**VIVA Questions**

**1)What is machine learning?**

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead.

**2)Define supervised learning**

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.

**3)Define unsupervised learning**

Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.

**4)Define semi supervised learning**

Semi-supervised learning is a class of machine learning tasks and techniques that also make use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data.

**5)Define reinforcement learning**

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

**6)What do you mean by hypotheses?**

Hypothesis is a candidate model that approximates a target function for mapping inputs to outputs.

**7)What is classification?**

A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.

**8)What is clustering?**

A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

**9)What is Association?**

**Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

**10)Define precision, accuracy and recall**

Precision is the fraction of relevant instances among the retrieved instances, Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition: Accuracy = Number of correct predictions Total number of predictions. Recall is the fraction of the total amount of relevant instances that were actually retrieved.

**11)Define entropy**

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

**12)Define regression**

A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

**13)How K-nn is different from k-means clustering**

K-Nearest Neighbors is a supervised classification algorithm, while k-means clustering is an unsupervised clustering algorithm.

While the mechanisms may seem similar at first, what this really means is that in order for K-Nearest Neighbors to work, you need labeled data you want to classify an unlabeled point into (thus the nearest neighbor part).

K-means clustering requires only a set of unlabeled points and a threshold: the algorithm will take unlabeled points and gradually learn how to cluster them into groups by computing the mean of the distance between different points.

**14)What is concept learning?**

The problem of searching through a predefined space of potential hypotheses for the hypothesis that best fits the training examples.

**15)Define specific boundary and general boundary**

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

**16)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.

**17)Define decision tree**

Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves).

**18)Define Find-S Algorithm**

FIND-S algorithm finds the most specific hypothesis within H that is consistent with the positive training examples. – The final hypothesis will also be consistent with negative examples if the correct target concept is in H, and the training examples are correct

**19)What is ANN?**

Artificial Neural networks (ANN) or neural networks are computational algorithms. It intended to simulate the behavior of biological systems composed of “neurons”. ... A neural network is a machine learning algorithm based on the model of a human neuron.

**20)Explain gradient descent approximation**

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model.

**21)State Bayes theorem**

Bayes' Theorem is the fundamental result of probability theory – it puts the posterior probability P(H|D) of a hypothesis as a product of the probability of the data given the hypothesis(P(D|H)), multiplied by the probability of the hypothesis (P(H)), divided by the probability of seeing the data.

**22)Define Bayesian belief networks**

A Bayesian network, Bayes network, belief network, decision network, ..... In other applications the task of defining the network is too complex for humans. ... (BN) is a challenge pursued within machine learning.

**23)Differentiate hard and soft clustering**

Hard Clustering: In hard clustering, each data point either belongs to a cluster completely or not. ... Soft Clustering: In soft clustering, instead of putting each data point into a separate cluster, a probability or likelihood of that data point to be in those clusters is assigned.

**24)Define variance**

Variance, in the context of Machine Learning, is a type of error that occurs due to a model's sensitivity to small fluctuations in the training set. High variance would cause an algorithm to model the noise in the training set. This is most commonly referred to as overfitting

**25)What is inductive machine learning?**

Inductive Learning. This involves the process of learning by example -- where a system tries to induce a general rule from a set of observed instances. This involves classification -- assigning, to a particular input, the name of a class to which it belongs. Classification is important to many problem solving tasks.

**26)Why K nearest neighbor algorithm is lazy learning algorithm?**

K-NN is a lazy learner because it doesn't learn a discriminative function from the training data but “memorizes” the training dataset instead. For example, the logistic regression algorithm learns its model weights (parameters) during training time.

Why naïve Bayes is naïve?

A naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. Basically, it's "naive" because it makes assumptions that may or may not turn out to be correct.

**27)Mention classification algorithms**

* Linear Classifiers: Logistic Regression, Naive Bayes Classifier
* Nearest Neighbor
* Support Vector Machines
* Decision Trees
* Boosted Trees
* Random Forest
* Neural Networks

**28)Define pruning**

Pruning is a technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances

**29)Mention clustering algorithms**

### K-Means

### EMC

### Affinity Propagation

**30)Define Bias**

Bias is a phenomenon that occurs when an algorithm produces results that are systematically prejudiced due to erroneous assumptions in the machine learning process.

**31)What is learning rate? Why it is need.**

The learning rate controls how quickly the model is adapted to the problem. Smaller learning rates require more training epochs given the smaller changes made to the weights each update, whereas larger learning rates result in rapid changes and require fewer training epochs.