Machine Learning

Itis the field of study that gives computers the capability to learn from data by finding patterns and structures , and based on this findings it predicts new observations

---

Types of ML algo -:

1) **supervised ML** -:
-it uses labelled data to make prediction of new observation.
- Labelled Data -: is data where in features are mapped with op feature (feature -: cols)
works on 3 steps
i) remember previous data
ii) formulate that data by finding patterns / structure
ii) make prediction based on findings

types of Supervised ML algo
i) **Regression algo** -: when target col is cont. then use regression algo
example - : predict salary , price of cars , price of house
ii) **classification** -: when target col is discrete or categorical in nature then use classification algo
example -: covid y/n , load y/n

----
2) **unsupervised ML -:**
- it used unlabelled data to make predictions.
- unlabelled - in this data we dont know the target col.

types unsupervised ML algo -:
i) **clustering** -: creating groups/clusters by finding patterns from i/p data.
example - : customer segmentation

----

3) Reinforcement machine learning
This method allows machines and software agents to automatically determine the ideal behaviour within a specific context to maximize its performance.

example -: robotics , self driving car

Linear Regression algo-:

-It is Supervised ML algo used to predict cont. target varaible.
-it uses label data for prediction.
-It tries to establish relationship between X (indepedent variable) and Y (target) by finding out
best fit line.
-Line can be drawn with y=mx+c
here, y --> target
m--> slope --> how much y is changing wrt x (it indicate steepness of a line)
x--> indepedent var.
c--> intercept. (a location where it intersects an axis)

what is best fit line ?

-when a line is covering maximum data points from dataset or most of the points are close to a line,
is known as best fit line.
- in best fit line error rate is very low and accuracy is high.

what is error?
-> gap/ distance between actual data point and predicted data point.
-> this is also known as residual.
-> error rate must be low as possible.

how to find best fit line??

Gradient descent helps us to draw best fit line by calculating best values of m and c by taking
partial derivative at each step.

step1 -: machine will randomly select m and c value
step2 -: based on this m and c it will draw a line using y=mx+c
step3-: now it will calculate error rate. and tries to minimize it.
step4-: it will calculate new values of m and c by taking partial derivative of old m &c
step5-: using these new values it draws a line.
step6 -:cont this process till error rate becomes low.

Logistic Regression:-

-> It is a supervised ML algo used to solve classification problems.

-> It predicts outcomes which are categorical in nature.

-> Logistic regression uses sigmoid/ logistic function to classify a data point.

-> Logistic/sigmoid function always return probabilistic value that lies between 0 to 1

-> In logistic regression , instead of fitting best fit line , we fit "s" shaped curve, which predicts
two maximum values (0 or 1)

->Curve indicates likelihood of something.

->Sigmoid function maps any value into a range of 0-1

->logistic function uses threshold which help to classify a data point.

->value above threshold will be considered as 1, value less than threshold will be considered as 0

-> It is widely used to solve binary classification problems.

Assumption of Logistic Regression.

1)Target must be categorical in nature.

2)NO multi-collinearity

Advantages

-Performs well on linear data.

-Results are easily interpretable.

-It works well on large data sets

-faster training because of sigmoid function.

- works well on binary datasets

Disadvantages

- it does not perform well on non-linear data.
- It does work well on high dimensional data (Large features)-
- It makes assumptions on data.
- It does not work well on multiclassification datasets

Hyper Tuning Parameters

1) penalty -: it adds penalty term. possible values are -{l1,l2,elasticnet,none}

2) solver -: liblinear,sag,saga,lbfgs

3) multi_class -: auto,ovr,mltinomial

ROC -AUC curve (Receiver operating characteristics Area Under Curve)

-> it is a performance metrics for the classification problem at various threshold settings.

->It tell how much the model is capable of classifying between classes.

->Higher the AUC, the better the model is at predicting 0 class as 0 and 1 class as 1.

-> high value of AUC means model is good and vice versa

-> It is graph which we plot with TPR vs FPR. where TPR is on Y axis and FPR is on X axis.

Interview Questions:

1) How logistic regression works? (imp)
2) what is sigmoid function /importance of sigmoid function(imp)
3) importance of threshold in logistic regression. (imp)
4) Can logistic regression works with large data? --> Yes, it requires large data
5) Explain Drawbacks of logistic regression.
6) When you will like to use logistic regression (imp)
7) Explain ROC- AUC curve.
8) How will you improvise logistic regression performance / what are hypertuners of logistic regression (imp)
9) Does logsitic regression uses regularization by default? ---> Yes , l2 by default
10) explain solver in logistic regression(imp)
11) Advantages and disadvantages of logistic regression
12) logistic regression vs linear regression (imp)

SVM (Support vector Machine)
-> It is supervised ML algo which can be used to solve classification as well as regression problems.

Objective -:
-SVM is based on the idea of finding a hyperplane/ Decision line in an N-Dimensional space that best seperate the features into different domains.

Hyperplane-:

-Hyperplanes are decision boundaries that classify the data points into classes. Data points falling either side of the hyperplane can be classified to different classes.
-Dimension of hyperplane is depends on number of features. i.e if no of features are 2, then hyperplane is line. if no of features are 3 or more than 3 then it is known as 2d hyperplane

----
support vectors-:
Support vectors are data points that are closer to the hyper-plane and influence the position of the hyperplane.

support vectors plays imp role to draw decision line/hyperplane.

Margin-: The distance of vectors from the hyperplane are called margins.
-distance from boundary line to decision line.

Best hyperplane ----> hyperplane with High margin is considered as best hyperplane.
----
kernel -: Kernel is used to handle non-linear dataset as we can not draw best decision line in non linear data.
kernel will add extra dimension to handle non-linear data by finding out best hyperplane in higher dimension space.

Advantages of SVM-:
1) it can handle linear as well non linear data. -: it handle linear data by finding a best decison line and it handles non linear data by using kernel trick.

2) it can be used to solve classification as well as regression problems.

3) stability -: A small change to the data does not affect the hyperplane.

Disadvantages

1) choosing a correct kernel type.
2) extensive memroy requirement - > High complex algo , high vol. of computation requires.
3) Long trining time on large non linear data
4) it requires Feature scaling
5) difficult to intrpret results of SVM

Decision Tree:

-it is a supervised ML algo that uses label data to classify a data point.
-It can be used to solve regression as well as classification problem.
-It is a graphical representation for getting all the possible solutions to problem/ decision based on given condition.
-It uses different nodes such as Root node, branch/decision node and leaf node.
-It tree like structured classifier , where internal nodes represent the features of a data set , branches represent the decision rules each leaf node represent the outcome.

On which basis DT select feature for further splitting?

sol 1) On the basis of impurity. DT select a feature with low impurity.
sol 2) INformation Gain.

How to calculate impurity?
1) Gini index - 1-p2-q2
where p is a prob of even will occure (like the movie) and q is the prob of event will not occure (not like the movie)

2) Entropy

Advantages of DT
-> Results of DT are easy to interpret.
-> DT are not affected by noisy data.
-> It can handle non linear data also.
-> IT can solve regression as well as classification problem.

Disadvantages of DT

->It is not suitable for large and high dimension datasets.
-> It is not flexible as it might lead to reconstruct DT.
->it always overfits. (IMP)

How to solve overfitting problem of DT?
--> use pruning techniques
1) max_depth -: The maximum depth of the tree. If None, then nodes are expanded until
    all leaves are pure. Default value is None.
2) min_sample_leaf-: The minimum number of samples required to be at a leaf node Default value -: 1
3) min_sample_split -: The minimum number of samples required to split an internal node Default -: 2

Unsupervised ML:

Clustering:

-CLustering is a unsupervised learning process of creating groups of data points based on similarity.
-HEre we dont have target column. we look at the data and then try to club similar observation and form different groups.

Application of clustering/ where to apply clustering>?

-customer segmentation.
-recommendation system.

How to perform clustering?
- We have two algorithms to perform clustering
1) K-Means clustering
2) Hierarchical clustering.

How K-Means works?
Here K is -: no of groups/clusters to make.

1) Decide the value of K.
(To decide the value of K we must have Domain knowledge).

2) Select K centroids
(Centroids can be selected randomly or can be selected from datapoints.)

3) By calculating the Euclidean distance assign the datapoint to the nearest centroids/cluster.Now again find the new centroid for that cluster and keep doing this process for inner iteration times (default value is 300).
and then calculate inertia.

4) Now again re-generate centroids and go to step no 3. Keep doing this process for Outer iteration times. (default value:- 10)

5) Final centroids/clusters are selected whose inertia value is low.

How Good clusters/final clusters are selected?
(refer whitboard)

How to select number of cluster to make?
1)You must domain knowledge
2)Use Elbow technique/ Method (refer whiteboard.)

Interview questions
-What is clustering?
-Why to use clustering / Application clustering?
-What is K in K-means
-Difference between Kmeans and KNN algo.
-How Kmeans works?
-How best clusters are selected
-what is inertia and importance of it
-How to select the best value fo K?