**Interview questions on naïve bayes:**

[**https://analyticsarora.com/11-unique-machine-learning-interview-questions-on-naive-bayes/**](https://analyticsarora.com/11-unique-machine-learning-interview-questions-on-naive-bayes/)

**https://www.mlstack.cafe/blog/naive-bayes-interview-questions**

1. **What mathematical concept Naive Bayes is based on?**

Naive Bayes is based on Bayes theorem in statistics. It calculates probabilities independently for each class based on conditions and without conditions and then predicts outcomes based on that.

#### ****What are the different types of Naive Bayes classifiers?****

Multinomial Naive Bayes  
Bernoulli Naive Bayes  
Gaussian Naive Bayes

#### ****Is Naive Bias a classification algorithm or regression algorithm?****

It is a classification algorithm. Naive Bayes is a supervised learning algorithm but it can also be trained as semi-supervised learning algorithm.

#### ****What are some benefits of Naive Bayes?****

It works better than simple algorithms like [logistic regression](https://analyticsarora.com/10-unique-machine-learning-interview-questions-on-logistic-regression/) etc. It also works well with categorical data and with numerical data as well. Additionally, It is very easy and fast to work with the Naive Bayes classifier. Complex and high dimensional data is well suited for Naive Bayes classifier. It can also be trained using a small labeled dataset with semi-supervised learning.

In other words:

• It performs well with both clean and noisy data.

• Training takes a few samples, but the fundamental assumption is that the training dataset is a genuine representation of the population.

• Obtaining the likelihood of a forecast is simple.

#### ****What are the cons of Naive Bayes classifier?****

Naive Bayes classifiers suffer from “[Zero Frequency](https://www.atoti.io/how-to-solve-the-zero-frequency-problem-in-naive-bayes/)” problem. This happens when a category is not present in the training set. It will give it 0 probability.

Its biggest downside is the consideration of features as independent of each other because in real life it is impossible to get independent features. All features are somehow co-related with each other.

#### ****What are the applications of Naive Bayes?****

#### Naive Bayes classifier is a very powerful technique. It is applied in various classification techniques which are used for real-time prediction. The algorithm is also widely used in NLP tasks like sentiment analysis of text sentences, applying spam filtering, text classification etc. It is also used to make recommendation systems and for collaborative filtering.

#### ****Is Naive Bayes is a discriminative classifier or generative classifier?****

#### Naive Bayes is a [generative classifier](http://www.chioka.in/explain-to-me-generative-classifiers-vs-discriminative-classifiers/#:~:text=A%20generative%20classifier%20tries%20to,learned%20captures%20the%20real%20model.). It learns from the actual distribution of the dataset by performing operations on it. It does not create a decision boundary to classify data.

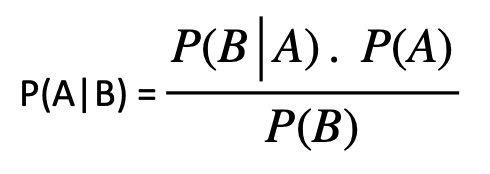
#### ****What is posterior probability and prior probability in Naïve Bayes?****

Probability of event A given event B is true, P(A|B), is called as posterior probability and independent probability of event A, P(A), is called as prior probability i.e.

P(A|B) = posterior probability

P(A) = prior probability

#### ****Define likelihood and evidence in Naive Bayes?****



Here, P(A) & P(B) are independent probabilities of event A and B,

P(A|B) = probability of event A given B is true,

P(B|A) = probability of event B given A is true.

The probability of event B given that event A is true is called likelihood and the independent probability of event B is called evidence.

P(B|A) = likelihood

P(B) = evidence

#### ****While calculating the probability of a given situation, what error can we run into in Naïve Bayes and how can we solve it?****

#### We might encounter the zero division error when the probability for a particular scenario in the numerator is zero. To mitigate this, we can use Laplace Smoothing which basically adds a number to the numerator and another number to the denominator.

#### ****What is the Bernoulli distribution in Naïve Bayes?****

This is a distribution that evaluates a particular outcome as binary. For example, in the Bernoulli Naïve Bayes classifier, given a word, that word can either be in a message or not.

#### ****How does Naïve Bayes treat numerical and categorical values?****

* For the categorical features, we can estimate our probability using a distribution such as multinomial or Bernoulli.
* For the numerical features, we can estimate our probability using a distribution such as Normal or Gaussian.

#### ****What is the best dataset scenario for the Naïve Bayes Classifier?****

If the training data is smaller or if the dataset has fewer number of observations (samples) and a high number of features. Naïve Bayes works well on this data because of its High bias – Low variance trade off.

#### ****What’s the difference between Generative Classifiers and Discriminative Classifiers?****

A **Generative Model** explicitly models each class’s underlying distribution. It ‌learns the joint probability distribution given a probabilistic interaction

i.e. P(message, spam) = P(spam) \* P(message|spam)

Where both P(spam) and P(message|spam) can be estimated from the dataset by computing class frequencies. An example of a generative model would be Naive Bayes.

A **Discriminative Model** models the decision boundary between the classes by ‌learning the conditional probability distribution P(spam|message) from the dataset. An example of a generative model would be Logistic Regression.

Some final thoughts

1. It’s assumption is every features are independent.
2. Scaling is not required.
3. Robust to outliers.
4. Can handle the missing value.
5. For numerical value it use normal or the gaussian distribution.
6. Very fast and easy.
7. Work nice for the small but high numbers of features dataset.

**Interview questions on KNN**

## **1. What is the KNN Algorithm?**

**KNN(K-nearest neighbours)** is a **supervised**learning and **non-parametric** algorithm that can be used to solve both classification and regression problem statements.

## **2. Why is KNN a non-parametric Algorithm?**

The term “**non-parametric”** refers to not making any assumptions on the underlying data distribution. These methods do not have any fixed numbers of parameters in the model.Similarly in KNN, the model parameters grow with the training data by considering each training case as a parameter of the model. So, KNN is a non-parametric algorithm.

## 3. What is “K” in the KNN Algorithm?

K represents the number of nearest neighbours you want to select to predict the class of a given item, which is coming as an unseen dataset for the model.

## 4. Why is the odd value of “K” preferred over even values in the KNN Algorithm?

The odd value of K should be preferred over even values in order to ensure that there are no ties in the voting. If the square root of a number of data points is even, then add or subtract 1 to it to make it odd.

## 5. How does the KNN algorithm make the predictions on the unseen dataset?

## 6. Is Feature Scaling required for the KNN Algorithm? Explain with proper justification.

Yes, feature scaling is required to get the better performance of the KNN algorithm.

**For Example,** Imagine a dataset having n number of instances and N number of features. There is one feature having values ranging between **0 and 1**. Meanwhile, there is also a feature that varies from **-999 to 999**. When these values are substituted in the formula of Euclidean Distance, this will affect the performance by giving higher weightage to variables having a higher magnitude.

## 8. Can the KNN algorithm be used for regression problem statements?

**Yes**, KNN can be used for regression problem statements.

In other words, the KNN algorithm can be applied  when the dependent variable is continuous. For regression problem statements, the predicted value is given by the average of the values of its k nearest neighbours.

## 9. Why is the KNN Algorithm known as Lazy Learner?

When the KNN algorithm gets the training data, it does not learn and make a model, it just stores the data. Instead of finding any discriminative function with the help of the training data, it follows **instance-based learning** and also uses the training data when it actually needs to do some prediction on the unseen datasets.

As a result, KNN does not immediately learn a model rather delays the learning thereby being referred to as Lazy Learner.

## 10. Why is it recommended not to use the KNN Algorithm for large datasets?

**The Problem in processing the data:**

KNN works well with smaller datasets because it is a lazy learner. It needs to store all the data and then make a decision only at run time. It includes the computation of distances for a given point with all other points. So if the dataset is large, there will be a lot of processing which may adversely impact the performance of the algorithm.

**Sensitive to noise:**

Another thing in the context of large datasets is that there is more likely a chance of noise in the dataset which adversely affects the performance of the KNN algorithm since the KNN algorithm is sensitive to the noise present in the dataset.

## 11. How to handle categorical variables in the KNN Algorithm?

To handle the categorical variables we have to create **dummy variables** out of a categorical variable and include them instead of the original categorical variable. Unlike regression, create k dummies instead of (k-1).

**For example,** a categorical variable named **“Degree”** has 5 unique levels or categories. So we will create 5 dummy variables. Each dummy variable has 1 against its degree and else 0.

## 12. How to choose the optimal value of K in the KNN Algorithm?

There is no straightforward method to find the optimal value of K in the KNN algorithm.

You have to play around with different values to choose which value of K should be optimal for my problem statement. Choosing the right value of K is done through a process known as **Hyperparameter Tuning**.

The optimum value of K for KNN is**highly dependent on the data** itself. In different scenarios, the optimum K may vary. It is more or less a hit and trial method.

There is no one proper method of finding the K value in the KNN algorithm. No method is the rule of thumb but you should try the following suggestions:

**1. Square Root Method:** Take the square root of the number of samples in the training dataset and assign it to the K value.

**2. Cross-Validation Method:** We should also take the help of cross-validation to find out the optimal value of K in KNN. Start with the minimum value of k **i.e, K=1**, and run cross-validation, measure the accuracy, and keep repeating till the results become consistent.

## 13. How can you relate KNN Algorithm to the Bias-Variance tradeoff?

* As the value of k increases, the bias will be increases
* As the value of k decreases, the variance will increases
* With the increasing value of K, the boundary becomes smoother

So, there is a tradeoff between **overfitting and underfitting** and you have to maintain a balance while choosing the value of K in KNN. Therefore,**K should not be too small or too large**.

## 14. Which algorithm can be used for value imputation in both categorical and continuous categories of data?

KNN is the only algorithm that can be used for the imputation of both categorical and continuous variables. It can be used as one of many techniques when it comes to handling missing values.

To impute a new sample, we determine the samples in the training set “nearest” to the new sample and averages the nearby points to impute. A **Scikit learn library of Python** provides a quick and convenient way to use this technique.

**Note:** NaNs are omitted while distances are calculated. Hence we replace the missing values with the average value of the neighbours. The missing values will then be replaced by the average value of their “neighbours”.

**15. Explain the statement- “The KNN algorithm does more computation on test time rather than train time”.**

The above-given statement is **absolutely true**.

The basic idea behind the kNN algorithm is to determine a k-long list of samples that are close to a sample that we want to classify. Therefore, the training phase is basically storing a training set, whereas during the prediction stage the algorithm looks for k-neighbours using that stored data. Moreover, KNN does not learn anything from the training dataset as well.

## 16. What are the things which should be kept in our mind while choosing the value of k in the KNN Algorithm?

If K is small, then results might not be reliable because the noise will have a higher influence on the result. If K is large, then there will be a lot of processing to be done which may adversely impact the performance of the algorithm.

**So, the following things must be considered while choosing the value of K:**

* K should be the square root of n (number of data points in the training dataset).
* K should be chosen as the odd so that there are no ties. If the square root is even, then add or subtract 1 to it.

## 18. What are the disadvantages of the KNN Algorithm?

Some of the disadvantages of the KNN algorithm are as follows:

**1. Does not work well with large datasets:** In large datasets, the cost of calculating the distance between the new point and each existing point is huge which decreases the performance of the algorithm.

**2. Does not work well with high dimensions:** KNN algorithms generally do not work well with high dimensional data since, with the increasing number of dimensions, it becomes difficult to calculate the distance for each dimension.

**3. Need feature scaling:** We need to do feature scaling (standardization and normalization) on the dataset before feeding it to the KNN algorithm otherwise it may generate wrong predictions.

**4. Sensitive to Noise and Outliers:** KNN is highly sensitive to the noise present in the dataset and requires manual imputation of the missing values along with outliers removal.

## **Skill test Questions and Answers**

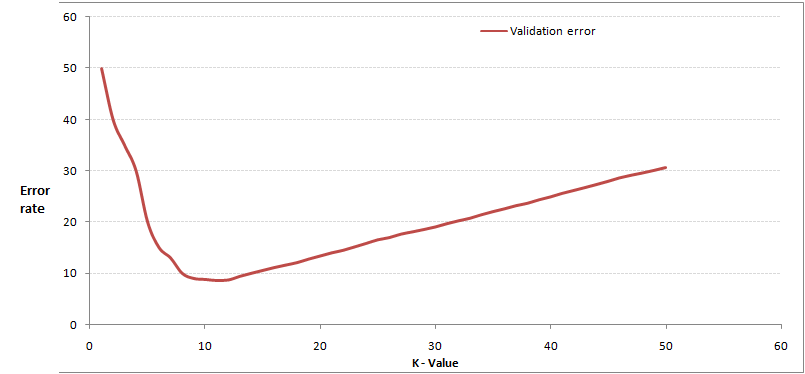
**1) [True or False] k-NN algorithm does more computation on test time rather than train time.**

A) TRUE  
B) FALSE

**Solution: A**

The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

**2) In the image below, which would be the best value for k assuming that the algorithm you are using is k-Nearest Neighbor.**

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2017/08/01195427/Pic21.jpg)

A) 3  
B) 10  
C) 20  
D 50 

**Solution: B**

Validation error is the least when the value of k is 10. So it is best to use this value of k

**3) Which of the following distance metric can not be used in k-NN?**

A) Manhattan  
B) Minkowski  
C) Tanimoto  
D) Jaccard  
E) Mahalanobis  
F) All can be used

**Solution: F**

All of these distance metric can be used as a distance metric for k-NN.

**4) Which of the following option is true about k-NN algorithm?**

A) It can be used for classification  
B) It can be used for regression  
C) It can be used in both classification and regression

**Solution: C**We can also use k-NN for regression problems. In this case the prediction can be based on the mean or the median of the k-most similar instances.

**5) Which of the following statement is true about k-NN algorithm?**

1. k-NN performs much better if all of the data have the same scale
2. k-NN works well with a small number of input variables (p), but struggles when the number of inputs is very large
3. k-NN makes no assumptions about the functional form of the problem being solved

A) 1 and 2  
B) 1 and 3  
C) Only 1  
D) All of the above

**Solution: D**The above mentioned statements are assumptions of kNN algorithm

**6) Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?**

A) K-NN  
B) Linear Regression  
C) Logistic Regression

**Solution: A**

k-NN algorithm can be used for imputing missing value of both categorical and continuous variables.

**7) Which of the following is true about Manhattan distance?**

A) It can be used for continuous variables  
B) It can be used for categorical variables  
C) It can be used for categorical as well as continuous  
D) None of these

**Solution: A**

Manhattan Distance is designed for calculating the distance between real valued features.

**8) Which of the following distance measure do we use in case of categorical variables in k-NN?**

1. Hamming Distance
2. Euclidean Distance
3. Manhattan Distance

A) 1  
B) 2  
C) 3  
D) 1 and 2  
E) 2 and 3  
F) 1,2 and 3

**Solution: A**

Both Euclidean and Manhattan distances are used in case of continuous variables, whereas hamming distance is used in case of categorical variable.

**15) Which of the following will be true about k in k-NN in terms of Bias?**

A) When you increase the k the bias will be increases  
B) When you decrease the k the bias will be increases  
C) Can’t say  
D) None of these

**Solution: A**

large K means simple model, simple model always condider as high bias

**17) The following two distances(Eucludean Distance and Manhattan Distance) have given to you which generally we used in K-NN algorithm. These distance are between two points A(x1,y1) and B(x2,Y2).**

**Your task is to tag the both distance by seeing the following two graphs. Which of the following option is true about below graph ?**

[](https://cdn.analyticsvidhya.com/wp-content/uploads/2017/08/01190245/Pic_17.jpg)

A) Left is Manhattan Distance and right is euclidean Distance  
B) Left is Euclidean Distance and right is Manhattan Distance  
C) Neither left or right are a Manhattan Distance  
D) Neither left or right are a Euclidian Distance**Solution: B**Left is the graphical depiction of how euclidean distance works, whereas right one is of Manhattan distance.

**18) When you find noise in data which of the following option would you consider in k-NN?**

A) I will increase the value of k  
B) I will decrease the value of k  
C) Noise can not be dependent on value of k  
D) None of these

**Solution: A**To be more sure of which classifications you make, you can try increasing the value of k.

**19) In k-NN it is very likely to overfit due to the curse of dimensionality. Which of the following option would you consider to handle such problem?**

1. Dimensionality Reduction
2. Feature selection

A) 1  
B) 2  
C) 1 and 2  
D) None of these

**Solution: C**

In such case you can use either dimensionality reduction algorithm or the feature selection algorithm

**20) Below are two statements given. Which of the following will be true both statements?**

1. k-NN is a memory-based approach is that the classifier immediately adapts as we collect new training data.
2. The computational complexity for classifying new samples grows linearly with the number of samples in the training dataset in the worst-case scenario.

A) 1  
B) 2  
C) 1 and 2  
D) None of these

**Solution: C**

Both are true and self explanatory

**23) A company has build a kNN classifier that gets 100% accuracy on training data. When they deployed this model on client side it has been found that the model is not at all accurate. Which of the following thing might gone wrong?**

**Note: Model has successfully deployed and no technical issues are found at client side except the model performance**  
A) It is probably a overfitted model  
B) It is probably a underfitted model  
C) Can’t say  
D) None of these**Solution: A**In an overfitted module, it seems to be performing well on training data, but it is not generalized enough to give the same results on a new data.

**24) You have given the following 2 statements, find which of these option is/are true in case of k-NN?**

1. In case of very large value of k, we may include points from other classes into the neighborhood.
2. In case of too small value of k the algorithm is very sensitive to noise

A) 1  
B) 2  
C) 1 and 2  
D) None of these

**Solution: C**

Both the options are true and are self explanatory.

**25) Which of the following statements is true for k-NN classifiers?**

A) The classification accuracy is better with larger values of k  
B) The decision boundary is smoother with smaller values of k  
C) The decision boundary is linear  
D) k-NN does not require an explicit training step

**Solution: D** Option A: This is not always true. You have to ensure that the value of k is not too high or not too low.

Option B: This statement is not true. The decision boundary can be a bit jagged

Option C: Same as option B

Option D: This statement is true

**26) True-False: It is possible to construct a 2-NN classifier by using the 1-NN classifier?**

A) TRUE  
B) FALSE**Solution: A**You can implement a 2-NN classifier by ensembling 1-NN classifiers

**27) In k-NN what will happen when you increase/decrease the value of k?**

A) The boundary becomes smoother with increasing value of K  
B) The boundary becomes smoother with decreasing value of K  
C) Smoothness of boundary doesn’t dependent on value of K  
D) None of these**Solution: A**The decision boundary would become smoother by increasing the value of K

**28) Following are the two statements given for k-NN algorthm, which of the statement(s)**

**is/are true?**

1. We can choose optimal value of k with the help of cross validation
2. Euclidean distance treats each feature as equally important

A) 1  
B) 2  
C) 1 and 2  
D) None of these

**Solution: C**

**Context 29-30:**

Suppose, you have trained a k-NN model and now you want to get the prediction on test data. Before getting the prediction suppose you want to calculate the time taken by k-NN for predicting the class for test data.  
Note: Calculating the distance between 2 observation will take D time.

**29) What would be the time taken by 1-NN if there are N(Very large) observations in test data?**

A) N\*D  
B) N\*D\*2  
C) (N\*D)/2  
D) None of these**Solution: A**The value of N is very large, so option A is correct

**30) What would be the relation between the time taken by 1-NN,2-NN,3-NN.**

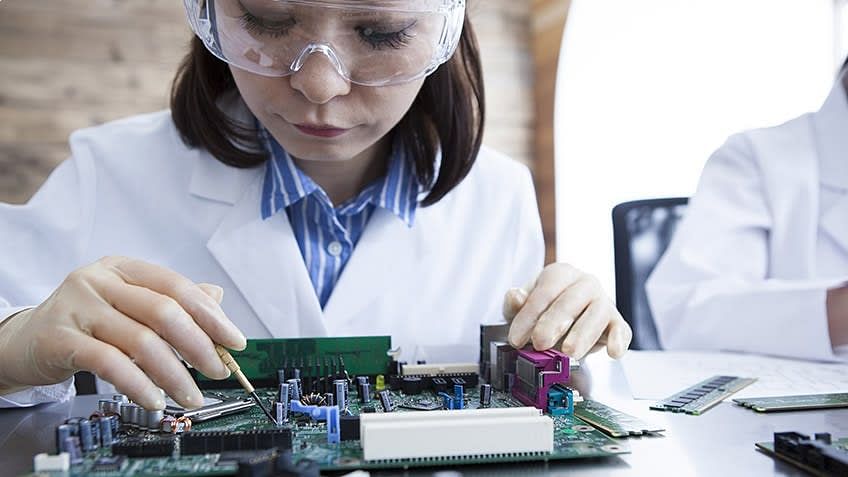
A) 1-NN >2-NN >3-NN  
B) 1-NN < 2-NN < 3-NN  
C) 1-NN ~ 2-NN ~ 3-NN  
D) None of these

**Solution: C** The training time for any value of k in kNN algorithm is the same.

# **Top 45 Machine Learning Interview Questions Answered for 2022**

Lesson 31 of 32[By Eshna Verma](https://www.simplilearn.com/authors/eshna-verma)

Last updated on Jul 18, 2022350375



[Previous](https://www.simplilearn.com/tutorials/machine-learning-tutorial/how-to-become-a-machine-learning-engineer)[Next](https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-quantum-computing)

## **Table of Contents**

[Top Machine Learning Interview Questions](https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-interview-questions#top_machine_learning_interview_questions)

[Become Part of the Machine Learning Talent Pool](https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-interview-questions#become_part_of_the_machine_learning_talent_pool)

Companies are striving to make information and services more accessible to people by adopting new-age technologies like [artificial intelligence](https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-artificial-intelligence) (AI) and machine learning. One can witness the growing adoption of these technologies in industrial sectors like banking, finance, retail, manufacturing, healthcare, and more.

[Data scientists](https://www.simplilearn.com/tutorials/data-science-tutorial/how-to-become-a-data-scientist), [artificial intelligence engineers,](https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/how-to-become-an-ai-engineer) [machine learning engineers](https://www.simplilearn.com/tutorials/machine-learning-tutorial/how-to-become-a-machine-learning-engineer), and [data analysts](https://www.simplilearn.com/what-does-a-data-analyst-do-article) are some of the in-demand organizational roles that are embracing AI. If you aspire to apply for these types of jobs, it is crucial to know the kind of machine learning interview questions that recruiters and hiring managers may ask.

This article takes you through some of the machine learning interview questions and answers, that you’re likely to encounter on your way to achieving your dream job.

## **Top Machine Learning Interview Questions**

Let's start with some commonly asked machine learning interview questions and answers.

### **1. What Are the Different Types of Machine Learning?**

There are three types of machine learning:

#### Supervised Learning

In [supervised machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-and-unsupervised-learning), a model makes predictions or decisions based on past or labeled data. Labeled data refers to sets of data that are given tags or labels, and thus made more meaningful.

Supervised Learning

#### Unsupervised Learning

In unsupervised learning, we don't have labeled data. A model can identify patterns, anomalies, and relationships in the input data.

Unsupervised Learning

#### Reinforcement Learning

Using [reinforcement learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/reinforcement-learning), the model can learn based on the rewards it received for its previous action.

Reinforcement Learning

Consider an environment where an agent is working. The agent is given a target to achieve. Every time the agent takes some action toward the target, it is given positive feedback. And, if the action taken is going away from the goal, the agent is given negative feedback.

### **2. What is Overfitting, and How Can You Avoid It?**

The [Overfitting](https://www.simplilearn.com/tutorials/machine-learning-tutorial/overfitting-and-underfitting) is a situation that occurs when a model learns the training set too well, taking up random fluctuations in the training data as concepts. These impact the model’s ability to generalize and don’t apply to new data.

When a model is given the training data, it shows 100 percent accuracy—technically a slight loss. But, when we use the test data, there may be an error and low efficiency. This condition is known as overfitting.

There are multiple ways of avoiding overfitting, such as:

* Regularization. It involves a cost term for the features involved with the objective function
* Making a simple model. With lesser variables and parameters, the variance can be reduced
* Cross-validation methods like k-folds can also be used
* If some model parameters are likely to cause overfitting, techniques for regularization like LASSO can be used that penalize these parameters

#### FREE Machine Learning Certification Course

To become a Machine Learning Engineer[EXPLORE COURSE](https://www.simplilearn.com/learn-machine-learning-basics-skillup?utm_source=frs&utm_medium=skillup-course-banner&utm_campaign=frs-tutorial-skillup-course-promotion)

### **3. What is ‘training Set’ and ‘test Set’ in a Machine Learning Model? How Much Data Will You Allocate for Your Training, Validation, and Test Sets?**

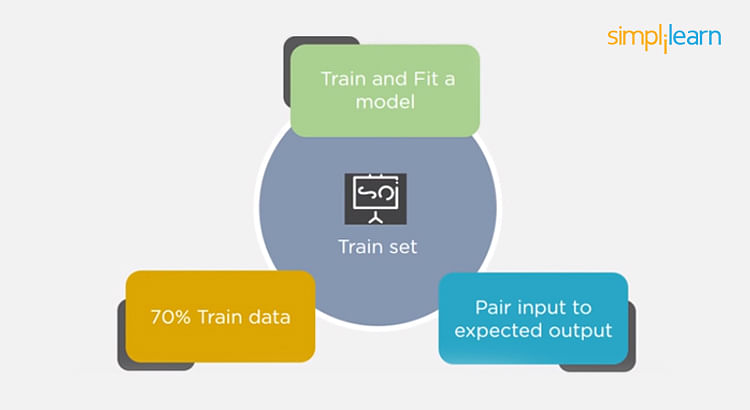
There is a three-step process followed to create a model:

1. Train the model
2. Test the model
3. Deploy the model

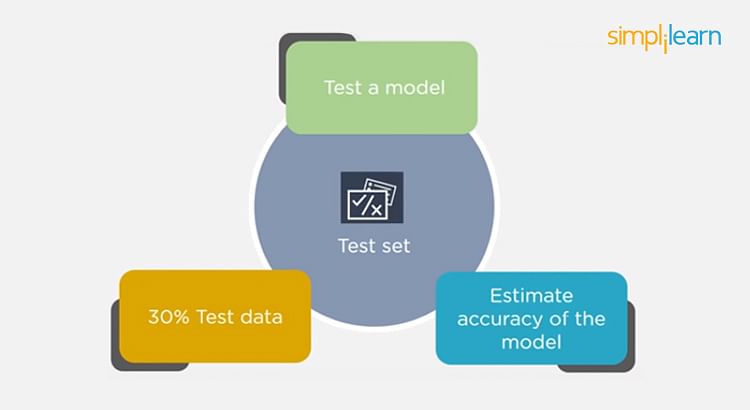
|  |  |
| --- | --- |
| Training Set | Test Set |
| * The training set is examples given to the model to analyze and learn * 70% of the total data is typically taken as the training dataset * This is labeled data used to train the model | * The test set is used to test the accuracy of the hypothesis generated by the model * Remaining 30% is taken as testing dataset * We test without labeled data and then verify results with labels |

Consider a case where you have labeled data for 1,000 records. One way to train the model is to expose all 1,000 records during the training process. Then you take a small set of the same data to test the model, which would give good results in this case.

But, this is not an accurate way of testing. So, we set aside a portion of that data called the ‘test set’ before starting the training process. The remaining data is called the ‘training set’ that we use for training the model. The training set passes through the model multiple times until the accuracy is high, and errors are minimized.



Now, we pass the test data to check if the model can accurately predict the values and determine if training is effective. If you get errors, you either need to change your model or retrain it with more data.



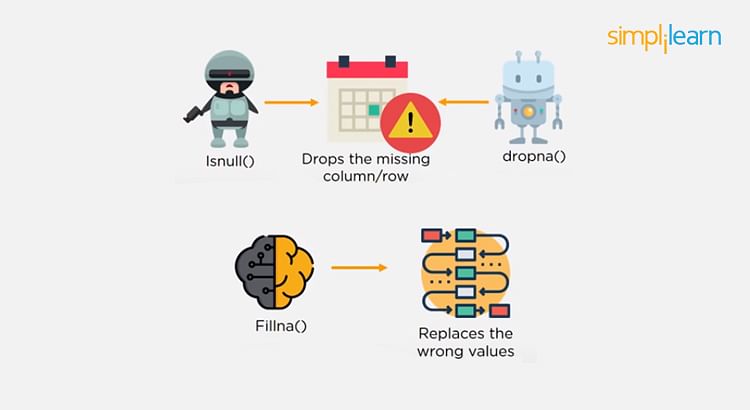
Regarding the question of how to split the data into a training set and test set, there is no fixed rule, and the ratio can vary based on individual preferences.

### **4. How Do You Handle Missing or Corrupted Data in a Dataset?**

One of the easiest ways to handle missing or corrupted data is to drop those rows or columns or replace them entirely with some other value.

There are two useful methods in Pandas:

* IsNull() and dropna() will help to find the columns/rows with missing data and drop them
* Fillna() will replace the wrong values with a placeholder value



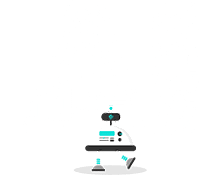
### **5. How Can You Choose a Classifier Based on a Training Set Data Size?**

When the training set is small, a model that has a right bias and low variance seems to work better because they are less likely to overfit.

For example, [Naive Bayes](https://www.simplilearn.com/tutorials/machine-learning-tutorial/naive-bayes-classifier) works best when the training set is large. Models with low bias and high variance tend to perform better as they work fine with complex relationships.

#### Artificial Intelligence Engineer

Your Gateway to Becoming a Successful AI Expert[VIEW COURSE](https://www.simplilearn.com/artificial-intelligence-masters-program-training-course?source=GhPreviewCTABanner)



### **6. Explain the Confusion Matrix with Respect to Machine Learning Algorithms.**

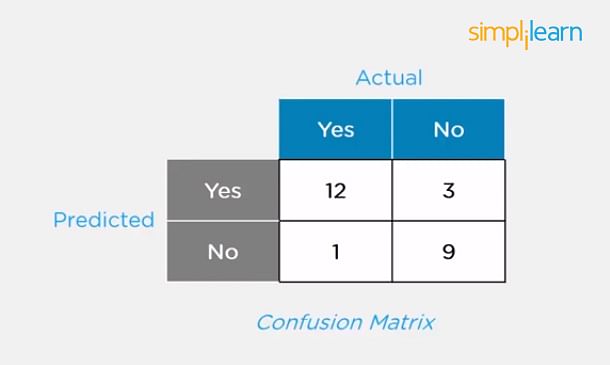
A [confusion matrix](https://www.simplilearn.com/tutorials/machine-learning-tutorial/confusion-matrix-machine-learning) (or error matrix) is a specific table that is used to measure the performance of an algorithm. It is mostly used in supervised learning; in unsupervised learning, it’s called the matching matrix.

The confusion matrix has two parameters:

* Actual
* Predicted

It also has identical sets of features in both of these dimensions.

Consider a confusion matrix (binary matrix) shown below:



Here,

For actual values:

Total Yes = 12+1 = 13

Total No = 3+9 = 12

Similarly, for predicted values:

Total Yes = 12+3 = 15

Total No = 1+9 = 10

For a model to be accurate, the values across the diagonals should be high. The total sum of all the values in the matrix equals the total observations in the test data set.

For the above matrix, total observations = 12+3+1+9 = 25

Now, accuracy = sum of the values across the diagonal/total dataset

= (12+9) / 25

= 21 / 25

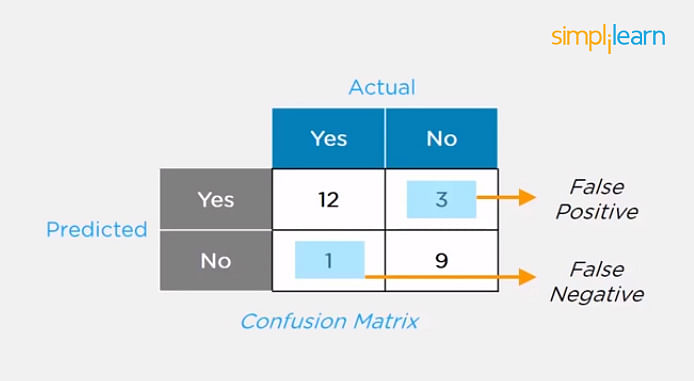
= 84%

### **7. What Is a False Positive and False Negative and How Are They Significant?**

False positives are those cases that wrongly get classified as True but are False.

False negatives are those cases that wrongly get classified as False but are True.

In the term ‘False Positive,’ the word ‘Positive’ refers to the ‘Yes’ row of the predicted value in the confusion matrix. The complete term indicates that the system has predicted it as a positive, but the actual value is negative.



So, looking at the confusion matrix, we get:

False-positive = 3

True positive = 12

Similarly, in the term ‘False Negative,’ the word ‘Negative’ refers to the ‘No’ row of the predicted value in the confusion matrix. And the complete term indicates that the system has predicted it as negative, but the actual value is positive.

So, looking at the confusion matrix, we get:

False Negative = 1

True Negative = 9

### **8. What Are the Three Stages of Building a Model in Machine Learning?**

The three stages of building a[machine learning model](https://www.simplilearn.com/machine-learning-models-article) are:

#### Model Building

Choose a suitable algorithm for the model and train it according to the requirement

#### Model Testing

Check the accuracy of the model through the test data

#### Applying the Model

Make the required changes after testing and use the final model for real-time projects

Here, it’s important to remember that once in a while, the model needs to be checked to make sure it’s working correctly. It should be modified to make sure that it is up-to-date.

#### Free Course: Introduction to ML with R

Master Machine learning with R Basics[ENROLL NOW](https://www.simplilearn.com/free-course-on-machine-learning-with-r-skillup?utm_source=frs&utm_medium=skillup-course-banner&utm_campaign=frs-tutorial-skillup-course-promotion)



### **9. What is Deep Learning?**

The [Deep learning](https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-deep-learning) is a subset of machine learning that involves systems that think and learn like humans using artificial neural networks. The term ‘deep’ comes from the fact that you can have several layers of neural networks.

One of the primary [differences between machine learning and deep learning](https://www.simplilearn.com/machine-learning-vs-deep-learning-major-differences-you-need-to-know-article) is that feature engineering is done manually in machine learning. In the case of deep learning, the model consisting of neural networks will automatically determine which features to use (and which not to use).

This is a commonly asked question asked in both Machine Learning Interviews as well as [Deep Learning Interview Questions](https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-interview-questions)

### **10. What Are the Differences Between Machine Learning and Deep Learning?**

|  |  |
| --- | --- |
| Machine Learning | Deep Learning |
| * Enables machines to take decisions on their own, based on past data * It needs only a small amount of data for training * Works well on the low-end system, so you don't need large machines * Most features need to be identified in advance and manually coded * The problem is divided into two parts and solved individually and then combined | * Enables machines to take decisions with the help of artificial neural networks * It needs a large amount of training data * Needs high-end machines because it requires a lot of computing power * The machine learns the features from the data it is provided * The problem is solved in an end-to-end manner |

### **11. What Are the Applications of Supervised Machine Learning in Modern Businesses?**

Applications of supervised machine learning include:

#### Email Spam Detection

Here we train the model using historical data that consists of emails categorized as spam or not spam. This labeled information is fed as input to the model.

#### Healthcare Diagnosis

By providing images regarding a disease, a model can be trained to detect if a person is suffering from the disease or not.

#### Sentiment Analysis

This refers to the process of using algorithms to mine documents and determine whether they’re positive, neutral, or negative in sentiment.

#### Fraud Detection

By training the model to identify suspicious patterns, we can detect instances of possible fraud.

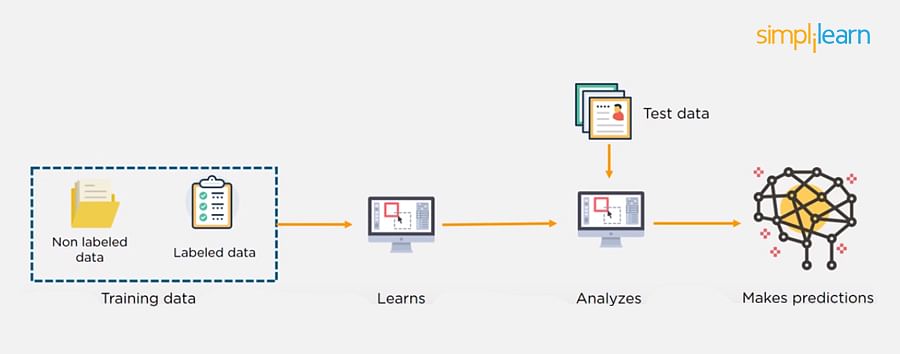
Related Interview Questions and Answers

[AI](https://www.simplilearn.com/artificial-intelligence-ai-interview-questions-and-answers-article?utm_source=table&utm_id=relatedQA) | [Data Science](https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-interview-questions)

### **12. What is Semi-supervised Machine Learning?**

Supervised learning uses data that is completely labeled, whereas unsupervised learning uses no training data.

In the case of semi-supervised learning, the training data contains a small amount of labeled data and a large amount of unlabeled data.

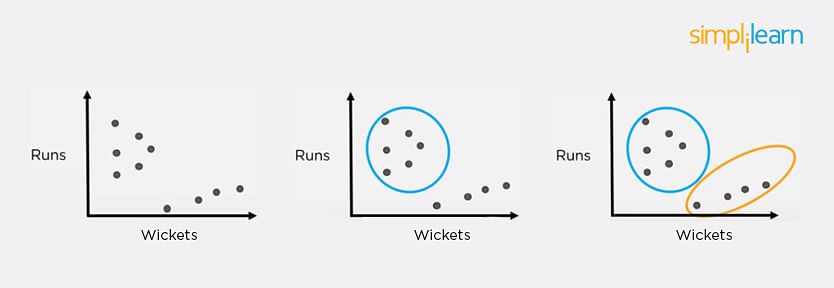


### **13. What Are Unsupervised Machine Learning Techniques?**

There are two techniques used in unsupervised learning: clustering and association.

#### Clustering

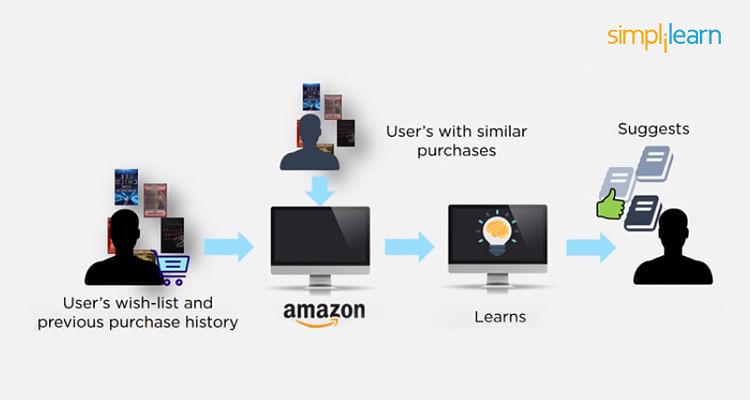
Clustering problems involve data to be divided into subsets. These subsets, also called clusters, contain data that are similar to each other. Different clusters reveal different details about the objects, unlike classification or regression.



#### Association

In an association problem, we identify patterns of associations between different variables or items.

For example, an e-commerce website can suggest other items for you to buy, based on the prior purchases that you have made, spending habits, items in your wishlist, other customers’ purchase habits, and so on.



### **14. What is the Difference Between Supervised and Unsupervised Machine Learning?**

* Supervised learning - This model learns from the labeled data and makes a future prediction as output
* Unsupervised learning - This model uses unlabeled input data and allows the algorithm to act on that information without guidance.

#### Free Course: Programming Fundamentals

Learn the Basics of Programming[ENROLL NOW](https://www.simplilearn.com/learn-programming-basics-free-course-skillup?utm_source=frs&utm_medium=skillup-course-banner&utm_campaign=frs-tutorial-skillup-course-promotion)



### **15. What is the Difference Between Inductive Machine Learning and Deductive Machine Learning?**

|  |  |
| --- | --- |
| Inductive Learning | Deductive Learning |
| * It observes instances based on defined principles to draw a conclusion * Example: Explaining to a child to keep away from the fire by showing a video where fire causes damage | * It concludes experiences * Example: Allow the child to play with fire. If he or she gets burned, they will learn that it is dangerous and will refrain from making the same mistake again |

### **16. Compare K-means and KNN Algorithms.**

|  |  |
| --- | --- |
| K-means | KNN |
| * [K-Means](https://www.simplilearn.com/tutorials/machine-learning-tutorial/k-means-clustering-algorithm) is unsupervised * K-Means is a clustering algorithm * The points in each cluster are similar to each other, and each cluster is different from its neighboring clusters | * [KNN](https://www.simplilearn.com/tutorials/machine-learning-tutorial/knn-in-python) is supervised in nature * KNN is a classification algorithm * It classifies an unlabeled observation based on its K (can be any number) surrounding neighbors |

### **17. What Is ‘naive’ in the Naive Bayes Classifier?**

The classifier is called ‘naive’ because it makes assumptions that may or may not turn out to be correct.

The algorithm assumes that the presence of one feature of a class is not related to the presence of any other feature (absolute independence of features), given the class variable.

For instance, a fruit may be considered to be a cherry if it is red in color and round in shape, regardless of other features. This assumption may or may not be right (as an apple also matches the description).

### **18. Explain How a System Can Play a Game of Chess Using Reinforcement Learning.**

Reinforcement learning has an environment and an agent. The agent performs some actions to achieve a specific goal. Every time the agent performs a task that is taking it towards the goal, it is rewarded. And, every time it takes a step that goes against that goal or in the reverse direction, it is penalized.

Earlier, chess programs had to determine the best moves after much research on numerous factors. Building a machine designed to play such games would require many rules to be specified.

With reinforced learning, we don’t have to deal with this problem as the learning agent learns by playing the game. It will make a move (decision), check if it’s the right move (feedback), and keep the outcomes in memory for the next step it takes (learning). There is a reward for every correct decision the system takes and punishment for the wrong one.

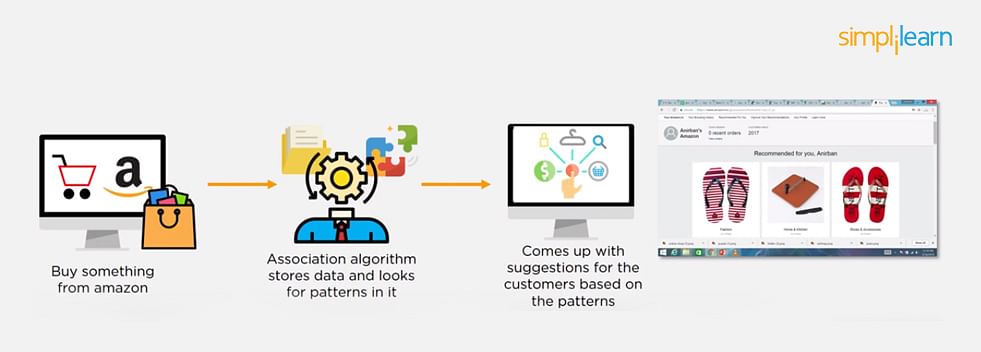
### **19. How Will You Know Which Machine Learning Algorithm to Choose for Your Classification Problem?**

While there is no fixed rule to choose an algorithm for a classification problem, you can follow these guidelines:

* If accuracy is a concern, test different algorithms and cross-validate them
* If the training dataset is small, use models that have low variance and high bias
* If the training dataset is large, use models that have high variance and little bias

### **20. How is Amazon Able to Recommend Other Things to Buy? How Does the Recommendation Engine Work?**

Once a user buys something from Amazon, Amazon stores that purchase data for future reference and finds products that are most likely also to be bought, it is possible because of the Association algorithm, which can identify patterns in a given dataset.



#### Free Course: Machine Learning Algorithms

Learn the Basics of Machine Learning Algorithms[ENROLL NOW](https://www.simplilearn.com/learn-machine-learning-algorithms-free-course-skillup?utm_source=frs&utm_medium=skillup-course-banner&utm_campaign=frs-skillup-course-promotion)



### **21. When Will You Use Classification over Regression?**

Classification is used when your target is categorical, while regression is used when your target variable is continuous. Both classification and regression belong to the category of supervised [machine learning algorithms](https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article).

Examples of classification problems include:

* Predicting yes or no
* Estimating gender
* Breed of an animal
* Type of color

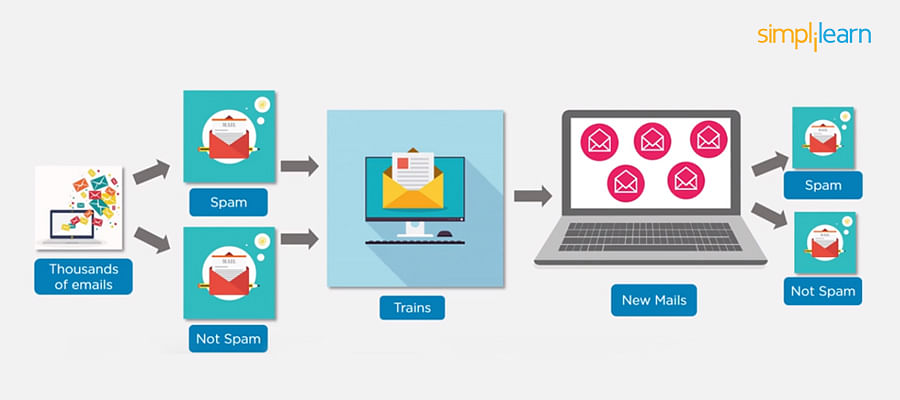
Examples of regression problems include:

* Estimating sales and price of a product
* Predicting the score of a team
* Predicting the amount of rainfall

### **22. How Do You Design an Email Spam Filter?**

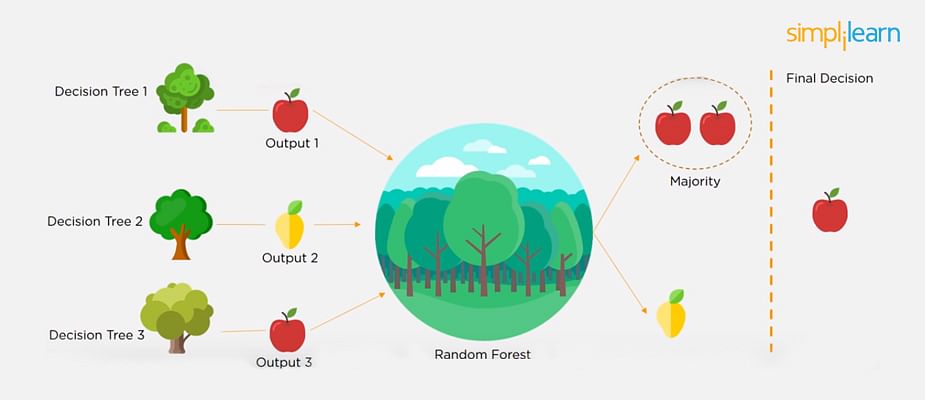
Building a spam filter involves the following process:

* The email spam filter will be fed with thousands of emails
* Each of these emails already has a label: ‘spam’ or ‘not spam.’
* The supervised machine learning algorithm will then determine which type of emails are being marked as spam based on spam words like the lottery, free offer, no money, full refund, etc.
* The next time an email is about to hit your inbox, the spam filter will use statistical analysis and algorithms like [Decision Trees](https://www.simplilearn.com/the-power-of-decision-trees-in-machine-learning-article) and [SVM](https://www.simplilearn.com/tutorials/data-science-tutorial/svm-in-r) to determine how likely the email is spam
* If the likelihood is high, it will label it as spam, and the email won’t hit your inbox
* Based on the accuracy of each model, we will use the algorithm with the highest accuracy after testing all the models



### **23. What is a Random Forest?**

A [‘random forest](https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm)’ is a supervised machine learning algorithm that is generally used for classification problems. It operates by constructing multiple decision trees during the training phase. The random forest chooses the decision of the majority of the trees as the final decision.

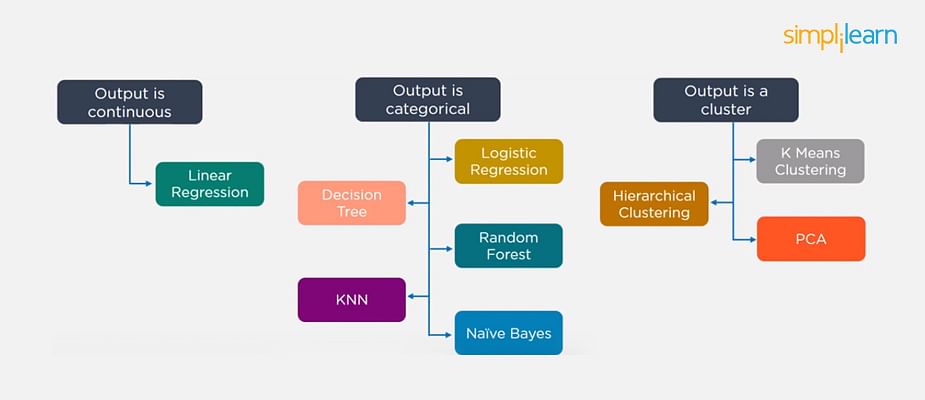


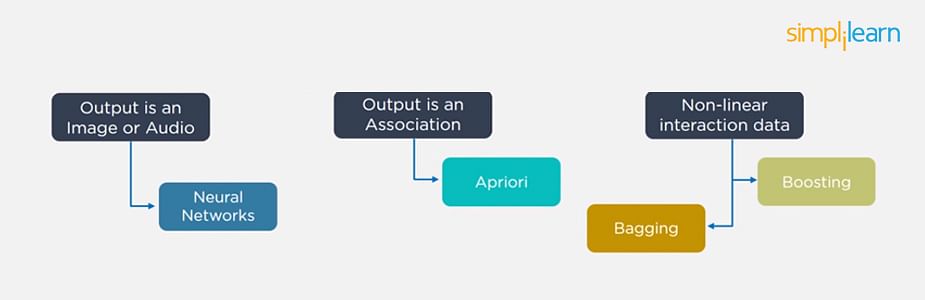
### **24. Considering a Long List of Machine Learning Algorithms, given a Data Set, How Do You Decide Which One to Use?**

There is no master algorithm for all situations. Choosing an algorithm depends on the following questions:

* How much data do you have, and is it continuous or categorical?
* Is the problem related to classification, association, clustering, or regression?
* Predefined variables (labeled), unlabeled, or mix?
* What is the goal?

Based on the above questions, the following algorithms can be used:





### **25. What is Bias and Variance in a Machine Learning Model?**

#### Bias

Bias in a machine learning model occurs when the predicted values are further from the actual values. Low bias indicates a model where the prediction values are very close to the actual ones.

Underfitting: High bias can cause an algorithm to miss the relevant relations between features and target outputs.

#### Variance

Variance refers to the amount the target model will change when trained with different training data. For a good model, the variance should be minimized.

Overfitting: High variance can cause an algorithm to model the random noise in the training data rather than the intended outputs.

### **26. What is the Trade-off Between Bias and Variance?**

The [bias-variance](https://www.simplilearn.com/tutorials/machine-learning-tutorial/bias-and-variance) decomposition essentially decomposes the learning error from any algorithm by adding the bias, variance, and a bit of irreducible error due to noise in the underlying dataset.

Necessarily, if you make the model more complex and add more variables, you’ll lose bias but gain variance. To get the optimally-reduced amount of error, you’ll have to trade off bias and variance. Neither high bias nor high variance is desired.

High bias and low variance algorithms train models that are consistent, but inaccurate on average.

High variance and low bias algorithms train models that are accurate but inconsistent.

### **27. Define Precision and Recall.**

#### Precision

Precision is the ratio of several events you can correctly recall to the total number of events you recall (mix of correct and wrong recalls).

Precision = (True Positive) / (True Positive + False Positive)

#### Recall

A recall is the ratio of the number of events you can recall the number of total events.

Recall = (True Positive) / (True Positive + False Negative)

### **28. What is a Decision Tree Classification?**

A [decision tree builds classification](https://www.simplilearn.com/the-power-of-decision-trees-in-machine-learning-article) (or regression) models as a tree structure, with datasets broken up into ever-smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

### **29. What is Pruning in Decision Trees, and How Is It Done?**

Pruning is a [technique in machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning) that reduces the size of decision trees. It reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.

Pruning can occur in:

* Top-down fashion. It will traverse nodes and trim subtrees starting at the root
* Bottom-up fashion. It will begin at the leaf nodes

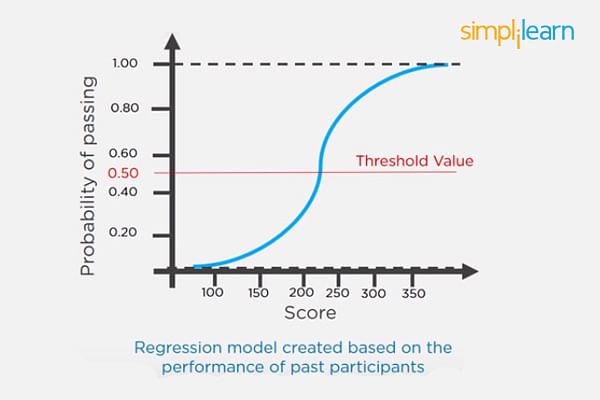
There is a popular pruning algorithm called reduced error pruning, in which:

* Starting at the leaves, each node is replaced with its most popular class
* If the prediction accuracy is not affected, the change is kept
* There is an advantage of simplicity and speed

### **30. Briefly Explain Logistic Regression.**

[Logistic regression](https://www.simplilearn.com/tutorials/machine-learning-tutorial/logistic-regression-in-python) is a classification algorithm used to predict a binary outcome for a given set of independent variables.

The output of logistic regression is either a 0 or 1 with a threshold value of generally 0.5. Any value above 0.5 is considered as 1, and any point below 0.5 is considered as 0.



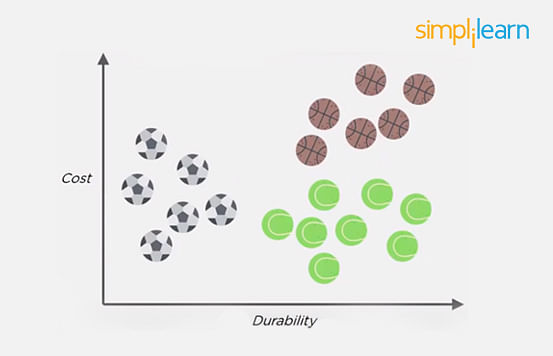
### **31. Explain the K Nearest Neighbor Algorithm.**

[K nearest neighbor algorithm](https://www.simplilearn.com/tutorials/machine-learning-tutorial/knn-in-python) is a classification algorithm that works in a way that a new data point is assigned to a neighboring group to which it is most similar.

In K nearest neighbors, K can be an integer greater than 1. So, for every new data point, we want to classify, we compute to which neighboring group it is closest.

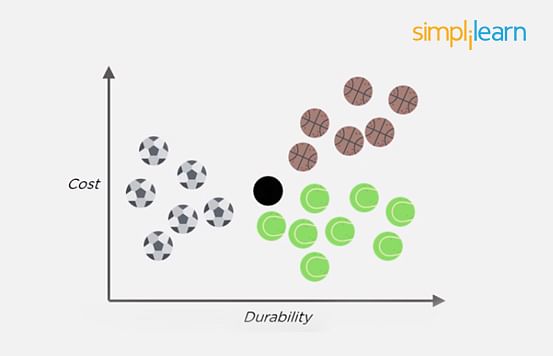
Let us classify an object using the following example. Consider there are three clusters:

* Football
* Basketball
* Tennis ball



Let the new data point to be classified is a black ball. We use KNN to classify it. Assume K = 5 (initially).

Next, we find the K (five) nearest data points, as shown.



Observe that all five selected points do not belong to the same cluster. There are three tennis balls and one each of basketball and football.

When multiple classes are involved, we prefer the majority. Here the majority is with the tennis ball, so the new data point is assigned to this cluster.

### **32. What is a Recommendation System?**

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system: It’s an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

### **33. What is Kernel SVM?**

Kernel SVM is the abbreviated version of the kernel support vector machine. Kernel methods are a class of algorithms for pattern analysis, and the most common one is the kernel SVM.

### **34. What Are Some Methods of Reducing Dimensionality?**

You can reduce dimensionality by combining features with feature engineering, removing collinear features, or using algorithmic dimensionality reduction.

Now that you have gone through these machine learning interview questions, you must have got an idea of your strengths and weaknesses in this domain.

### **35. What is Principal Component Analysis?**

Principal Component Analysis or PCA is a multivariate statistical technique that is used for analyzing quantitative data. The objective of PCA is to reduce higher dimensional data to lower dimensions, remove noise, and extract crucial information such as features and attributes from large amounts of data.

### **36. What do you understand by the F1 score?**

The F1 score is a metric that combines both Precision and Recall. It is also the weighted average of precision and recall.

The F1 score can be calculated using the below formula:

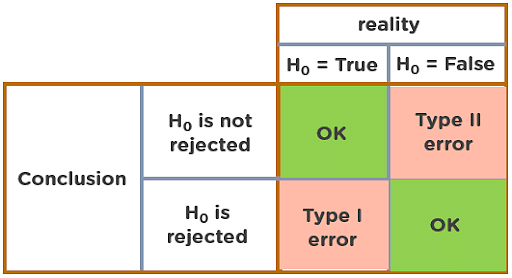
F1 = 2 \* (P \* R) / (P + R)

The F1 score is one when both Precision and Recall scores are one.

### **37. What do you understand by Type I vs Type II error?**

Type I Error: Type I error occurs when the null hypothesis is true and we reject it.

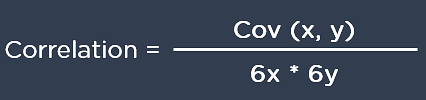
Type II Error: Type II error occurs when the null hypothesis is false and we accept it.



### **38. Explain Correlation and Covariance?**

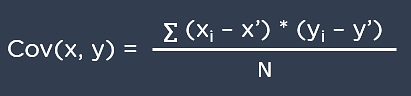
Correlation: Correlation tells us how strongly two random variables are related to each other. It takes values between -1 to +1.

Formula to calculate Correlation:



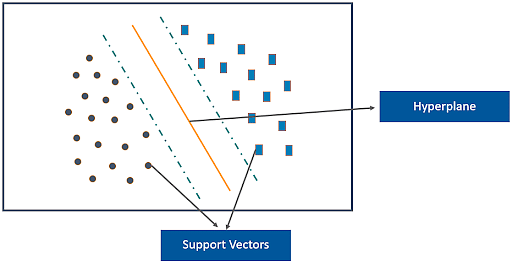
Covariance: Covariance tells us the direction of the linear relationship between two random variables. It can take any value between - ∞ and + ∞.

Formula to calculate Covariance:



### **39. What are Support Vectors in SVM?**

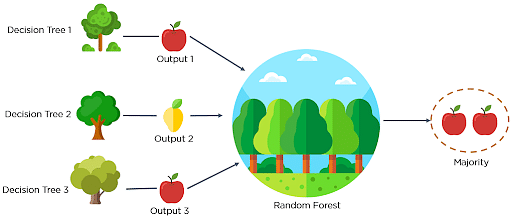
Support Vectors are data points that are nearest to the hyperplane. It influences the position and orientation of the hyperplane. Removing the support vectors will alter the position of the hyperplane. The support vectors help us build our support vector machine model.



### **40. What is Ensemble learning?**

Ensemble learning is a combination of the results obtained from multiple machine learning models to increase the accuracy for improved decision-making.

Example: A Random Forest with 100 trees can provide much better results than using just one decision tree.



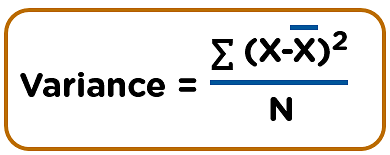
### **41. What is Cross-Validation?**

Cross-Validation in Machine Learning is a statistical resampling technique that uses different parts of the dataset to train and test a machine learning algorithm on different iterations. The aim of cross-validation is to test the model’s ability to predict a new set of data that was not used to train the model. Cross-validation avoids the overfitting of data.

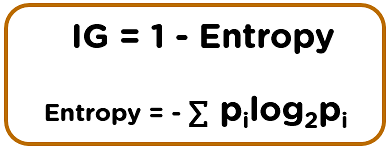
K-Fold Cross Validation is the most popular resampling technique that divides the whole dataset into K sets of equal sizes.

### **42. What are the different methods to split a tree in a decision tree algorithm?**

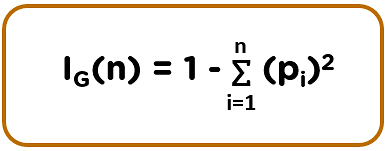
Variance: Splitting the nodes of a decision tree using the variance is done when the target variable is continuous.



Information Gain: Splitting the nodes of a decision tree using Information Gain is preferred when the target variable is categorical.



Gini Impurity: Splitting the nodes of a decision tree using Gini Impurity is followed when the target variable is categorical.



### **43. How does the Support Vector Machine algorithm handle self-learning?**

The [SVM algorithm](https://www.simplilearn.com/tutorials/data-science-tutorial/svm-in-r) has a learning rate and expansion rate which takes care of self-learning. The learning rate compensates or penalizes the hyperplanes for making all the incorrect moves while the expansion rate handles finding the maximum separation area between different classes.

### **44. What are the assumptions you need to take before starting with linear regression?**

There are primarily 5 assumptions for a Linear Regression model:

* Multivariate normality
* No auto-correlation
* Homoscedasticity
* Linear relationship
* No or little multicollinearity

### **45. What is the difference between Lasso and Ridge regression?**

Lasso(also known as L1) and Ridge(also known as L2) regression are two popular regularization techniques that are used to avoid overfitting of data. These methods are used to penalize the coefficients to find the optimum solution and reduce complexity. The Lasso regression works by penalizing the sum of the absolute values of the coefficients. In Ridge or L2 regression, the penalty function is determined by the sum of the squares of the coefficients.