNet
* VGG net
* Google net

33)

| | Biological neuron | Artificial neuron |
|---|---|---|
| processing unit | Neuron | nodes |
| input | dendrites | Arcs |
| output | axons | Arcs |
| Speed | Slower | Faster |
| Complexity | more complex | less complex |
| Control unit | not present | present |

34) Activation functions
* ReLU
* Hyperbolic
* Sigmoid
* ELU

35) ANN to perform OR operation




C = A ∨ B

36) Feature of AlexNet
* It is 8 layers deep
* Uses ReLU activation function
* Overlapping pooling

37) Features of Google net
* It has 22 layers
* Uses ReLU Activation function
* Non-overlapping pooling.

2 marks

1) Compare tradional learning & ML

ML model learns by itself, it can handle new data. Traditional programming is more fixed approach where programmers design solutions, while ML is more flexible & learns from data to generate a solution

2)

| Online learning | Batch learning |
|---|---|
| * Learning is done Incrementally on dataset | * Learning is done once on dataset |
| * Complex to develop | * Less complex |
| * Adaptable | * not adaptable |
| * Expensive | * Less expensive |

3)

| Instance based learning | Model based learning |
|---|---|
| * Learn from examples and make predictions by finding similar examples | * Learn from underlying relationships and patterns in data creating mathematical representation |

4)

| Supervised learning | Unsupervised learning |
|---|---|
| * Input is provided along with o/p | * Only i/p is provided |
| * Uses training dataset | * Uses input dataset |
| * Used for prediction | * Used for analysis |
| * known no. of classes | * Unknown no. of classes |
| * Off-line analysis of data | * Real-time analysis of data |

5)

| Bias | Variance |
|---|---|
| * Training error of a model | * Difference between training and testing error of a model |

6) Machine learning

Machine learning is a subset of artificial intelligence that enables a machine or system to learn and improve from experience

7)

| Overfitting | Underfitting |
|---|---|
| → Too complex model | → Too simple model |
| → Accurate for training set | → No accurate for training set |
| → Need to reduce complexity | → need to increase complexity |
| → Not accurate for validation set | → Not accurate for validation set |

8) Batch algorithm:

In batch learning, data is accumulated over a period of time. The machine learning model is then trained with the data from time to time in batches

9) Online learning:

Qs. ⑥

10) Four challenges in ML

→ Insufficient Quantity of data

→ Poor Quality of data

→ Overfitting

→ Underfitting

11) Supervised, classification

12) For reliable machine learning model, you need to split dataset into training, validation & testing. Otherwise Result will be biased

13) Two supervised & unsupervised algorithms

↙ ↓

* Linear regression
* Logistic

* k-means
* Isolation forest

14) Steps in end-end ML project:

* Look at the big picture
* Get the data
* Discover and visualize data to get insights
* Prepare data for ML algorithms
* Select a model and train it
* Fine tune your model
* Present your solution to the test set
* Launch, monitor and maintain your system

15) A model should learn from data and able to predict the median housing price in any district given all metrices. Suggest type of learning

Ans: Supervised learning, Regression task

16)

| Probability sampling | Non probability sampling |
|---|---|
| → Probability selection of each individual is predetermined | → Probability selection of each individual is not known |
| → Types | → Types |
| 1) Simple random sampling | 1) Quota sampling |
| 2) Systematic " | 2) Network " |
| 3) Stratified " | 3) Judgemental " |
| 4) Cluster " | 4) Grab " |
| 5) Multistage " | |

17) Hypothesis

Hypothesis is a system's prediction function

18) MAE

→ Mean Absolute Error

→ $MAE = \frac{1}{n} \sum_{i=1}^{n} |h(x^{(i)}) - y^{(i)}|$

19) pipeline

A sequence of data processing components is called data pipeline

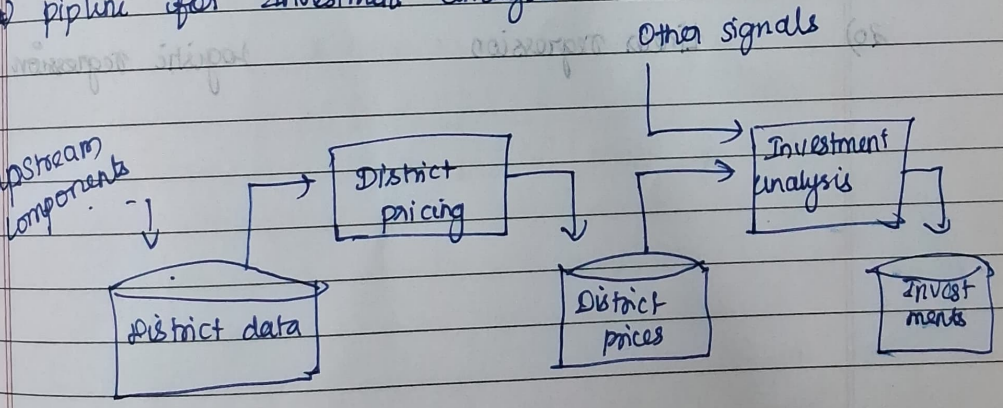
20) Ratio of training & testing dataset
→ Training data should be high for more accurate prediction
→ Eg: 80-20, 70-30

21) head() → used to get first n rows
info() → print information about dataframe

22) Open source dataset
* kaggle
* Amazon web services
* Google dataset search
* Visual Data

23) Essential libraries in python
* Pandas
* numpy
* Matplotlib
* Jupyter
* Scikit-learn

24) pipline for Investment analysis




25) Steps in framing process
* Determining whether ML is the right approach for solving problem
* Framing problem in ML terms

26) RMSE

Root mean square error $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(h(x(i)) - y(i)\right)^2}$

$n \rightarrow$ no. of instances

$x(i) \rightarrow$ feature values of $i^{th}$ instance

$y(i) \rightarrow$ label

$h \rightarrow$ hypothesis

27) Histogram

Histogram shows the no. of instances that have given value range

28) describe() $\rightarrow$ gives statistical summary of dataframe

29) Feature scaling methods

* min-max scaling
* Standardization

30) Linear regression Logistic regression