

growth and innovation because it enables companies to detect customer behavior and operational challenges.

Machine Learning engineers are critical to most industries as they play important roles in organizational growth and improving customer experience. If you are a job seeker looking for a [Machine Learning Engineer job](#) or a recruiter looking to [hire the best ML engineer](#), the following list of 100 machine learning interview questions and answers will help you prepare for the interview or assess the candidates.

Table of contents

[Basic Machine Learning interview questions and answers \(51\)](#)

[Intermediate Machine Learning interview questions and answers \(28\)](#)

[Advanced Machine Learning interview questions and answers \(29\)](#)

BASIC MACHINE LEARNING INTERVIEW QUESTIONS AND ANSWERS

1. Differentiate between Training Sets and Test Sets?

[Hide Answer](#)

Training Set

- The data in the training set are the examples provided to the model to train that particular model.

I'm hiring developers

I'm looking for jobs

than testing data is recommended.

- To train the model, the **training set is the labeled data that is used.**

Test Set

- The data in the test are used to test the model accuracy of the already trained model.
- The Test Set contains around 20%-30% of the total data. This data is then further used to test the accuracy of the trained model.

2. Define Bias and Variance.

[Hide Answer](#)

Bias

When a model makes predictions, a disparity between the model's prediction values and actual values arises, and this difference is known as bias. Bias is the incapacity of machine learning algorithms like Linear Regression to grasp the real relationship between data points.

Variance

If alternative training data were utilized, the variance would describe the degree of variation in the prediction. In layman's terms, variance describes how far a random variable deviates from its predicted value.

3. You have come across some missing data in your dataset. How will you handle it?

[Hide Answer](#)

In order to handle some missing or corrupted data, the easiest way is to just replace the corresponding rows and columns, which contain the incorrect data, with some different values. The two most useful functions in Panda for this purpose are `isnull()` and `fillna()`.

- `isnull()`: is used to find missing values in a dataset
- `fillna()`: is used to fill missing values with 0's

4. Explain Decision Tree Classification.

[Hide Answer](#)

A decision tree uses a tree structure to generate any regression or classification models. While the decision tree is developed, the datasets are split up into ever-

5. How is a logistic regression model evaluated?

[Hide Answer](#)

One of the best ways to evaluate a logistic regression model is to use a confusion matrix, which is a very specific table that is used to measure the overall performance of any algorithm.

Using a confusion matrix, you can easily calculate the Accuracy Score, Precision, Recall, and F1 score. These can be extremely good indicators for your logistic regression model.

If the recall of your model is low, then it means that your model has too many False Negatives. Similarly, if the precision of your model is low, it signifies that your model has too many False Positives. In order to select a model with a balanced precision and recall score, the F1 Score must be used.

6. To start Linear Regression, you would need to make some assumptions. What are those assumptions?

[Hide Answer](#)

To start a Linear Regression model, there are some fundamental assumptions that you need to make:

- The model should have a multivariate normal distribution
- There should be no auto-correlation
- Homoscedasticity, i.e, the dependent variable's variance should be similar to all of the data
- There should be a linear relationship
- There should be no or almost no multicollinearity present

7. What is multicollinearity and how will you handle it in your regression model?

[Hide Answer](#)

present complications.

There are various ways to check and handle the presence of multicollinearity in your regression model. One of them is to calculate the **Variance Inflation Factor (VIF)**. If your model has a VIF of less than 4, there is no need to investigate the presence of multicollinearity. However, if your VIF is more than 4, an investigation is very much required, and if VIF is more than 10, there are serious concerns regarding multicollinearity, and you would need to correct your regression model.

8. Explain why the performance of XGBoost is better than that of SVM?

[Hide Answer](#)

XGBoost is an ensemble approach that employs a large number of trees. This implies that when it repeats itself, it becomes better.

If our data isn't linearly separable, SVM, being a linear separator, will need to use a Kernel to bring it to a point where it can be split. Due to there not being an ideal Kernel for every dataset, this can be limiting.

9. Why is an encoder-decoder model used for NLP?

[Hide Answer](#)

An encoder-decoder model is used to create an output sequence based on a given input sequence. The final state of the encoder is used as the initial state of the decoder, and this makes the encoder-decoder model extremely powerful. This also allows the decoder to access the information that is taken from the input sequence by the encoder.

10. What are Machine Learning and Artificial Intelligence?

[Hide Answer](#)

[machine learning vs artificial intelligence vs deep learning.](#)

11. Differentiate between Deep Learning and Machine Learning?

[Hide Answer](#)

Machine Learning adopts algorithms to **learn from data sets and apply this to future decision making**. **Deep Learning** is a subset of Machine Learning that uses **large amounts of data and complex algorithms to create neural networks that can learn and make decisions on their own**.

12. What is cross validation?

[Hide Answer](#)

Cross validation is a concept used to evaluate models' performances to avoid **overfitting**. It is an easy method to compare the predictive capabilities of models and is best suitable when limited data is available.

13. What are the types of Machine learning?

[Hide Answer](#)

There are mainly three types of Machine Learning, viz:

Reinforcement learning: It is about taking the best possible action to maximize reward in a particular scenario. It is used by various software and machines to find the best path it should take in a given situation.

Supervised learning: Using labeled datasets to train algorithms to classify data easily for predicting accurate outcomes.

[Hide Answer](#)

Supervised algorithms are those that use labeled data to learn a mapping function from input variables to output variables. Unsupervised algorithms learn from unlabeled data and discover hidden patterns and structures in the data.

15. What is Selection Bias?

[Hide Answer](#)

Selection Bias is a statistical error that brings about a bias in the sampling portion of the experiment. This, in turn, causes more selection of the sampling portion than other groups, which brings about an inaccurate conclusion.

16. What is the difference between correlation and causality?

[Hide Answer](#)

Correlation is the relation of one action (A) to another action (B) when A does not necessarily lead to B, but Causality is the situation where one action (A) causes a result (B).

17. What is the difference between Correlation and Covariance?

[Hide Answer](#)

Correlation quantifies the relationship between two random variables with three values: 0,1 and -1.

Covariance is the measure of how two different variables are related and how changes in one impact the other. Read [correlation vs covariance](#) to know about these two and for a further in-depth comparison.

Supervised learning algorithms are trained using labeled data, while reinforcement learning algorithms are trained using a reward function. Supervised learning algorithms are used to predict a given output, while reinforcement learning algorithms are used to maximize a reward by taking a series of actions.

19. What are the requirements of reinforcement learning environments?

[Hide Answer](#)

State, reward data, agent, and environment. It is entirely different from other machine learning paradigms. Here we have an agent and an environment. The environment refers to a task or simulation; the agent is an algorithm that interacts with the environment and tries to solve it.

20. What different targets do classification and regression algorithms require?

[Hide Answer](#)

Regression algorithms require categorical and numerical targets. Here, regression finds correlations between dependent and independent variables. It helps predict continuous variables such as market trends and weather patterns.

On the other hand, classification is an algorithm that segregates the dataset into classes on various parameters. Here classification algorithms work to predict the willingness of bank customers to pay their loans, email, or spam classification.

21. What are five popular algorithms used in Machine Learning?

[Hide Answer](#)

represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.

K-nearest neighbor: K-nearest neighbor (KNN) is a supervised learning algorithm used for classification and regression. The algorithm finds the k-nearest data points in the training dataset and uses them to make predictions. It works by calculating the distance between the query point and the k-nearest data points and then uses the labels of these points to make a prediction.

Support Vector Machines (SVM): It is used to create the best line or decision boundary that can segregate n-dimensional space into classes to put the new data point in the correct category quickly.

Probabilistic networks: They are graphical models of interactions among a set of variables, where the variables are represented as nodes of a graph and the interaction as directed edges between the vertices. It allows a compact description of complex stochastic relationships among several random variables.

22. What is the confusion matrix?

[Hide Answer](#)

The [confusion matrix](#) consists of an error matrix table used for concluding the performance of a classification algorithm. It determines the classification models' performance for a given test data set. It has multiple categorical outputs, but it can only be determined if the actual values for test data are known.

23. List the differences between KNN and k-means clustering.

[Hide Answer](#)

This is a supervised technique	This is an unsupervised technique
--------------------------------	-----------------------------------

24. What are the differences between Type I error and Type II error?

[Hide Answer](#)

Type I	Type II
False positive	False negative
This states that something has happened when it has not happened	This states that nothing has happened when it has actually happened

25. What is semi-supervised learning?

[Hide Answer](#)

A semi-supervised learning happens when a small amount of labeled data is introduced to an algorithm. The algorithm then studies that data and uses it on unlabeled data. Semi-supervised learning combines the efficiency of unsupervised learning and the performance of supervised learning.

26. Where are semi-supervised learning applied?

[Hide Answer](#)

Some areas it is applied include labeling data, fraud detection, and machine translation.

Stemming is a normalization technique that removes any affix joined to a word, leaving it in its base state. It makes text easier to process. It is commonly used in information retrieval, an important step in text pre-processing and text mining applications. Stemming can be used in various NLP tasks such as text classification, information retrieval, and text summarization.

28. What is Lemmatization?

[Hide Answer](#)

This is a normalization technique that converts a word into a lemma form, or the root word, which is not the stem word. It is a process in which a word is reduced to its base form, but not similar to stemming; it considers the context of the word and produces a valid word. Lemmatization is quite difficult compared to stemming because it requires a lot more knowledge about the structure of a language; it's a much more intensive process than just trying to set up a heuristic stemming algorithm. Lemmatization is often used in natural language processing (NLP) applications to improve text analysis and feature extraction.

29. What is a PCA?

[Hide Answer](#)

PCA means [Principal Component Analysis](#), and is mainly used for dimension reduction. It is a statistical technique used to reduce the dimensionality of large datasets while retaining as much information as possible. In other words, it identifies patterns and correlations among variables and summarizes them into a smaller set of uncorrelated variables called principal components.

PCA is commonly used in data preprocessing and exploratory data analysis to simplify data visualization, identify outliers, and reduce noise in the data. It is also

[Hide Answer](#)

Support vectors are the data points in a dataset that are closest to the hyperplane (the line that separates the classes in the dataset) and are used to build the classifier.

31. In terms of access, how are array and linked lists different?

[Hide Answer](#)

Linked lists allow users to transverse the full linked lists, even up to the element in a sequential access pattern. However, an array provides access to elements directly using their index value.

32. What is P-value?

[Hide Answer](#)

P-value or probability value indicates the probability of obtaining the observed data or more extreme values by random chance. A small P-value suggests that the observed result is unlikely and that observed data is consistent with the null hypothesis and provides evidence to support the alternative hypothesis.

33. What techniques are used to find resemblance in the recommendation system?

[Hide Answer](#)

Cosine and Pearson Correlation are techniques used to find resemblance in recommendation systems. Where the Pearson correlation coefficient is the

[Hide Answer](#)

Classification is a concept used to produce discrete results and to classify data into specific regions. On the other hand, regression is used to assess the relationship between independent variables and dependent variables.

35. What does the area under the ROC curve indicate?

[Hide Answer](#)

ROC stands for Receiver Operating Characteristic. It measures the usefulness of a test where the larger the area, the more useful the test. These areas are used to compare the effectiveness of the tests. A higher AUC (area under the curve) generally indicates that the model is better at distinguishing between the positive and negative classes. AUC values range from 0 to 1, with a value of 0.5 indicating that the model is no better than random guessing, and a value of 1 indicating perfect classification.

36. What is a neural network?

[Hide Answer](#)

Much like a human brain, the neural network is a network of different neurons connected in a way that helps information flow from one neuron to the other. It is a function that maps input to desired output with a given set of inputs. Structurally, it is organized into an input layer, an output layer, and one or more hidden layers.

37. What is an Outlier?

[Hide Answer](#)

38. What is another name for a Bayesian Network?

[Hide Answer](#)

Casual network, Belief Network, Bayes network, Bayes net, Belief Propagation Network, etc. are some of its other names. It is a probabilistic graphical model that showcases a set of variables and their conditional dependencies.

39. What is ensemble learning?

[Hide Answer](#)

Ensemble learning is a method that merges multiple machine learning models to create various powerful models. The aim is to provide better performance by combining models rather than sticking to a single model.

40. What is clustering?

[Hide Answer](#)

Clustering is a process of grouping sets of items into several groups. Items or objects must be similar within the cluster and different from other objects in other clusters. The goal of clustering is to identify patterns and similarities in the data that can be used to gain insights and make predictions. Different clustering algorithms use different methods to group data points based on their features and similarity measures, such as distance or density. Clustering is commonly used in various applications such as customer segmentation, image and text classification, anomaly detection, and recommendation systems.

Collinearity is when two predictor variables in a multiple regression share some correlations.

42. What is overfitting?

[Hide Answer](#)

Overfitting is said to happen when a statistical model observes and learns the details in the training data to the point that it starts negatively impacting the model's performance on new datasets.

43. What is the Bayesian Network?

[Hide Answer](#)

Bayesian network represents a graphical model between sets of variables. We say it probabilistic because these networks are built on a probability distribution and also use probability theory for prediction and anomaly detection. Bayesian networks are used in for reasoning, diagnostic, anomaly detection, prediction to list a few.

44. What is the time series?

[Hide Answer](#)

Time series is a particular sequence of data observations in successive order collected over a period. It usually does not need any maximum or minimum time input. It basically forecasts target values based solely on a known history of target values. It is used to predict time-based predictions such as signal processing,

[Hide Answer](#)

Dimension reduction is the reduction of variables put under consideration. It lessens the number of features in a dataset while saving as much information as possible. This can be done for various reasons, such as to improve the performance of a learning algorithm, reduce the complexity of a model, or make it easier to visualize the data.

46. What is underfitting?

[Hide Answer](#)

Underfitting is a type of error in ML models where the model fails to capture the underlying pattern of the data. It occurs when a model is too simplistic and is unable to capture the complexity of the data, leading to poor generalization performance on unseen data. In other words, the model is not complex enough to accurately capture the relationship between the input and output variables. This often leads to high bias and low variance.

47. What is sensitivity?

[Hide Answer](#)

This is the probability that the prediction outcome of the model is true when the value is positive. It can be described as the metric for evaluating a model's ability to predict the true positives of each available category.

Sensitivity = $TP / (TP + FN)$ (i.e. True Positive / True Positive + False Negative)

48. What is specificity?

[Hide Answer](#)

49. What are the differences between stochastic gradient descent (SGD) and gradient descent (GD)?

[Hide Answer](#)

Both of these gradients are algorithms used to ascertain the parameters that will minimize a loss function. However, in the case of GB, all training samples are evaluated for each set of parameters. On the contrary, for SGB, one training sample is always evaluated for a set of parameters.

50. What is an Array?

[Hide Answer](#)

An array is a collection of data elements of the same type, such as integers, strings, or floating-point numbers, stored in contiguous memory locations. Every component of an array is identified by an index that represents its position in the array.

51. What is a linked list?

[Hide Answer](#)

This is an ordered collection of similar data type elements joined with pointers. It consists of several individual allocated nodes or a series of connected nodes. Each node contains data plus a pointer or the address of the next node in the list.

Apply Now

INTERMEDIATE MACHINE LEARNING INTERVIEW QUESTIONS AND ANSWERS

1. Explain Decision Tree in ML.

[Hide Answer](#)

Decision tree is a supervised machine learning method that shows a map of possible outcomes of related decisions, similar to a tree structure using datasets divided into smaller subsets as the decision tree develops.

2. Why is the Naive Bayes Method 'Naive'?

[Hide Answer](#)

The **Naive Bayes** is called naive because, as a supervised learning algorithm, it makes assumptions by applying the Bayes Theorem that all attributes are independent of each other.

3. State the Bayes' theorem for Naive Bayes Method.

[Hide Answer](#)

The Bayes' theorem for Naive Bayes Method is stated as follows:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

Where,

$P(A|B)$ is the posterior probability of class (target) given predictor (attribute)

$P(B|A)$ is the likelihood which is the probability of predictor given class

4. What is Static Memory Allocation?

[Hide Answer](#)

This is when memory is allocated as soon as the array is declared at compilation time. This allocation offers more efficiency and speed in operation as there is no overhead. However, once allocated, the assigned memory space cannot be deallocated during compilation. Static memory is automatically deallocated when the program terminates, and the memory is returned to the operating system.

5. How would you define Dynamic Memory Allocation?

[Hide Answer](#)

Memory allocation is said to be dynamic if the allocation occurs at runtime whenever a new node is added. Nowadays, modern operating systems use dynamic memory allocation.

DMA is useful in scenarios such as:

- You don't know the amount of memory needed beforehand.
 - You want data structures without any upper limit
 - For efficient use of memory space
-

6. How would you define Precision and Recall?

[Hide Answer](#)

Precision and Recall are two commonly used metrics in the field of ML to evaluate the performance of a binary classification model.

Precision measures how many of the items predicted to belong to the positive class are actually positive. It is defined as the ratio of true positives (TP) to the total number of predicted positives.

actual positives.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

7. What are some tools used to discover outliers?

[Hide Answer](#)

Outliers are errors or extreme values that differ from the data in a set. They can impact the accuracy of your result if not detected through the outlier detection tools. Some popular tools to discover outliers include Z-score, Scatter plot, and Box plot.

8. What is the Fourier Transform?

[Hide Answer](#)

This mathematical technique is used to transform a function of time into a function of frequency. It is necessary for ML and deep learning because a convolution in the time domain is a multiplication in the frequency domain.

9. What is Inductive Logic Programming?

[Hide Answer](#)

Inductive logic programming (ILP) is a part of AI that uses logic programming and serves as a uniform representation for examples, background knowledge, and hypotheses. It involves developing algorithms and systems that can automatically learn logic programs from input-output examples, which can then be used to make predictions on new input data.

We can apply it to various domains, including software engineering, natural language processing, and bioinformatics.

Kernel SVM stands for Kernel Support Vector Machine. In SVM, a kernel is a function that aids in problem-solving. They provide shortcuts to help you avoid doing complicated math. The amazing thing about kernel is that it allows us to go to higher dimensions and execute smooth computations. Kernel SVMs can work with a variety of kernel functions, including linear, polynomial, and radial basis function (RBF) kernels, among others.

11. What are some types of clustering?

[Hide Answer](#)

These are some types of clustering:

Fuzzy clustering: Each data point can belong to multiple clusters.

: It divides a set of n observations into k clusters. Used in situations when one doesn't have existing group labels.

Hierarchical clustering: It is an algorithm that groups similar objects into segments called clusters. Here, each cluster differs from the others, and the objects within each cluster are the same.

Density-based clustering: It clusters data based on their density and hence requires uniform density within the cluster. Here density drops n =between clusters.

12. How would you describe reinforcement learning?

[Hide Answer](#)

Reinforcement learning is a type of ML where an agent interacts with an environment and learns to make decisions based on feedback from the environment in the form of rewards or penalties. Unlike supervised and unsupervised learning, the agent is not given labeled data but instead learns from its own experiences through trial and error. The objective of reinforcement

13. What are some methods to improve inference time?

[Hide Answer](#)

Some methods to improve inference time are as follows:

- Cache predictions
- Knowledge distillation
- Reduction of parameters by pruning
- Model size reduction
- Parallel computing

14. What is content-based filtering and collaborative filtering?

[Hide Answer](#)

Content-based filtering recommends items to a user based on their past behavior, preferences, and interests. For example, if a user has shown interest in action movies, then the content-based filtering algorithm will recommend more action movies to that user.

Collaborative filtering recommends items to a user based on the preferences of other users who have similar tastes. For example, if a user has watched several action movies, the collaborative filtering algorithm will recommend other action movies that were also watched by users who have similar movie preferences.

15. What is deductive learning and inductive learning?

[Hide Answer](#)

Deductive learning is the process of using conclusions to form observations, while inductive learning is using observations to draw conclusions.

Inductive learning, on the other hand, is a bottom-up approach where specific observations or data are used to form a general principle or theory. This approach involves observing patterns and regularities in the data, and then using those observations to form a hypothesis or general principle.

We use inductive learning when there is little or no data available to derive conclusions. On the contrary, some theories are available in the deductive learning approach.

16. How would you differentiate data mining from Machine learning?

[Hide Answer](#)

Data mining is the process of using structured data to extract information or find anomalies, correlations, and patterns within large datasets to predict outcomes using machine learning algorithms.

But Machine Learning is the development of algorithms to provide abilities to processors to learn without being programmed.

17. Why is the ROC in a curve important?

[Hide Answer](#)

The ROC curve is important because it is a visual representation of how well a model can distinguish between two classes, and it can be used to compare different models. The area under the curve (AUC) is a measure of how well the model performs, with a higher AUC indicating a better model. Additionally, the shape of the curve can indicate whether a model is biased towards one class or another.

18. Why does overfitting occur in ML?

[Hide Answer](#)

data.

19. What are the functions of unsupervised learning?

[Hide Answer](#)

The following are some functions of unsupervised learning:

- Dimensionality reduction
 - Finding clusters of data
 - Anomaly deduction
-

20. What are some functions of supervised learning?

[Hide Answer](#)

Some functions of supervised learning as follows:

- Regression
 - Feature selection
 - Classification
 - Time series forecasting
-

21. What are the two components of Bayesian logic?

[Hide Answer](#)

Two components of Bayesian logic are as follows:

Prior distribution: This is the information or beliefs that you have about the world before observing any new data. Prior knowledge is typically represented as a probability distribution over possible states of the world.

22. How would you describe a Recommender system?

[Hide Answer](#)

A recommendation engine can be seen as a system to predict users' likes and interests and recommend products that fit their tastes. Data used in this system can be derived from user ratings of movies and songs and search engine history.

23. What functions are used to perform categorical conversion into factor?

[Hide Answer](#)

ML algorithms need numeric inputs. It requires converting categorical values into factors to get the latter. The functions `factor()` and `as.factor()` are used to perform categorical conversion into factor.

24. Are elements stored consecutively in linked lists?

[Hide Answer](#)

No. Elements can be stored anywhere in linked lists. A linked list contains nodes where each node is composed of a data field and a reference (link) to the next node.

25. What is Regularization in ML?

[Hide Answer](#)

Regularization is a technique used to prevent overfitting in ML models. Overfitting occurs when a model is too complex and fits the training data too closely, resulting in poor performance on new, unseen data. Regularization techniques add a penalty term to the model's objective function, which encourages the

which differ in the way they penalize the coefficients.

26. Do you consider decision trees to have any advantages or disadvantages? If yes, what are they?

[Hide Answer](#)

Decision tree algorithms are favored for their interpretability, scalability, and accuracy. They can be used to classify data points quickly and accurately, and are great for dealing with large datasets. They are also simple to understand, making them great for explaining complex models to non-technical stakeholders.

27. What do you understand about the exploding gradient technique problem in the back propagation technique?

[Hide Answer](#)

The exploding gradient problem occurs when large error gradients gather in training, resulting in large changes in the neural network weights. It can result in NaN values due to the values of weights being too large.

28. What do you understand about the Bias-Variance Tradeoff?

[Hide Answer](#)

If your model is simple with only a few parameters, it will have high bias and low variance. With many parameters, it will have low bias and high variance. One needs to find the right balance without overfitting and underfitting the data. This is why there is a tradeoff between bias and variance. Bias-variance tradeoff is important as it helps us know the limitations of our model and how to improve on it.

Apply Now

ADVANCED MACHINE LEARNING INTERVIEW QUESTIONS AND ANSWERS

1. How would you describe an F1 score, and how would you use it?

[Hide Answer](#)

This is an evaluation metric in ML that sums up the predictive performance of a model by combining two competing metrics, namely, Precision and Recall. In binary classification, we assume F1 score to measure a model's accuracy because it is the weighted average of precision and recall scores.

Scores of F1 range from 0 to 1, with 1 being the best score and 0 being the worst. These F1 scores are used in information retrieval to see how well a model recovers results and how the model performs.

2. Explain the difference between the Loss functions and Cost functions.

[Hide Answer](#)

We usually call it the Loss function when calculating loss considering only a single point. However, we call it the Cost function when calculating the sum of errors for multiple data.

In simpler terms, the cost function sums the difference for the training database, while the loss function shows the difference between the actual and predicted values for a single record.

3. How do you handle outlier values?

- Dropping them
 - Marking them as outliers and including them as a feature
 - Transforming the feature to reduce the effect of the outlier
-

4. What is a random forest, and how does it work?

[Hide Answer](#)

Random forest is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is built using a different subset of the training data and a random subset of the features. This randomness helps to prevent overfitting and improve the generalization of the model.

Random forest works in the following steps:

1. Take up a sample size from the training data
2. Start with a particular node
3. Run the algorithm from the start node viz: If the number of observations is smaller than node size, then stop Select random variables Spot the variable that is best at splitting the observations Divide the observations into two nodes Call step 'a' on each of these two nodes

During the process of building each decision tree, the algorithm splits the data at each node based on the feature that provides the best split. This split is determined using a metric such as information gain or Gini impurity.

Once all the decision trees are built, they are used together to make predictions. The idea is that each tree has its strengths and weaknesses, but by combining them, the weaknesses are averaged out, and the strengths are amplified. This ensemble approach is often more accurate than using a single decision tree.

5. What ensemble techniques can be used to aggregate multiple models?

[Hide Answer](#)

Two ensemble techniques that can be used include Bagging and Boosting:

Boosting: This is an iterative ensemble technique where the models are trained sequentially, and each subsequent model is trained on the instances where the previous model performed poorly. The idea behind boosting is to reduce the bias of the model by focusing on the misclassified instances

6. What methods can be used to find the threshold of a classifier?

[Hide Answer](#)

The threshold for a classifier can be derived using precision-recall Curves and ROC Curves. In other instances, a grid search can be used to adjust the threshold for finding the best value.

7. How can you check for the Normality of a dataset?

[Hide Answer](#)

Plots can be used to check for the normality of datasets, and some of these checks are:

- Kolmogorov-Smirnov Test
 - Shapiro-Wilk Test
 - Anderson-Darling Test
 - D'Agostino Skewness Test
-

8. How can you differentiate between a parametric and a non-parametric model?

[Hide Answer](#)

Parametric models have limited parameters and only need to know the model's parameters to predict new data, whereas non-parametric models have unlimited

9. Do you think logistic regression can be used for more than two classes?

[Hide Answer](#)

No. Logistic regression is a binary classifier; thus, it cannot be applied to more than two classes. However, it can be employed in multinomial logistic regression of solving multi-class classification problems.

10. What differences exist between Softmax and Sigmoid functions?

[Hide Answer](#)

The sigmoid function is a mathematical function that maps any input value to a value between 0 and 1 and it is used for binary classification. On the other hand, the softmax function is used for multi-classification, and the sum of the probabilities will be 1.

11. How can overfitting be avoided in ML?

[Hide Answer](#)

Some of the methods that you can use to avoid overfitting in machine learning are:

- Cross validation
 - Custom feature selection
 - Data augmentation
 - Using larger datasets
 - Data simplification
 - Ensembling
 - Regularization
-

It depends on the situation. A false negative is risky in fields like medicine when a virus is scanned and reported absent when it is present. A false positive is risky in situations like spam email detection because it can classify an important email as spam.

13. How would you handle a dataset suffering from high variance?

[Hide Answer](#)

The bagging algorithm can be used to split the data into subgroups sampling replicated from random data. After this split, the random data is utilized to establish rules using a training algorithm. Thereafter, the polling technique is used to combine all the predicted outcomes of the model.

14. Explain Genetic programming as it relates to Evolutionary algorithm.

[Hide Answer](#)

Genetic programming is similar to Evolutionary algorithm. It is a subset of evolutionary algorithms that implement the model's parameter, random mutation, crossover, fitness function, and various layers of evolution to resolve a user-defined task.

Through repeated generations of evaluation and modification, genetic programming can evolve increasingly complex and sophisticated programs that are optimized for the task at hand. This approach can be used for a wide range of applications, from artificial intelligence and machine learning to robotics and optimization problems.

15. What are some classification methods that SVM can handle?

[Hide Answer](#)

- Multilabel classification
-

16. Why do you think instance-based learning algorithm is sometimes referred to as Lazy learning algorithm?

[Hide Answer](#)

Instance-based learning algorithm is sometimes referred to as lazy learning algorithm because it does not require any training data or generalization of the data. Instead, it uses the data the algorithm is given and stores it in memory for use in future predictions. The algorithm does not actively learn from the data, but rather relies on the stored instances to make predictions. Thus, its learning process can be seen as "lazy" as compared to other algorithms that actively learn from the data.

17. Explain the reason for pruning of a decision tree.

[Hide Answer](#)

Pruning is the process of removing unnecessary nodes in a decision tree in order to reduce the complexity and improve the generalization accuracy of the model. It is used to prevent overfitting. Pruning helps in reducing the complexity of the decision trees by removing the nodes that do not have a great impact on the prediction accuracy.

18. How Regularization reduces the cost term?

[Hide Answer](#)

Regularization is a technique used to prevent overfitting by adding a penalty term to the cost function. The penalty term is designed to discourage the model from

[Hide Answer](#)

Converting categorical variables into factors is necessary because many ML algorithms cannot work with categorical variables in their raw form. ML algorithms work with numerical values, and factors are a way of representing categorical variables as numerical values.

20. Do you believe treating a categorical variable as a continuous variable will result in a better predictive model?

[Hide Answer](#)

For a better predictive model, the variable has to be ordinal in nature.

21. Why do we need the confusion matrix?

[Hide Answer](#)

The confusion matrix is usually a summary of predictions on a classification model that enables us to visualize the performance of an algorithm or model. It measures the performance of an algorithm or model.

22. What differentiates Gradient boosting and Random forest?

[Hide Answer](#)

In gradient boosting, the trees are trained sequentially one after another, thereby correcting errors of the previous one.	In Random forest, the decision trees are constructed independently
GB gives better performance when there is unbalanced data	Random forest performs better for multiclass object detection

23. How does a Box-Cox transformation occur?

[Hide Answer](#)

This happens when there is a power transformation of non-normal dependent variables into normal variables. It involves a lambda parameter which, when set to zero, indicates that this transform is equivalent to log-transform.

24. What functions can be used to identify, drop, and replace missing or corrupted values?

[Hide Answer](#)

Usually, missing or corrupted values happen when no data is available for certain variables; this can happen due to incomplete data entry, lost files, etc. Some functions that identify, drop, and replace missing or corrupted values are `IsNull()`, `dropna()`, and `Fillna()` functions in Pandas.

25. How is data divided in cross-validation?

[Hide Answer](#)

Cross validation divides data into three parts which are training, testing, and validation data. Data is further divided into k numbers of sub-parts, and the model is trained on k-1 of those sub-parts and then tested on the remaining fold. This process is repeated k times, with each of the k folds used exactly once for

26. A game exists where you are required to roll two six-faced dice. You stand a chance to win \$50 if the sum of the two dice equals 7. However, you must pay \$4 to play each time you roll both dice. Would you partake in this game?

Note: If you play 6 times, what is the probability of making money from this game?

[Hide Answer](#)

- The first condition states that if the sum of the values on the 2 dice is equal to 7, then you win \$50. But for all the other cases, you must pay \$4.
- To begin with, let's estimate the number of possible cases if we throw the two dice. The total number of cases is $6 \times 6 = 36$.
- In these 36 cases, how many produce a sum of 7?
- Possible combinations that produce a sum of 7 are (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1)..
- This indicates that out of 36 chances, 6 will produce a sum of 7. The ratio now is $6/36 = 1/6$
- So this suggests that we have a chance of winning \$50, once in every 6 games.
- So to answer the question, if I play 6 times, I will win \$50 for one game, whereas for the other 5 games I will have to pay \$4 each, which is \$20 for all five games. Therefore, I will be at a winning advantage if I win \$50 but end up paying \$20. Statistically, I'll be making a profit of \$30 every 6 games.

27. Assuming you have a two-column table of users and their friends coupled with a two-column table of users and their liked pages. Write a SQL query to make recommendations using pages that your friends have liked when it should not recommend pages that you have liked.

[Hide Answer](#)

```
ON f.friend_id = l.user_id  
WHERE l.page_id NOT IN (SELECT page_id FROM like  
WHERE user_id = f.user_id)
```

28. When recommendations like “...people who did this also bought this” comes up on a website. What sort of algorithm propels that?

[Hide Answer](#)

This is a recommendation based on collaborative filtering. Collaborative filtering compares users' activities on a website and recommends similar actions to users with similar interests or behavior.

29. What would you do if the model you have trained has a high variance and low bias?

[Hide Answer](#)

To solve high variance problems, we will use regularization techniques to rectify this or select the higher features in the feature importance graph to train the model.

For low bias, we could use the bagging algorithm to divide the data into sub-parts using randomized sampling. Thereafter, we can create models with these samples using just one algorithm. Finally, we join the model prediction using voting classification or, perhaps, averaging.

Looking for remote developer job at US companies?

Work at Fortune 500 companies and fast-scaling startups from the comfort of your home

[Apply Now](#)