

ML interview question:

- 1.What is Machine Learning ? Explain different types of Machine Learning.
- 2.Explain linear regression algorithm in detail.
- 3.Explain overfitting and underfitting .Also explain the concept of a good model.
- 4.Explain logistic regression algorithm in detail.
- 5.Explain naïve bayesian classifier algorithm in detail.
- 6.Explain decision tree algorithm in detail.
- 7.Explain K nearest neighbor algorithm in detail.
- 8.Explain principal component analysis algorithm in detail.

Name :- Rushikesh Jadhav
Roll No. : 3G159
Sub :- ML Assignment.

Q.7] What is Machine Learning? Explain different types of Machine learning.

→ i) Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data. It is a branch of artificial intelligence based on the idea that systems can learn from data.

Machine learning is a method of data analysis that automates analytical model building. Identify patterns and make decisions with minimal human intervention.

ii) There are ~~four~~ different types of machine learning:

- i) Supervised learning
- ii) Unsupervised learning
- iii) Reinforcement learning.
- iv) Semi-supervised learning.

v) Supervised learning :- supervised learning is the machine learning task of learning a function. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.

Some popular examples of machine learning

- i) Linear regression.
- ii) Random forest
- iii) Support Vector Machines.

ii) Unsupervised learning :- unsupervised learning is a type of machine learning in which the algorithm is not provided with any pre-assigned labels or scores for the training data.

Some examples of unsupervised learning.

- i) k-Means clustering
- ii) Principal Component Analysis.

iii) semi-supervised learning :- Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a unlabeled data. semi-supervised learning falls between unsupervised learning and supervised learning.

Some application of semi-supervised learning.

- i) text document classifier.
- ii) speech analysis.

iv) reinforcement learning :- Reinforcement learning is about the learning optimal behaviour in an environment to obtain maximum reward. It takes action in order to maximize the notion of cumulative reward.

a)

Explain linear regression algorithm in detail.

→ Linear Regression is an ML algorithm used for supervised learning. Linear regression performs the task to predict to a dependent variable (target) based on the given independent variable. So this regression technique finds out a linear relationship b/w dependent variable and other given independent variables.

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent (x) and dependent (y) variable. The linear regression model gives a sloped straight line describing the relationship within the variables.

a)

Explain overfitting and underfitting. Also explain the concept of a good model.

→

Bias :- Assumptions made by a model to make function easier to learn.

Variance :- If you train your data on training data and obtain a very low error, upon changing the data and then training the same previous model you experience a high error, this is variance.

→ Overfitting :- A statistical model is said to be overfitted when we train it with a lot of data when a model gets trained with so much data, it starts learning from the noise and inaccurate data.

entries in our data set. In the overfitting High Variance and low bias are present.

Underfitting :- A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data. Its occurrence simply means that our model or the algorithm does not fit the data well enough. In the underfitting High bias and low variance are present.

Good model :- A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data. In good models contain one accuracy or interpretability of the output. and contains linearity.

Q) Explain linear regression algorithm in detail.

→ Linear Regression Algorithm is a machine learning algorithm based on supervised learning. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable. So, this regression technique finds out

a linear relationship between x and y (output).
Hence, the name is Linear Regression.

- Q) Explain naive bayesian classifier algorithm in detail.

→ The Naive Bayes classification algorithm is a probabilistic classifier. It is based on probability models. You can derive probability models by using Bayes' theorem.

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.

$P(A)$ is the priori of A. The evidence is an attribute value of an unknown instance.

$P(A|B)$ is a posteriori probability of A, i.e. probability of event after decision is seen.

Now, we apply Bayes' theorem in following way:

$$P(y|x) = \frac{P(x|y) P(y)}{P(x)}$$

Q) Explain decision tree algorithm in detail.

→ Decision tree is a supervised learning technique that can be used for both classification and Regression problems. but mostly it is preferred for solving classification problems. It is a tree-structured classifier, where internal nodes represent the feature of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a decision tree, there are two nodes, which are the decision node and leaf node. Decision nodes are used to make any decision and have multiple branches, whereas leaf nodes are the output of those decisions and do not contain any further branches.

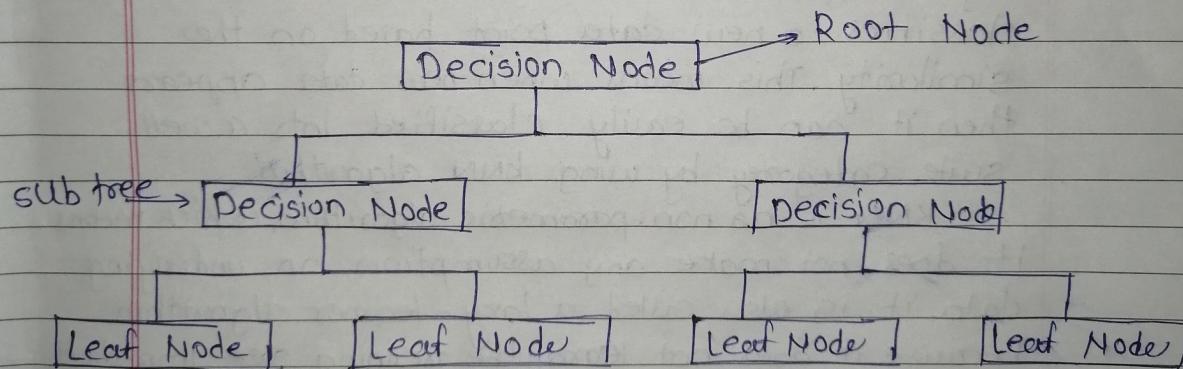
Decision trees usually mimic human thinking ability while making a decision, so it's easy to understand. The logic behind the decision tree can be easily understood because it shows a tree-like structure.

Decision - Tree Terminologies :-

- Root Node :- Root Node is from where the decision tree starts, it represents the entire dataset.

- Leaf Node :- Leaf nodes are the final output nodes, and tree cannot be segregated further after getting a leaf node.

- splitting :- splitting is the process of dividing the decision node/root node into sub-nodes according to given conditions.
- Branch/sub Tree :- A tree formed by splitting the tree.
- pruning :- pruning is the process of removing the unwanted branches from the tree.
- parent/child node :- The root node of the tree is called the parent node, and other nodes are called the child nodes.



Q) Explain k-nearest neighbor algorithm in detail
→

K-nearest neighbors (KNN) algorithm is a type of supervised machine learning algorithm which can be used for both classification as well as regression problems. However, it is mainly used for classification problems. However, in the industry.

KNN is ~~assumes~~ assumes the similarity between the new case data and available cases and put the new case into the category that is most similar to the available categories. KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using KNN algorithm.

KNN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy-learner algorithm because it does not learn the training set immediately. It stores the dataset and at the time of classification, it performs action on the dataset.

Q) Explain principal component analysis algorithm in detail.

→ Large datasets are increasingly common and are often difficult to interpret. Principal components analysis (PCA) is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It can also be used as an exploratory tool for data analysis.

Principal component analysis (PCA) is a technique for reducing the dimensionality of such datasets.

Steps to perform PCA are as following:

- 1) Standardize the data.
- 2) Compute the covariance matrix of the features from the dataset.
- 3) Perform eigenvalue decomposition on the covariance matrix.
- 4) Order the eigenvectors in decreasing order based on the magnitude of their corresponding eigenvalues.
- 5) Determine k , the number of top principal components to select.
- 6) Construct the projection matrix from the chosen number of top principal components.
- 7) Compute the new k -dimensional feature space.