General VIVA Questions

1. What is machine learning?

The capability of a machine to imitate intelligent human behavior.

2. Define supervised, unsupervised and reinforcement learning?

a. Supervised learning requires a labeled dataset for training.
b. Unsupervised learning identifies hidden data patterns from an unlabeled dataset,
c. Reinforcement learning does not require data as it learns by interacting with the environment.

3. What are the examples of Supervised and Unsupervised Learning?

Unsupervised Learning areas of application include market basket analysis, semantic clustering, recommender systems, etc.

The most commonly used Supervised Learning algorithms are decision tree, logistic regression, linear regression, support vector machine.

4. What do you mean by hypothesis?

It is defined as the approximate function that best describes the target in supervised machine learning algorithms.

5. What is classification?

In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data.

6. What is clustering?

Grouping unlabeled examples is called clustering.

7. Define precision, accuracy and recall.

Accuracy tells you how many times the ML model was correct overall. Precision is how good the model is at predicting a specific category. Recall tells you how many times the model was able to detect a specific category.

8. Define Entropy.

Entropy, as it relates to machine learning, is a measure of the randomness in the information being processed.

9. Define regression.

Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables.

10. How KNN is different from K-means and Clustering?

KNN is a supervised machine learning algorithm used for classification, whereas KMeans is an unsupervised machine learning algorithm used for clustering.

11. What is concept learning?

Concept learning describes the process by which experience allows us to partition objects in the world into classes for the purpose of generalization, discrimination, and inference.

12. Define specific and general boundary.

The general boundary, with respect to hypothesis space and training data, is the set of maximally general members of consistent with. The specific boundary, with respect to hypothesis space and training data, is the set of minimally general (i.e., maximally specific) members of consistent with.

13. Define target function

A target function, in machine learning, is a method for solving a problem that an AI algorithm parses its training data to find. Once an algorithm finds its target function, that function can be used to predict results (predictive analysis).

14. Define Decision tree?

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure.

15. What is ANN

Artificial Neural Network. Its a network made up of neurons which is artificial. Just kidding, Artificial Neural Networks (ANN) are algorithms based on brain function and are used to model complicated patterns and forecast issues.

16. Explain gradient descent approximation

Gradient descent is an algorithm that numerically estimates where a function outputs its lowest values.

17. State Bayes Theorem

Bayes' theorem named after Thomas Bayes, describes the probability of an event, based on prior knowledge of conditions.

18. Define Bayesian Belief Networks

Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty.

19. Differentiate between hard and soft clustering

Hard clustering is method to grouping the data items such that each item is only assigned to one cluster, K-Means is one of them. While Soft clustering is method to grouping the data items such that an item can exist in multiple clusters, Fuzzy C-Means (FCM) is an example.

20. Define Variance

The variance is a measure of variability. It is calculated by taking the average of squared deviations from the mean.

21. What is inductive machine learning.

Inductive Learning Algorithm (ILA) is an iterative and inductive machine learning algorithm which is used for generating a set of a classification rule, which produces rules of the form “IF-THEN”, for a set of examples, producing rules at each iteration and appending to the set of rules.

22. Why KNN is lazy learning?

Because it does no training at all when you supply the training data.

23. Why is naive bayes naive?

Naive Bayes is called naive because it assumes that each input variable is independent.

24. Mention classification algorithms

i. Logistic Regression.
ii. Naive Bayes.
iii. K-Nearest Neighbors.
iv. Decision Tree.
v. Support Vector Machines.

25. Define pruning.

Pruning is the process of removing weight connections in a network to increase inference speed and decrease model storage size.

26. Differentiate clustering and classification

Classification is used for supervised learning whereas clustering is used for unsupervised learning.

27. Mention clustering algorithms

i. Density-based. ...
ii. Distribution-based. ...
iii. Centroid-based. ...
iv. Hierarchical-based. ...
v. K-means clustering algorithm.

28. Define Bias

Bias is a phenomenon that skews the result of an algorithm in favor or against an idea.