1. What is the concept of supervised learning? What is the significance of the name?

Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Basically, supervised learning is when we teach or train the machine using data that is well labelled. Which means some data is already tagged with the correct answer.

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher.

2. In the hospital sector, offer an example of supervised learning.

Disease Identification and Diagnosis

Medical Imaging Diagnosis

Robotic Surgery

3. Give three supervised learning examples.

House prices Prediction

Is it a cat or a dog?

How’s the weather today?

Who are the unhappy customers?

4. In supervised learning, what are classification and regression?

Regression and Classification algorithms are Supervised Learning algorithms. Both the algorithms are used for prediction in Machine learning and work with the labelled datasets.

Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.

5. Give some popular classification algorithms as examples.

Logistic Regression.

Naive Bayes.

K-Nearest Neighbors.

Decision Tree.

Support Vector Machines.

6. Briefly describe the SVM model.

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

7. In SVM, what is the cost of misclassification?

8. In the SVM model, define Support Vectors.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

9. In the SVM model, define the kernel.

Kernels or kernel methods (also called Kernel functions) are sets of different types of algorithms that are being used for pattern analysis. They are used to solve a non-linear problem by using a linear classifier.

Kernel Function is a method used to take data as input and transform into the required form of processing data. “Kernel” is used due to set of mathematical functions used in Support Vector Machine provides the window to manipulate the data.

10. What are the factors that influence SVM's effectiveness?

11. What are the benefits of using the SVM model?

SVM works relatively well when there is a clear margin of separation between classes. SVM is more effective in high dimensional spaces. SVM is effective in cases where the number of dimensions is greater than the number of samples. SVM is relatively memory efficient.

12. What are the drawbacks of using the SVM model?

SVM algorithm is not suitable for large data sets. SVM does not perform very well when the data set has more noise i.e. target classes are overlapping. In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.

13. Notes should be written on

1. The kNN algorithm has a validation flaw.

2. In the kNN algorithm, the k value is chosen.

Important thing to note in k-NN algorithm is the that the number of **features** and the number of **classes** both **don't play a part in determining the value of k** in k-NN algorithm. k-NN algorithm is an ad-hoc classifier used to classify test data based on distance metric, i.e a test sample is classified as Class-1 if there are more number of Class-1 training samples closer to the test sample compared to other Classes training samples. For Eg: If value of k = 5 samples, then the 5 closest training samples are selected based on a distance metric and then a voting for most number of samples per class is done. So if 3 samples belong to Class-1 and 2 belong to Class-5, then that test sample is classified as Class-1. So the value of **k** indicates the number of training samples that are needed to classify the test sample.

Coming to your question, the value of **k** is non-parametric and a general rule of thumb in choosing the value of **k** is **k = sqrt(N)/2**, where **N** stands for the **number of samples in your training dataset**. Another tip that I suggest is to try and keep the value of k odd, so that there is no tie between choosing a class but that points to the fact that training data is highly correlated between classes and using a simple classification algorithm such as k-NN would result in poor classification performanc

3. A decision tree with inductive bias

The inductive bias (also known as learning bias) of a learning algorithm is **the set of assumptions that the learner uses to predict outputs of given inputs that it has not encountered**. In machine learning, one aims to construct algorithms that are able to learn to predict a certain target output.

**Shorter trees are preferred over longer trees**. Trees that place high information gain attributes close to the root are preferred over those that do not.

14. What are some of the benefits of the kNN algorithm?

**Some Advantages of KNN**

* Quick calculation time.
* Simple algorithm – to interpret.
* Versatile – useful for regression and classification.
* High accuracy – you do not need to compare with better-supervised learning model

15. What are some of the kNN algorithm's drawbacks?

**Some Disadvantages of KNN**

* Accuracy depends on the quality of the data.
* With large data, the prediction stage might be slow.
* Sensitive to the scale of the data and irrelevant features.
* Require high memory – need to store all of the training data.
* Given that it stores all of the training, it can be computationally expensive.

16. Explain the decision tree algorithm in a few words.

A decision tree is **a graphical representation of all the possible solutions to a decision based on certain conditions**. Tree models where the target variable can take a finite set of values are called classification trees and target variable can take continuous values (numbers) are called regression trees.

A decision tree is simply a set of cascading questions. When you get a data point (i.e. set of features and values), you use each attribute (i.e. a value of a given feature of the data point) to answer a question. The answer to each question decides the next question.

Decision Tree algorithm belongs to the family of supervised learning algorithms. The goal of using a Decision Tree is **to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data**(training data).

17. What is the difference between a node and a leaf in a decision tree?

A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

18. What is a decision tree's entropy?

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample.

Entropy, as it relates to machine learning, is **a measure of the randomness in the information being processed**. The higher the entropy, the harder it is to draw any conclusions from that information. Flipping a coin is an example of an action that provides information that is random.

19. In a decision tree, define knowledge gain.

Information gain is the reduction in entropy or surprise by transforming a dataset and is often used in training decision trees. Information gain is calculated by comparing the entropy of the dataset before and after a transformation.

Information gain **helps to determine the order of attributes in the nodes of a decision tree**. The main node is referred to as the parent node, whereas sub-nodes are known as child nodes. We can use information gain to determine how good the splitting of nodes in a decision tree.

20. Choose three advantages of the decision tree approach and write them down.

* Easy to read and interpret. One of the advantages of decision trees is that their outputs are easy to read and interpret without requiring statistical knowledge. ...
* Easy to prepare. ...
* Less data cleaning required.

21. Make a list of three flaws in the decision tree process.

* Overfitting the data
* Guarding against bad attribute choices
* Handling continuous valued attributes
* Handling missing attribute values
* Handling attributes with differing costs

22. Briefly describe the random forest model.

Random forest is **a supervised learning algorithm**. The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

A random forest is **a machine learning technique that's used to solve regression and classification problems**. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees.