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

Ans:  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. After that, the machine is provided with a new set of examples(data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

The significance of the name is allows collecting data and produces data output from previous experiences and helps to optimize performance criteria with the help of experience.

Some common real-life examples are:

* Customer Analysis – Provides better recommendations by identifying patterns in customers’ purchase history
* Predictive Analysis – Offers accurate insights into business data points
* Fraud Detection –  Identifies fraudulent transactions by evaluating datasets
* Image Classification – Identifies, locates, and categorizes objects from videos and uses them during vision analysis to accurately identify images

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

### Ans: Managing Medical Data:Machine learning is actually advancing the health care industry by implementing cognitive technology in order to unwind a huge amount of medical records and also in order to perform any power diagnosis. Machine learning helps predict the intent of a user. Implementing machine learning in an organization’s workflow can develop a personalized user experience that allows the company to make better decisions and better actions that enhance the customer’s experience which benefits the organization. Therefore, machine learning helps to store, collect and reform data.

### Helps in Medical Diagnosis:According to the Global Market Insights, Medical imaging and diagnosis powered by AI should witness more than 40% growth to surpass $2.5 billion by 2024. With the help of machine learning and deep learning models, AI is actually revolutionizing the image diagnosis field in medicine. The one major application of AI in medical diagnosis is MRI scans. AI has taken over the complex analysis of MRI scans and it has made it a much simpler process.

### Detecting Diseases at an Earlier Stage:Machine learning played a very important role in the early predictions of medical conditions such as heart attacks and diabetes. There are many AI-based wearables that are being developed to monitor the health of a person and display any warnings when the devices observe something unusual or unlikely. For eg:- Fitbit and Apple watch. These devices monitor a person’s heart rate, sleep cycle, breathing rate, activity level, blood pressure, so on. It keeps the record of these measures 24×7.

### Personalized Medicine:The predictive analysis of Machine learning can help users to get personalized treatment. Generally, nurses are bound to choose from a specific set of diagnoses or predict the risk to the patient using a fixed formula based on the history and available genetic information. Whereas, machine learning in medicine predicts the data of the patient by analyzing the medical history to generate multiple treatment options. Due to these treatments being based on the user’s data they’re more likely to suit the patient and are more personalized.

### Personalized Medicine:The predictive analysis of Machine learning can help users to get personalized treatment. Generally, nurses are bound to choose from a specific set of diagnoses or predict the risk to the patient using a fixed formula based on the history and available genetic information. Whereas, machine learning in medicine predicts the data of the patient by analyzing the medical history to generate multiple treatment options. Due to these treatments being based on the user’s data they’re more likely to suit the patient and are more personalized.

3. Give three supervised learning examples.

Ans:

* Email Filtering Supervised learning is commonly used in email filtering to classify incoming emails as spam or legitimate.
* Credit Scoring In credit scoring, supervised learning is used to predict the creditworthiness of loan applicants.
* Voice Recognition Supervised learning is utilized in voice recognition to help virtual assistants and other applications recognize and understand spoken commands

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

Ans: Supervised learning is classified into two categories of algorithms:

* Classification: A classification problem is when the output variable is a category, such as “Red” or “blue” , “disease” or “no disease”.
* Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

5. Give some popular classification algorithms as examples.

Ans: Logistic regression

* + - Support vector machine
    - Decision Tree
    - KNN
    - Naïve Bayes

6. Briefly describe the SVM model.

Ans: Support Vector Machine(SVM) is a [supervised machine learning](https://www.geeksforgeeks.org/supervised-unsupervised-learning/) algorithm used for both classification and regression. he objective of the SVM algorithm is to find a [hyperplane](https://www.geeksforgeeks.org/separating-hyperplanes-in-svm/) in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.T he objective of the SVM algorithm is to find a [hyperplane](https://www.geeksforgeeks.org/separating-hyperplanes-in-svm/) in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three. finds the maximum margin as done with previous data sets along with that it adds a penalty each time a point crosses the margin. So the margins in these types of cases are called soft margins. When there is a soft margin to the data set, the SVM tries to minimize (1/margin+∧(∑penalty)). Hinge loss is a commonly used penalty. If no violations no hinge loss. violations hinge loss proportional to the distance of violation. The SVM kernel is a function that takes low-dimensional input space and transforms it into higher-dimensional space, ie it converts nonseparable problems to separable problems. Simply put the kernel, does some extremely complex data transformations and then finds out the process to separate the data based on the labels or outputs defined.A screenshot of a computer

Description automatically generated

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

Ans: soft margin which allows "misclassification, for SVM is that some dataset are not linear, which means we can not find the hyperplane separating the different classes of the data.

Mathematically, we rewrite the constrain from

yi(w⋅xi−b)≥1 1≤i≤n

to

yi(w⋅xi−b)≥1−ξi 1≤i≤n1

The slack variable ξi able to states the "misclassification".

However, we can not allow too much "misclassification". we rewrite the objective function,

argminw,ξ,b{12∥w∥2+C∑i=1nξi}

The tuning parameter C which you claim "the price of the misclassification" is exactly the weight for penalizing the "soft margin".

8. In the SVM model, define Support Vectors.

Ans: **S**upport Vectors are the points that are closest to the hyperplane. A separating line will be defined with the help of these data points.

9. In the SVM model, define the kernel.

Ans:  A **kernel** is nothing but a measure of similarity between data points. The **kernel function** in a kernelized SVM tells you, that given two data points in the original feature space, what the similarity is between the points in the newly transformed feature space. There are various kernel functions available, Radial Basis Function Kernel (RBF) and Polynomial Kernel etc.

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

Ans: The effectiveness of SVM depends on the selection of kernel, the kernel's parameters, and soft margin parameter λ .�A common choice is a Gaussian kernel, which has a single parameter *�*. The best combination of � and � is often selected by a [grid search](https://en.wikipedia.org/wiki/Grid_search) with exponentially growing sequences of λ � and *�.*Typically, each combination of parameter choices is checked using [cross validation](https://en.wikipedia.org/wiki/Cross-validation_(statistics)), and the parameters with best cross-validation accuracy are picked.

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

Ans: Benefits of using SVM model:

* SVM works better when the data is Linear
* It is more effective in high dimensions
* With the help of the kernel trick, we can solve any complex problem
* SVM is not sensitive to outliers
* Can help us with Image classification

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

Ans: The drawbacks of using the SVM model are:

* Choosing a good kernel is not easy
* It doesn’t show good results on a big dataset
* The SVM hyperparameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact

13. Notes should be written on

1. The kNN algorithm has a validation flaw.

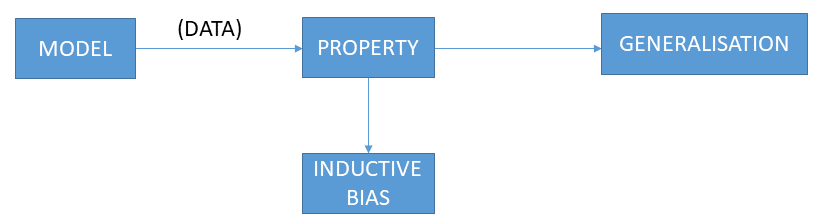
Ans: kNN produces predictions by looking at the knearest neighbours of a case x to predict its y. In particular, the kNN model basically consists of its training cases - but that's the cross validation procedure doesn't care about at all.

2. In the kNN algorithm, the k value is chosen: The value of k determines the number of neighbors to look at. In classification problems, it can be helpful to use odd values of k, since it requires a majority vote (which can be more difficult with an even number). Using **Euclidian distance**, which represents the shortest distance between two points, the point is added in the cluster and updating of cluster and centroid takes place. When we decrease the value of K to 1, our predictions become less stable. The accuracy decreases and the metric “F-Measure” becomes more sensitive to outliers. For better results, increase the value of K until the F-Measure value is higher than the threshold.Elbow method is used for finding out optimal value of K.

3. A decision tree with inductive bias

Ans: Before learning a model given a data and a learning algorithm, there are a few assumptions a learner makes about the algorithm. These assumptions are called the inductive bias. It is like the property of the algorithm.

For eg. in the case of decision trees, the depth of the tress is the inductive bias. If the depth of the tree is too low, then there is too much generalisation in the model. Similarly, if the depth of the tree is too much, there is too less generalisation and while testing the model on a new example, we might reach a particular example used to train the model. This may give us incorrect results.



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

Ans:

* It is simple to implement.
* No training is required before classification.

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

Ans:

* Can be cost-intensive when working with a large data set.
* A lot of memory is required for processing large data sets.
* Choosing the right value of K can be tricky.

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

Ans: Decision trees are a type of supervised learning algorithm which are used for mainly classification and regression.They have a tree like structure in which the internal nodes are "tests" for attributes and the branches are the results of the "tests". The leaf nodes will be the class labels i.e., the output of the learner. Given below is the basic structure of a decision tree.



A tree can be “learned” by splitting the source set into subsets based on Attribute Selection Measures. Attribute selection measure (ASM) is a criterion used in decision tree algorithms to evaluate the usefulness of different attributes for splitting a dataset. The goal of ASM is to identify the attribute that will create the most homogeneous subsets of data after the split, thereby maximizing the information gain. This is done using information gain, Cal gini index or entropy. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of a decision tree classifier does not require any domain knowledge or parameter setting and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high-dimensional data.

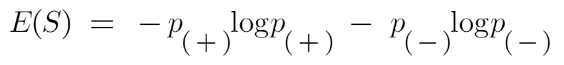
17. What is the difference between a node and a leaf in a decision tree?

Ans:

* Root Node: It is the topmost node in the tree,  which represents the complete dataset. It is the starting point of the decision-making process.
* Decision/Internal Node: A node that symbolizes a choice regarding an input feature. Branching off of internal nodes connects them to leaf nodes or other internal nodes.
* Leaf/Terminal Node: A node without any child nodes that indicates a class label or a numerical value.

18. What is a decision tree's entropy?

Ans: Entropy is nothing but the uncertainty in dataset or measure of disorder. The formula for Entropy is shown below:

 Here p+ is the probability of positive class

p– is the probability of negative class

S is the subset of the training example

Apure sub-split means that either you should be getting “yes”, or you should be getting “no”. In order to make a decision tree, we need to calculate the impurity of each split, and when the purity is 100%, we make it as a leaf node. The higher the Entropy, the lower will be the purity and the higher will be the impurity. Goal is to decrease the uncertainty or impurity in the dataset

19. In a decision tree, define knowledge gain.

Ans: Information gain measures the reduction in entropy or variance that results from splitting a dataset based on a specific property. it is also a deciding factor for which attribute should be selected as a decision node or root node. It is used in decision tree algorithms to determine the usefulness of a feature by partitioning the dataset into more homogeneous subsets with respect to the class labels or target variable. The higher the information gain, the more valuable the feature is in predicting the target variable.

information gain Decision tree algorithm

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

Ans: Decision trees are able to generate understandable rules.

* Decision trees perform classification without requiring much computation.
* Decision trees are able to handle both continuous and categorical variables.
* Decision trees provide a clear indication of which fields are most important for prediction or classification.
* Ease of use: Decision trees are simple to use and don’t require a lot of technical expertise, making them accessible to a wide range of users.
* Scalability: Decision trees can handle large datasets and can be easily parallelized to improve processing time.
* Missing value tolerance: Decision trees are able to handle missing values in the data, making them a suitable choice for datasets with missing or incomplete data.
* Handling non-linear relationships: Decision trees can handle non-linear relationships between variables, making them a suitable choice for complex datasets.
* Ability to handle imbalanced data: Decision trees can handle imbalanced datasets, where one class is heavily represented compared to the others, by weighting the importance of individual nodes based on the class distribution.

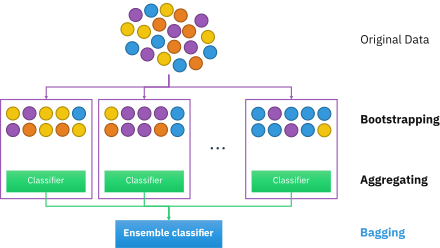
21. Make a list of three flaws in the decision tree process.

Ans: Decision trees are able to generate understandable rules.

* Decision trees perform classification without requiring much computation.
* Decision trees are able to handle both continuous and categorical variables.
* Decision trees provide a clear indication of which fields are most important for prediction or classification.
* Ease of use: Decision trees are simple to use and don’t require a lot of technical expertise, making them accessible to a wide range of users.
* Scalability: Decision trees can handle large datasets and can be easily parallelized to improve processing time.
* Missing value tolerance: Decision trees are able to handle missing values in the data, making them a suitable choice for datasets with missing or incomplete data.
* Handling non-linear relationships: Decision trees can handle non-linear relationships between variables, making them a suitable choice for complex datasets.
* Ability to handle imbalanced data: Decision trees can handle imbalanced datasets, where one class is heavily represented compared to the others, by weighting the importance of individual nodes based on the class distribution.

22. Briefly describe the random forest model.

Ans: Random forest works on the Bagging principle which chooses a random sample/random subset from the entire data set. Hence each model is generated from the samples (Bootstrap Samples) provided by the Original Data with replacement known as row sampling. This step of row sampling with replacement is called bootstrap. Now each model is trained independently, which generates results. The final output is based on majority voting after combining the results of all models. This step which involves combining all the results and generating output based on majority voting, is known as aggregation.



#### Steps Involved in Random Forest Algorithm

Step 1: In the Random forest model, a subset of data points and a subset of features is selected for constructing each decision tree. Simply put, n random records and m features are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression, respectively.