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

Answer:-->Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.Therefore, the goal of supervised learning is to learn a function that, given a sample of data and desired outputs, best approximates the relationship between input and output observable in the data.In Supervised learning, you train the machine using data which is well "labeled." Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Supervised learning allows you to collect data or produce a data output from the previous experience.Supervised machine learning algorithms allow for the discovery of insights to better understand relationships and patterns within a labeled training data set. A labeled training data set already contains the known value, or answer, for the target variable of each record.

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

Answer:-->The Healthcare industry is an essential industry that offers care to millions of citizens, while at the same time, contributing to the local economy. Artificial Intelligence is benefiting the healthcare industry in numerous ways. Information technology is revolutionizing the healthcare industry by providing a helping hand. As we know, AI is the development of computing systems that are capable of performing tasks that typically require human intelligence. These involve heavy tasks like decision making, solving complex problems, object detection, and much more. The benefits of technologies, such as, increased level of accuracy and high level of computation, which take days for humans to solve manually are getting implemented into the healthcare business to enhance the services and keep the data organized.Machine learning is an application of AI which has impacted various domains including marketing, finance, the gaming industry, and even the musical arts. However, the largest impact of Artificial intelligence is on the field of the healthcare industry.

3. Give three supervised learning examples.

Answer:-->The most commonly used Supervised Learning algorithms are decision tree, logistic regression, linear regression, support vector machine.Supervised learning is when the model is getting trained on a labelled dataset. A labelled dataset is one that has both input and output parameters

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

Answer-->The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc. and Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.Classification tries to find the decision boundary, which divides the dataset into different classes. Regression algorithms solve regression problems such as house price predictions and weather predictions

5. Give some popular classification algorithms as examples.

Answer:-->Popular Classification Algorithms: Logistic Regression. Naive Bayes. K-Nearest Neighbors. Decision Tree. Support Vector Machines. Naive Bayes classifier algorithm gives the best type of results as desired compared to other algorithms like classification algorithms like Logistic Regression, Tree-Based Algorithms, Support Vector Machines. Hence it is preferred in applications like spam filters and sentiment analysis that involves text.

6. Briefly describe the SVM model.

Answer:-->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.Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of pointsSupport Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points

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

Answer:-->Hard margin SVM does not allow any misclassification to happen. In case our data is non-separable/ nonlinear then the Hard margin SVM will not return any hyperplane as it will not be able to separate the data.Two types of misclassifications are tolerated by SVM under soft margin: The dot is on the wrong side of the decision boundary but on the correct side/ on the margin (shown in left) The dot is on the wrong side of the decision boundary and on the wrong side of the margin (shown in right)Misclassification occurs when individuals are assigned to a different category than the one they should be in. This can lead to incorrect associations being observed between the assigned categories and the outcomes of interest.

8. In the SVM model, define Support Vectors.

Answer:-->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.Support Vectors: These are the points that are closest to the hyperplane. A separating line will be defined with the help of these data points. Margin: it is the distance between the hyperplane and the observations closest to the hyperplane (support vectors). In SVM large margin is considered a good margin

9. In the SVM model, define the kernel.

Answer:__:"Kernel" is used due to a set of mathematical functions used in Support Vector Machine providing the window to manipulate the data. So, Kernel Function generally transforms the training set of data so that a non-linear decision surface is able to transform to a linear equation in a higher number of dimension spaces.Kernelscale is literally a scaling parameter for the input data. The input data is recommended to be scaled with respect to a feature before being applied to the Kernel function. When the absolute values of some features range widely or can be large, their inner product can be dominant in the Kernel calculation.

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

Answer:-->The effectiveness of SVM depends on the selection of kernel, kernel's parameters and soft margin parameter C. . Each pair of parameters is checked using cross validation, and the parameters with best cross validation accuracy are picked.A hyperplane is a decision boundary that differentiates the two classes in SVM. A data point falling on either side of the hyperplane can be attributed to different classes. The dimension of the hyperplane depends on the number of input features in the dataset. Add more data. Having more data is always a good idea. ... Treat missing and Outlier values. ... Feature Engineering. ... Feature Selection. ... Multiple algorithms. ... Algorithm Tuning. ... Ensemble methods.

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

Answer:-->After giving an SVM model sets of labeled training data for each category, they're able to categorize new text. Compared to newer algorithms like neural networks, they have two main advantages: higher speed and better performance with a limited number of samples (in the thousands).The advantages of SVM and support vector regression include that they can be used to avoid the difficulties of using linear functions in the high-dimensional feature space, and the optimization problem is transformed into dual convex quadratic programs.SVM's are very good when we have no idea on the data. Works well with even unstructured and semi structured data like text, Images and trees. The kernel trick is real strength of SVM. With an appropriate kernel function, we can solve any complex problem.

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

Answer:--> It does not execute very well when the data set has more sound i.e. target classes are overlapping. In cases where the number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.It does not execute very well when the data set has more sound i.e. target classes are overlapping. In cases where the number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.It does not execute very well when the data set has more sound i.e. target classes are overlapping. In cases where the number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.

13. Notes should be written on
14. The kNN algorithm has a validation flaw.
2. In the kNN algorithm, the k value is chosen.
3. A decision tree with inductive bias

Answer:-->1.

1. The kNN algorithm has a validation flaw-->In particular, the model created via kNN is just the available labeled data, placed in some metric space. In other words, for kNN, there is no training step because there is no model to build. Template matching & interpolation is all that is going on in kNN. Neither is there a validation step.1.Training error Rate 2.Validation Error Rate. If we observe the training error rate graph it can be seen that error increases for increasing value of K,also error is zero for K=1. This is because the closest point to any training data point is itself. Hence the prediction is always accurate with K=1
2. In the kNN algorithm, the k value is chosen.--->The k value in the k-NN algorithm defines how many neighbors will be checked to determine the classification of a specific query point. For example, if k=1, the instance will be assigned to the same class as its single nearest neighbor.The optimal K value usually found is the square root of N, where N is the total number of samples. Use an error plot or accuracy plot to find the most favorable K value. KNN performs well with multi-label classes, but you must be aware of the outliers.In KNN, finding the value of k is not easy. A small value of k means that noise will have a higher influence on the result and a large value make it computationally expensive. Data scientists usually choose as an odd number if the number of classes is 2 and another simple approach to select k is set k=sqrt(n).

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. But inductive bias is absolutely essential to machine learning (and human learning, for that matter). Without inductive bias, a learner can't generalize from observed examples to new examples better than random guessing

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

Answer:-->The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.K nearest neighbors (kNN) is an efficient lazy learning algorithm and has successfully been developed in real applications. It is natural to scale the kNN method to the large scale datasets.Feature scaling is essential for machine learning algorithms that

calculate distances between data. If not scaled, the feature with a higher value range starts dominating when calculating distances. KNN which uses Euclidean distance is one such algorithm which essentially require scaling.Here are some of the advantages of using the k-nearest neighbors algorithm: It's easy to understand and simple to implement. It can be used for both classification and regression problems. It's ideal for non-linear data since there's no assumption about underlying data

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

Answer:-->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. It's main disadvantages are that it is quite computationally inefficient and its difficult to pick the “correct” value of K. However, the advantages of this algorithm is that it is versatile to different calculations of proximity, it's very intuitive and that it's a memory based approach. Because it does no training at all when you supply the training data. At training time, all it is doing is storing the complete data set but it does not do any calculations at this point.

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

Answer_-->A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes).A decision tree is a flowchart that starts with one main idea and then branches out based on the consequences of your decisions. It's called a “decision tree” because the model typically looks like a tree with branches Root Nodes – It is the node present at the beginning of a decision tree from this node the population starts dividing according to various features.

Decision Nodes – the nodes we get after splitting the root nodes are called Decision Node

Leaf Nodes – the nodes where further splitting is not possible are called leaf nodes or terminal nodes

Sub-tree – just like a small portion of a graph is called sub-graph similarly a sub-section of this decision tree is called sub-tree.

Pruning – is nothing but cutting down some nodes to stop overfitting.

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

Answer:-->Decision Node: When a sub-node splits into further sub-nodes, it's a decision node. Leaf Node or Terminal Node: Nodes that do not split are called leaf or terminal nodes. Pruning: Removing the sub-nodes of a parent node is called pruning. A tree is grown through splitting and shrunk through pruning.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.All Tree nodes are connected by links called edges . It's an important part of trees , because it's manages the relationship between nodes . Leaves are the last nodes on a tree. They are nodes without children.

18. What is a decision tree's entropy?

Answer:-->Entropy is an information theory metric that measures the impurity or uncertainty in a group of observations. It determines how a decision tree chooses to split data.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.The entropy here is approximately 0.88. This is considered a high entropy , a high level of disorder (meaning low level of purity). Entropy is measured between 0 and 1. (Depending on the number of classes in your dataset, entropy can be greater than 1 but it means the same thing , a very high level of disorder. entropy in machine learning is related to randomness in the information being processed in your machine learning project

19. In a decision tree, define knowledge gain.

Answer:-->Information gain is the basic criterion to decide whether a feature should be used to split a node or not. The feature with the optimal split i.e., the highest value of information gain at a node of a decision tree is used as the feature for splitting the node.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.Information Gain, or IG for short, measures the reduction in entropy or surprise by splitting a dataset according to a given value of a random variable. A larger information gain suggests a lower entropy group or groups of samples, and hence less surprise.

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

Answer:-->Compared to other Machine Learning algorithms Decision Trees require less data to train. They can be used for Classification and Regression. They are simple. They are tolerant to missing values.It ensures a comprehensive analysis of the consequences of each branch while also recognizing which nodes might need further analyzing. Therefore, it is easy to validate the algorithm using statistical tests. This makes Decision Trees an accountable mode.Decision trees are composed of three main parts—decision nodes (denoting choice), chance nodes (denoting probability), and end nodes (denoting outcomes)

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

Answer:-->One of the limitations of decision trees is that they are largely unstable compared to other decision predictors. A small change in the data can result in a major change in the structure of the decision tree, which can convey a different result from what users will get in a normal event.Overfitting (where a model interprets meaning from irrelevant data) can become a problem if a decision tree's design is too complex. They are not well-suited to continuous variables (i.e. variables which can have more than one value, or a spectrum of values)

22. Briefly describe the random forest model.

Answer:-->The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual treeThe random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual treeRandom forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It can perform both regression and classification tasks. A random forest produces good predictions that can be understood easily. It can handle large datasets efficiently. The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.It builds decision trees on different samples and takes their majority vote for classification and average in case of regressionStep 1: Select random samples from a given data or training set. Step 2: This algorithm will construct a decision tree for every training data. Step 3: Voting will take place by averaging the decision tree. Step 4: Finally, select the most voted prediction result as the final prediction result