|  |  |
| --- | --- |
| 1 | What is the concept of supervised learning? What is the significance of the name? |
| Ans. | 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 to classify data or predict outcomes accurately. |
| 2 | In the hospital sector, offer an example of supervised learning. |
| Ans. | Supervised learning uses a training set to teach models to yield the desired output. This training dataset includes inputs and correct outputs, which allow the model to learn over time. The algorithm measures its accuracy through the loss function, adjusting until the error has been sufficiently minimized. |
| 3 | Give three supervised learning examples. |
| Ans. | * Linear Regression. * Nearest Neighbor. * Gaussian Naive Bayes. * Decision Trees. * Support Vector Machine (SVM) * Random Forest. |
| 4 | In supervised learning, what are classification and regression? |
| Ans. | Fundamentally, classification is about predicting a label and regression is about predicting a quantity. That classification is the problem of predicting a discrete class label output for example. That regression is the problem of predicting a continuous quantity output for example. |
| 5 | Give some popular classification algorithms as examples. |
| Ans. | * k-Nearest Neighbors * Decision Trees. * Naive Bayes * Support Vector Machine * Logistic Regression |
| 6 | Briefly describe the SVM model. |
| Ans. | Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.  The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. |
| 7 | In SVM, what is the cost of misclassification? |
| Ans. | In cost-sensitive learning instead of each instance being either correctly or incorrectly classified, each class (or instance) is given a misclassification cost. Thus, instead of trying to optimize the accuracy, the problem is then to minimize the total misclassification cost. |
| 8 | In the SVM model, define Support Vectors. |
| Ans. | The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector. |
| 9 | In the SVM model, define the kernel. |
| Ans. | The kernel functions play a very important role in SVM. Their job is to take data as input and transform it into any required form. They are significant in SVM as they help in determining various important things.  A kernel is a function used in SVM for helping to solve problems. They provide shortcuts to avoid complex calculations. The amazing thing about kernel is that we can go to higher dimensions and perform smooth calculations with the help of it.  We can go up to an infinite number of dimensions using kernels.  Sometimes, we cannot have a hyperplane for certain problems. This problem arises when we go up to higher dimensions and try to form a hyperplane.  A kernel helps to form the hyperplane in the higher dimension without raising the complexity. |
| 10 | What are the factors that influence SVM's effectiveness? |
| Ans. | 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. |
| 11 | What are the benefits of using the SVM model? |
| Ans. | 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? |
| Ans. | 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 |
| Ans. | **The kNN algorithm has a validation flaw.**  The relatively low accuracy of kNN is caused by several factors. One of them is that every characteristic of the method has the same result on calculating distance. The solution to this problem is to give weight to each data characteristic.  **In the kNN algorithm, the k value is chosen.**  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.  **A decision tree with inductive bias**  Shorter trees are preferred over longer ones. 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? |
| Ans. | 1. No Training Period: KNN is called Lazy Learner (Instance-based learning). It does not learn anything in the training period. It does not derive any discriminative function from the training data. In other words, there is no training period for it. It stores the training dataset and learns from it only at the time of making real-time predictions. This makes the KNN algorithm much faster than other algorithms that require training e.g. SVM, Linear Regression, etc.  2. Since the KNN algorithm requires no training before making predictions, new data can be added seamlessly which will not impact the accuracy of the algorithm.  3. KNN is very easy to implement. There are only two parameters required to implement KNN i.e. the value of K and the distance function (e.g. Euclidean or Manhattan etc.) |
| 15 | What are some of the kNN algorithm's drawbacks? |
| Ans. | 1. Does not work well with the large dataset: In large datasets, the cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithm.  2. Does not work well with high dimensions: The KNN algorithm doesn't work well with high dimensional data because, with a large number of dimensions, it becomes difficult for the algorithm to calculate the distance in each dimension.  3. Need feature scaling: We need to do feature scaling (standardization and normalization) before applying the KNN algorithm to any dataset. If we don't do so, KNN may generate wrong predictions.  4. Sensitive to noisy data, missing values, and outliers: KNN is sensitive to noise in the dataset. We need to manually impute missing values and remove outliers. |
| 16 | Explain the decision tree algorithm in a few words. |
| Ans. | 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. |
| 17 | What is the difference between a node and a leaf in a decision tree? |
| Ans. | 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). |
| 18 | What is a decision tree's entropy? |
| Ans. | Entropy helps us to build an appropriate decision tree for selecting the best splitter. Entropy can be defined as a measure of the purity of the sub split. Entropy always lies between 0 to 1. The entropy of any split can be calculated by this formula. |
| 19 | In a decision tree, define knowledge gain. |
| Ans. | 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. |
| 20 | Choose three advantages of the decision tree approach and write them down. |
| Ans. | 1. 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. 2. Easy to prepare. 3. Less data cleaning is required. |
| 21 | Make a list of three flaws in the decision tree process. |
| Ans. | 1. Overfitting the data. 2. Guarding against bad attribute choices. 3. Handling continuous-valued attributes. 4. Handling missing attribute values. 5. Handling attributes with differing costs. |
| 22 | Briefly describe the random forest model. |
| Ans. | 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 tree. |