Assignment Answer

Answer 1.1

Anwer

Different between Static varaibles and Dynamic variables.

Static Variables	Dynamic Variables
1) Data types are checked at compile time.	1) Data type are checked at runtime.
2) Catch type error at compile time.	2) Type error cannot be detected
3) Better performance.	3) More Flexibility
4) Required eplicit declaration of data type.	4) Does not required explicit dec-laration of data type.
5) Varibles cannot change types.	5) Variables cannot changed the types.

Answer

Dictionary Method

pop() Removea and returns the last element from a python list.

example:- l = [1,2,3]

l.pop(2)

print(l)

output [1,2]

Clear() Removes all items from the dictionary.

Examples:- cubes = {1:1,2:8,3:27}

print(cubes)

{1:1,2:8,3:27}

cubes.clear()

output { }

popitem() Remove and return an arbitrary item (key,value) raises key error if the dictionary is empty.

Example :- Squares = {1:1,2:4,3:9}

print(squares) {1:1,2:4,3:9}

squares.popitem() {2:4,3:9}

Answer

FROZEN SET

Frozen set is just an immutable version of a python set object.

*While elements of a set can be modified at any time.Element of the Frozen set remain the same after Creation.

*Unordered

*Unindex

*No duplicate

Example:-

Vowels = {'a','e','i','o','u'}

Fset = Frozenset{Vowels)

Frozenset ({'i','o','u','e','a'})

Answer

Different between muand immutable data type in python

Mutable	Immutable
1) Object can be changed after Creation.	1) Object cannot be changed afte Creation.
2) Generally provides a method to add or remove elements.	2)It does not provide any method to add or to remove and changes the element.
3) Slower to access compare to immutable object.	3)Quiker to access compare to mutable object.
4)Changing mutable object is is easy and efficient.	4) can't be changed
e.g:-Dictionary,sets,lists etc.	e.g:-String,Byte,frozenset,tuple.etc

Answer

__init__

The method __init__ is a special method in python that is called automatically when creating a new intance of a class.

When we use this method to initialize the instance and it can be used to perform any action before the instance is applied.

Example:-

```
class student:
        def__init__(self,name,age)          (here __init__ is used)
        self.name = name
        self.age = age
class student_marks:
        def__init__(self,name,english,maths)          (here __init__ is
used
                self.name = name
                self.english = english
                self.maths = maths
        def Calculate__avg__marks(self):
                total_marks = self.english + self.maths.
                return avg_marks
Student1 = student_marks("Tanveer",20,12)
Student2 = student_marks("Sultan",10,20)

print(Studen1.Calculate__avg__marks() )
print(studen2.Calculate__avg__marks() )
```

output tanverr 16

Sultan 15

Answer

Docstring in Python

Docstring is a short form of documentation string.These are the string literals that are used within a particular program or a piece of code.

Docstring and Comments are used to make program more understandable.

Example:- Def add two number(a,b):

"""python program to add two numbers."""

(**docstring**)

print(a+b)

print(add two number.__doc__)

Answer

unit test in python

Unit testing is the process of checking small pieces of code to ensure that the individual parts of a program work properly on their own.

- unit tests are used to test individual blocks(unit) of functionality.
- unit testing is done by developers.

```
Example:- def price_calculator(items):
                if items <= 0:
                    return    0
                elif 1 <= items <= 10
                    price_per_item = 10
                else: #The implicity covers the case where items >= 11
                    price_per_item = 9
                return item * price_per_item.
```

#when we first write a function like this,we might manually test it by entering some numbers to see, if we get the expected price.

Answer

Break,Continue and **Pass** are used to control the Flow of loops in python.

Break

'Break' statement is used to exit a loops prematurely when a specific condition is met.

```
Example list1 = range(9)
            for x in list1:
                if (x == 5):
                    break
                print(x)
```

Continue

'Continue' statement is used to skip the current iteration of a loop when a specific condition is met.

pass

'Pass' statment is a no-op statement in python meaning it does nothing when executed.

Answer

Use of self keyword in python

Self keyword is like a variable which refers to current instance (i.e object)

of a class in order to access its attributes and methods inside class.

Example :-

```
class Bike:
    a = 9
  def__sample_one(self):
  print("Inside sample_one method")
  def__sample_two(self):
            print(self.a)
```

Answer

Global,protected and private attributes in python

Global attributes:-

Global attributes are defined outside of a class and can be accessed from anywhere in the code,including inside classes.

Protected attributes:-

Protected attributes are denoted with a single underscore(_) before the attributes name and are intended to beaccessed only within the class and its subclasses.

Private attributes:-

Private attributes are denoted with double underscores(__) before the attributes name and are intended to be accessed only within the class in which they are defined.

Example:- calss student:

```
"""My student class"""
  def__init__(self,name)
```

Answer

Module and package in Python

Module:-

A module is a pythonic statement which contain multiple function in it.Modules act as a pre-defined library in the code.Which is accessible to both coder and user.

Example:- 1. import Math

2.from math import pow

3. pow(2,8)

4. print(pow)

output <buit in function pow>

Packages:-

A package is the form of collection of tools which helps in the intiation of the

code.A python package act as a user-variables interface for any source code.This make a python package work at a defined time for any functionable code in the runtime.

Example:-

```
1. import math
2.print("math package")
```

output math package.

Answer

Lists and Tuples

Lists:-

A lists is a data structure in python that is a mutable or changeable ordered sequence of elements.

Tuples:-

A tuples is a type of data structure that is very similar to lists.The main difference between the two is that tuples are immutable and unordered.

Answer

Interpreted language and Dynamically type language

Interpreted language:-

An Interpreted language is a language in which the implementations execute instruction directlty without earlier compiling a program into machine language.

Dynamically type language:-

In dynamically typed languages,type checking takes place at runtime or execution time.

Ex:-Python,Javascript.

Answer

Dict Comprehensions.

Dict Comprehension are used to creatwe new dictionaries in a similar manner.They allow you to specify key-value pairs using expressions.

for Example, To create a dictionary that maps number to their squares for even numbers from 1 to 10.

```
Squares_dict = {x:x**2 for x in range (1,11)     if x%2 == 0}
```

List Comprehension.

List Comprehension are used to create new lists by applying an expression to each item in an iterable.(e.g a list,tuple,or string) and filtering the items based on a condition.

for Example, To create a list of msquare of even numbers from 1 to 10.

Square_of_even = [x**2 for x in range (1,11) if x%2 == 0]

Answer

Decorators

Decorators are very powerful and useful tiils in python as they are the specific range change that we make in python **syntax** to alter function easily.

Use cases for Decorators

Decorators are widely used for a variety of purpose such as:-

- Timing function execution.
- Catching results (memorization).
- Authorization and authentication.
- logging and error handling.
- Input validation.

Answer

Memory Management in Python.

Python uses its private head space to manage the memory,Basically all the objects and data structures are stored in the private head space.Even the programmer can not access this private space.Python also has an inbuilt garbage collector,Which recycles all the unused memory and frees the memory and makes it available to the head space.

Answer

lambda and its uses in Python

A lambda function is an anonymous function.This function can have any number of parameters,but can have just one statement.

For Example a = lambda x,y : x*y

print(a(7,19))

output 133

uses

Lambda is a keyword in python for defining the anonymous function.argument(s) is a placeholder, that is a variable that will be used to hold the vale you want to pass into the function expression.

A lambda function can have multiple variables depending on what ypu want to achieve.

Answer

Function of split() and join() in Python

Split()

The split() function can be used to divide a string into a list of smaller strings based on a specific delimeter.It is the code below string.split() breaks the string

into a list of words.using the space character as the delimeter.

Join()

The join functiion concatenates elements within an iterable. Such as a list into a single string.In the code snippet.

Join(string_list) rejoin the previously split words into a cohesive sentence by placing a space between each word.

Answer

Iterators,iterable and Generators in python

Iterators:-

An iterators is an objects that can be iterated upon. Thus,iterators contain a countable number of values.

iterable:-

Iterables is an object which can be looped over or iterated over with the help of a for loop.Object like lists,tuples,sets,string etc are called iterables.

Generator:-

A special type of function which does not return a single value,a function or expression that will process a given iterables one object at a time on demand.A generator will return a generator objects.

Answer

Difference between Xrange and range Function.

Range and Xrange are two function that could be used to itrate a certain number of times in for loops in Python.

- Range():- This return a list of numbers created using the range() Function.
- Xrange():-This function return the generator object that can be used to display numbers only by looping.The only particular range is displayed by looping.The only particular range is displayed on demand amd hence called lazy evaluation.

Answer

Pillars of Oops?

The four pillars of Object-Oriented Programming.

1. Abstraction.
2. Encapsulation.
3. Inheritance.
4. Polymorphism.

Answer

Method to find if a class is a child of another class.

In python, you can check if a class is a sub-class of another class by using the

built-in issubclass() function.
The issubclass() function takes two argument the first argument is the class you want to cheack and the second argument is the potential superclass.
Here's an example

```
1. class student:
2.             def__init__(self,fname,lname)
3.                         self.first name = fname
4.                         self.lastname = lname]
5.           def printname(self):
6.                         print(self.firstname,self.lastname)
7.  class person(student):
8.               def__init__(self,fname,lname)
9.  X =     person("vipin","kumar")
10.                        X.printname()
11.                        Vipin kumar                              (output)
```

Answer

Inheritance in Python

When a child inherited the properties of a parent class,it is called inheritance.
There are five types of inheritance in python.

1. Single Inheritance.
2. multi-level inheritance.
3. Multiple Inheritance.
4. Heirarchical Inheritance.
5. Hybrid Inheritance.

Single Inheritance:-

When a class have only one parent class.

```
1.  classfather:
2.  def father_property(self):
3.                  print("This is the father property")
4.  class son(father):
5.               def job(self):
6.                           print("son has also a job")
7.  child_obj=son()
8.  child_obj.job()
9.  son has also a job
10. child_obj.father property()
11.                  This is the father property.
```

Multi-level Inheritance:-

In multi-level inheritance the transfer of the properties of characteristics is done to more than one class hierarchically.

To get a better visualization we can consider it as an ancestor of grand children relation.

```
1. class Grandfather:
2.         def prop_grand_father(self):
3.             print("I am your grand father")
4. class father(Grand father):
5.         def prop_father(self):
6.             print("I am your father having property of mine and
   your Grand father as well")
7. class son(father):
8.         def prop_son(self):
9.             print("I am son and have property of both my father and
   Grand father")
```

Multiple inheritance

This inheritance enables a child class to inherit from more than one parent class.

Hierarchical Inheritance

This inheritance allows a class to host as a parent class for more than one child class or subclass.

Hybrid Inheritance

An inheritance is said to hybrid inheritance if more than one type of inheritance is implemented in the same code.

<u>Answer</u>

Encapsulation

Encapsulation means binding the code and data together.A python class is an example of encapsulation.

Generally, we restrict data access outside the class in encapsulation.

Example

```
1.  class student:
2.      def__init__(self,name,age,grade)
3.          self.__name = name
4.          self.__age = age
5.          self.__grade = grade
6.  #Getter Method
7.      def get_name(self):
```

```
8.                                  return self.__name.
9.              #Setter Method
10.                         def set_grade(self,grade):
11.                          if 0 <= grade <= 100:
12.                                  self.__grade = grade
13.         #Create student object
14.        student1 = student("Alice",18,19)
15.         #Accessing attributes through methods
16.                      print(student1.get__name())
17.         #output              "Alice"
```

Answer

Polymorphism

Polymorphism means same function name(buit different signature) being uses for different types.Same function have different form.

Example:-

- l = [10,20,30,40] (performing of Lists)
- print(len(l))
- 4
- s = "Welcome" (performing on string)
- print(len(s))
- 7

Answer 1.2

1st_Room:- Invalid, Identifier cannot start with a number.

Hundred$:- Invalid,Identifier cannot contain special characters like $.

total-marks:-Invalid,Identifier cannot contain hypens(-).

Percentag:- Invalid,Identifier cannot start or end with an underscore() should be fully spelled.

Answer 21

Skewness

Skewness of a distribution is defined as the lack of symmetry. In a symmetrical distribution, the Mean, Meadian and Mode are equal.The normal distribution has a skewness of 0.

Skewness tell us about distribution of our data.

Skewness is of two types

Positive skewness: When the tail on the right side of the distribution is longer or fatter, we say the data is positively skewed. For a positive skewness mean > median > mode.
Negative skewness:: When the tail on the left side of the distribution is longer or fatter, we say that the distribution is negatively skewed. For a negative skewness mean < median < mode.
Uses of Skewness:-
There are a few different ways that skewed data can be useful.

One way that skewed data can be useful is if you are trying to identify a trend. If you have a lot of data that is skewed in one direction, it can be helpful in identifying a trend. This is because the trend will be more pronounced in the data that is skewed in one direction.
Another way that skewed data can be useful is if you are trying to identify outliers. If you have a lot of data that is skewed in one direction, it can be helpful in identifying outliers. This is because the outliers will be more pronounced in the data that is skewed in one direction.

Answer 22

PROBABILITY MASS FUNCTION (PMF):-PMF is used to describe the probability distribution of a discrete random variable. It assigns probabilities to each possible value of the random variable.
PROBABILITY DENSITY FUNCTION:-A PDF is used to describe the probability distribution of a continuous random variable. It provides the relative likelihood of the random variable taking on different values within a range.

Probability mass function (PMF)	Probability density distribution (PDF)
Describes discrete random variables	Describes continuous random variables
Y-axis is a probability	Y-axis is a probability density
Sum of probabilities for all x is equal to 1	Area under PDF is equal to 1

Answer 23

Correlation

A correlation is a statistical measure of the relationship between two variables. The measure is best used in variables that demonstrate a linear relationship between each other. The fit of the data can be visually represented in a scatterplot. Using a scatterplot, we can generally assess the relationship between the variables and determine whether they are correlated or not.

The correlation coefficient is a value that indicates the strength of the relationship between variables. The coefficient can take any values from -1 to 1. The interpretations of the values are:

- -1: Perfect negative correlation. The variables tend to move in opposite directions (i.e., when one variable increases, the other variable decreases).
- 0: No correlation. The variables do not have a relationship with each other.
- 1: Perfect positive correlation. The variables tend to move in the same direction (i.e., when one variable increases, the other variable also increases).

Methods of Determining Correlation:

There are various methods of determining correlation in two variables. Here we discuss only following methods:

1. Karl Pearson's Coefficient of correlation

2. Spearman's rank Correlation method

3. Method of Least Squares.

Answer25

Certainly! Correlation and regression are both statistical techniques used to analyze the relationships between variables, but they serve different purposes and have different characteristics. Here are four key differences between them:

Purpose and Interpretation:

Correlation: Measures the strength and direction of the linear relationship between two variables. It quantifies how closely two variables move together, but it does not imply causation. The correlation coefficient, typically denoted as
r
r, ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship.
Regression: Assesses the nature of the relationship between an independent variable (predictor) and a dependent variable (response). It is used to make

predictions or to understand how changes in the predictor variable affect the response variable. Regression analysis provides an equation that models this relationship, typically in the form
$Y=\beta 0+\beta 1X+\epsilon$
Y=β 0+β 1X+ϵ, where $\beta 0$β 0and $\beta 1$β 1

are coefficients, and ϵ
ϵ is the error term.
Directionality:

Correlation: Does not imply a direction of influence. It simply measures how variables move together without suggesting which variable affects the other. For example, if you find a high correlation between height and weight, it doesn't indicate whether height affects weight or vice versa.
Regression: Establishes a direction of influence. In regression, you specify which variable is the predictor (independent) and which is the response (dependent). This allows you to interpret how changes in the predictor variable affect the response variable.
Mathematical Representation:

Correlation: Is represented by a single number, the correlation coefficient
r
r, which summarizes the strength and direction of the linear relationship. It does not provide a formula or equation for predicting values.
Regression: Provides an equation (e.g.,
$Y=\beta 0+\beta 1X$Y=β0+β 1

X) that describes the relationship between variables. This equation can be used to make predictions about the dependent variable based on the independent variable.
Use Cases:

Correlation: Primarily used when the goal is to explore and quantify the degree of association between two variables. It is useful for preliminary analysis to understand relationships but does not suggest a causal relationship.
Regression: Used when the goal is to predict the value of one variable based on another or to understand the nature and strength of the causal relationship

between variables. It is more suitable for making predictions and conducting causal inference, assuming appropriate model conditions are met.
In summary, while correlation gives a measure of how two variables are related, regression provides a detailed model for understanding and predicting relationships between variables.

Answer28

Normal Distribution is a fundamental concept in statistics and probability theory. It describes a continuous probability distribution that is symmetric about its mean, with a shape known as the "bell curve." The normal distribution is characterized by its mean (μ) and standard deviation (σ), and it is used in various fields including natural and social sciences to model real-world phenomena.

Four Assumptions of Normal Distribution:
Continuous Variable:

Assumption: The variable being measured is continuous rather than discrete. This means that the variable can take any value within a given range and is not limited to specific, isolated values.
Explanation: Normal distribution applies to variables where you can theoretically obtain an infinite number of possible values within any given range. For instance, height, weight, and temperature are continuous variables, whereas categorical data like gender or political affiliation would not be modeled by a normal distribution.
Symmetry:

Assumption: The distribution of the variable is symmetric about its mean.
Explanation: This implies that the data follows a pattern where deviations from the mean are equally likely to be above or below the mean. This symmetry ensures that the distribution has no skewness. If a dataset is not symmetric, it may not follow a normal distribution.
Mean and Standard Deviation:

Assumption: The distribution is defined by its mean (μ) and standard deviation (σ), and the entire shape of the distribution is determined by these two parameters.
Explanation: The mean (μ) specifies the center of the distribution, while the

standard deviation (σ) measures the spread or dispersion of the data. The normal distribution curve is fully described by these parameters, where about 68% of the data falls within one standard deviation of the mean, about 95% within two standard deviations, and about 99.7% within three standard deviations (this is known as the 68-95-99.7 rule).
Independent and Identically Distributed Data:

Assumption: The data points are independent of each other and are drawn from the same probability distribution.
Explanation: Each data point in the sample is assumed to be independent of the others, meaning that the value of one observation does not affect the value of another. Additionally, all observations are assumed to come from the same normal distribution with the same mean and standard deviation. This ensures that the properties of the normal distribution apply uniformly across the dataset.

Answer29

The normal distribution curve, also known as the Gaussian curve, has several distinctive characteristics and properties. These properties are central to understanding how data behaves in many natural and social phenomena. Here are the key characteristics of the normal distribution curve:

1. **Bell-Shaped Curve:**
- The shape of the normal distribution curve is bell-shaped, symmetric about its mean. The curve rises gradually from the tails to a peak at the mean, then falls symmetrically.

2. **Symmetry:**
- The normal distribution is perfectly symmetric around its mean. This means that the left side of the curve is a mirror image of the right side. Consequently, the mean, median, and mode of the distribution are all equal and located at the center of the distribution.

3. **Mean, Median, and Mode:**
- In a normal distribution, the mean (μ), median, and mode are all equal and situated at the center of the distribution. They represent the highest point of the curve.

4. **Asymptotic**:

- The tails of the normal distribution curve approach the horizontal axis but never actually touch it. This means the curve extends infinitely in both directions.

5. **Empirical Rule (68-95-99.7 Rule):**

- Approximately 68% of the data within a normal distribution falls within one standard deviation (σ) of the mean (μ).
- About 95% of the data falls within two standard deviations of the mean.
- Around 99.7% of the data falls within three standard deviations of the mean. This rule helps in understanding the spread of data in a normal distribution.

6. **Probability Density Function (PDF):**

- The probability density function (PDF) of a normal distribution is given

7. **Area Under the Curve**:

- The total area under the normal distribution curve equals 1. This represents the total probability of all possible outcomes in the distribution.

8. **Standardization:**

- Any normal distribution can be standardized to a standard normal distribution with a mean of 0 and a standard deviation of 1 by transforming the variable using:
- The resulting distribution, called the standard normal distribution, has the same properties as the normal distribution but is standardized.

9. **68-95-99.7 Rule:**

- This empirical rule describes the percentage of data within one, two, and three standard deviations from the mean:
  - Approximately 68% of the data falls within one standard deviation of the mean.
  - Approximately 95% falls within two standard deviations.
  - Approximately 99.7% falls within three standard deviations.

10. **No Skewness:**

- The normal distribution has zero skewness, meaning that it is not skewed to the left or right. The distribution is perfectly balanced around the mean.

11. **Kurtosis:**

- The normal distribution has a kurtosis of 3, which is considered

"mesokurtic." This means it has a moderate peak and tails compared to other distributions. Higher kurtosis indicates a "leptokurtic" distribution (more peaked with heavier tails), while lower kurtosis indicates a "platykurtic" distribution (flatter with lighter tails).

12. **Maximum Entropy:**

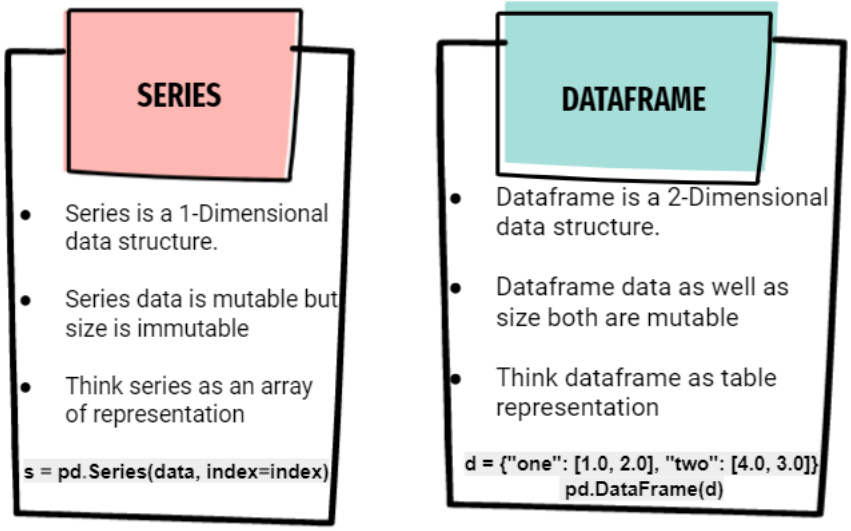
- Among all continuous distributions with a given mean and variance, the normal distribution has the maximum entropy. This means it is the most "uninformative" distribution, assuming no other information is known.

Understanding these characteristics helps in applying statistical techniques correctly and interpreting the behavior of data that follows a normal distribution.

Answer50

Machine learning

Answer



Answer

Pandas iloc VS loc

Feature	iloc	loc
Indexing	Label-based (names)	Integer-based (positions)
Flexibility	More flexible with labels	Less flexible, but precise with positions
Slicing	Inclusive (includes endpoint)	Exclusive (excludes endpoint)
Error handling	KeyError for invalid labels	IndexError for invalid positions

Answer

Difference Between Supervised and Unsupervised learning.

Supervised learning	Unsupervised learning
Input data is labelled	Input data is unlabelled
There is a training phase	There is no training phase
Data is modelled based on training dataset	Uses properties of given data for classification
Divided into two types: Classification and Regression	Most popular types: Clustering and Dimensionality reduction
Known number of classes (for classification)	Unknown number of classes

Answer

Bias-Variance Tradeoff

Bias-Variance Tradeoff is a fundamental concept in machine learning that deals with the balance between model bias and variance. In simpler terms, it refers to the tradeoff between a model's ability to accurately represent the underlying data patterns (low bias) and its susceptibility to fluctuations with changes in the training data (high variance).

Answer

Precision and Recall.

Precision and Recall are two evaluation metrics used to determine the effectiveness of a classification model. These metrics help to measure the quality of the model by looking at the number of true positives, true negatives, false positives, and false negatives.

Precision is the ability of a classification model to identify only the relevant data points. It is the fraction of the true positives predicted by the model out of all the true and false positives. For example, if a spam filter marks an email as spam and it is indeed spam, then this is counted as a true positive. Precision is calculated as:

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

Recall, on the other hand, is the ability of a classification model to find all the relevant data points. It is the fraction of the true positives predicted by the model out of all the true positives and false negatives. For example, if a spam filter misses a spam email, then this is counted as a false negative. Recall is calculated as:

Recall = True Positives / (True Positives + False Negatives)

- **Accuracy** shows how often a classification ML model is correct overall.
- **Precision** shows how often an ML model is correct when predicting the target class.
- **Recall** shows whether an ML model can find all objects of the target class.

Answer

Overfitting

In machine learning, overfitting occurs when an algorithm fits too closely or even exactly to its training data, resulting in a model that can't make accurate predictions or conclusions from any data other than the training data.
You can prevent overfitting by diversifying and scaling your training data set or using some other data science strategies, like those given below.

- Early stopping. Early stopping pauses the training phase before the machine learning model learns the noise in the data. ...
- Pruning. ...
- Regularization. ...
- Ensembling. ...

- Data augmentation.

Answer

cross-validation

Data scientists rely on several reasons for using cross-validation during their building process of Machine Learning (ML) models. For instance, tuning the model hyperparameters, testing different properties of the overall datasets, and iterate the training process. Also, in cases where your training dataset is small, and the ability to split them into training, validation, and testing will significantly affect training accuracy. The following main points can summarize the reason we use a CV, but they overlap. Hence, the list is presented here in a simplified way:

Answer

Classification vs Regression

Regression Algorithm	Classification Algorithm
The output variable has to be a real value or continuous in nature.	The output variable is discrete in the Classification SML problem.
Regression algorithm helps to map the input value and the continuous output variable.	Classification algorithm helps in mapping the input value with the output variable which is discrete in nature.
Regression algorithms are only used for data that is continuous.	Classification algorithms are only used for data that is discrete.
Finds the best fit for accurately predicting the output data in Regression.	Finds the decision boundary which helps in classifying the dataset into different classes in Classification
Regression algorithms help in solving the problems like predicting the weather or the price of Houses in a neighbourhood according to a real estate trend.	Classification algorithms are useful for problems like identifying spam emails, recognition of speech, identifying cancer cells, etc.

Answer

Ensemble Learning

Ensemble learning is a machine learning method that combines the predictions from multiple models to improve predictive performance. It strategically combines classifiers or expert models to address problems such as regression and classification

Answer

gradient descent

Gradient Descent stands as a cornerstone orchestrating the intricate dance of model optimization. At its core, it is a numerical optimization algorithm that aims

to find the optimal parameters—weights and biases—of a neural network by minimizing a defined cost function.

Gradient Descent (GD) is a widely used optimization algorithm in machine learning and deep learning that minimises the cost function of a neural network model during training. It works by iteratively adjusting the weights or parameters of the model in the direction of the negative gradient of the cost function until the minimum of the cost function is reached.

Answer

Difference between batch gradient descent and stochastic gradient descent.

Batch Gradient Descent	***Stochastic Gradient Descent***
Computes gradient using the whole Training data	Computes gradient using a single Training data
Slow and computationally expensive algorithm	Faster and less computationally expensive than Batch GD
Not suggested for large dataset	Can be used for large datasets
Deterministic in nature.	Stochastic(random) in nature.
Gives optimal solution given sufficient time to	Gives good solution but not optimal.
No random shuffling of points are required.	The data sample should be in a random order, have to shuffle the training set for every epoch.
Used with convex cost functions	Used with non-convex cost functions
Can't escape shallow local minima easily.	SGD can escape shallow local minima more easily due to its randomness
Convergence is slow.	Converges much faster.

Answer

curse of dimensionality in machine learning.

- The Curse of Dimensionality refers to the phenomenon where the efficiency and effectiveness of algorithms deteriorate as the dimensionality of the data increases exponentially.
- In high-dimensional spaces, data points become sparse, making it challenging to discern meaningful patterns or relationships due to the vast amount of data required to adequately sample the space.
- The Curse of Dimensionality significantly impacts machine learning algorithms in various ways. It leads to increased computational complexity, longer training times, and higher resource requirements. Moreover, it escalates the risk of overfitting and spurious correlations, hindering the algorithms' ability to generalize well to unseen data.

Answer

difference bet_een L1 and L2 regularization"

L1 Regularisation	L2 Regularisation
Sum of absolute value of weights	Sum of square of weights
Sparse solution is the outcome	Non-sparse (more segregated) solution
Multiple solutions are possible	Only one solution
Built-in feature selection in the penalty term	No specific feature selection mechanism
Robust to outliers	Not robots to outliers due to square term
Used in datasets with sparse features	Used in dataset with complex features

Answer

confusion matrix

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is a means of displaying the number of accurate and inaccurate instances based on the model's predictions. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance.

Uses of Confusion matrix

- **True Positive (TP):** The model correctly predicted a positive outcome (the actual outcome was positive).
- **True Negative (TN):** The model correctly predicted a negative outcome (the actual outcome was negative).
- **False Positive (FP):** The model incorrectly predicted a positive outcome (the actual outcome was negative). Also known as a Type I error.
- **False Negative (FN):** The model incorrectly predicted a negative outcome (the actual outcome was positive). Also known as a Type II error.

Answer

AUC-ROC curve

The AUC-ROC curve, or Area Under the Receiver Operating Characteristic curve, is a graphical representation of the performance of a binary classification model at various classification thresholds. It is commonly used in machine learning to assess the ability of a model to distinguish between two classes, typically the positive class (e.g., presence of a disease) and the negative class (e.g., absence of a disease).

Let's first understand the meaning of the two terms ROC and AUC.

ROC: Receiver Operating Characteristics

AUC: Area Under Curve

Receiver Operating Characteristics (ROC) Curve:

ROC stands for Receiver Operating Characteristics, and the ROC curve is the graphical representation of the effectiveness of the binary classification model. It plots the true positive rate (TPR) vs the false positive rate (FPR) at different classification thresholds.

Area Under Curve (AUC) Curve:

AUC stands for the Area Under the Curve, and the AUC curve represents the area under the ROC curve. It measures the overall performance of the binary classification model. As both TPR and FPR range between 0 to 1, So, the area will always lie between 0 and 1, and A greater value of AUC denotes better model performance. Our main goal is to maximize this area in order to have the highest TPR and lowest FPR at the given threshold. The AUC measures the probability that the model will assign a randomly chosen positive instance a higher predicted probability compared to a randomly chosen negative instance.

It represents the probability with which our model can distinguish between the two classes present in our target.

Answer

K-Nearest Neighbors

KNN is one of the most basic yet essential classification algorithms in machine learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining, and intrusion detection.

It is widely disposable in real-life scenarios since it is non-parametric, meaning it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data). We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

Answer

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it's best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the

margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Answer

Kernel Trick in Support Vector Machine.

Kernel functions in SVMs are math tricks that help the model understand data better. By making it look like it's in a higher-dimensional space. They're important because they let SVMs draw more complicated lines between different groups of data points. With kernels, SVMs can handle all kinds of relationships between data points, making accurate predictions. Common types of kernels include linear, polynomial, RBF, and sigmoid, each good for different kinds of data. Picking the right kernel is super important because it decides how well the SVM will work. As well as how good it is at figuring out new data.

Answer

Types of Kernel in SVM

Here are some common types of kernels in support vector machine algorithms:

1. Linear Kernel

The linear kernel is the simplest and is used when the data is linearly separable. It calculates the dot product between the feature vectors.

2. Polynomial Kernel

The polynomial kernel is effective for non-linear data.

It computes the similarity between two vectors in terms of the polynomial of the original variables.

3. Radial Basis Function (RBF) Kernel

The RBF kernel is a common type of Kernel in SVM for handling non-linear decision boundaries.

It maps the data into an infinite-dimensional space.

4. Sigmoid Kernel

The sigmoid SVM kernel types can be used as an alternative to the RBF kernel.

It is based on the hyperbolic tangent function and is suitable for neural networks and other non-linear classifiers.

5. Custom Kernels

In addition to the standard kernels mentioned above, SVMs allow the use of

custom kernels tailored to specific problems.
Custom kernels can be designed based on domain knowledge or problem-specific requirements.
Picking the right kernel from different types of Kernel in SVM depends on things. Like what the data looks like, and how complicated the boundary between classes is. As well as how fast you need the model to be. Also, you have to try different kernels and adjust settings to get the best results for what you're trying to do.

Answer

Hyperplane in SVM

In SVMs, a hyperplane is a subspace of one dimension less than the original feature space. In two-dimensional space, a hyperplane is a line, while in three-dimensional space, it is a plane. In general, in n-dimensional space, a hyperplane is an (n-1)-dimensional flat affine subspace.

The hyperplane in SVMs is represented by the equation:

$$w^T x + b = 0$$

where w is the weight vector, x is a data point, and b is the bias term. The weight vector determines the "shape" of the hyperplane, and the bias term determines the position of it.

The decision rule for SVMs is based on the sign of the equation above. If the result is positive, the data point is classified as one class, and if it is negative, the data point is classified as the other class.

Answer

Pros and cons of using a Support Vector Machine (SVM)

Advantages of Support Vector Machine:

1. SVM works relatively well when there is a clear margin of separation between classes.
2. SVM is more effective in high dimensional spaces.
3. SVM is effective in cases where the number of dimensions is greater than the number of samples.
4. SVM is relatively memory efficient

Disadvantages of Support Vector Machine:

1. SVM algorithm is not suitable for large data sets.
2. SVM does not perform very well when the data set has more noise i.e. target classes are overlapping.
3. In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.
4. As the support vector classifier works by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification.

Answer

Difference between hard margin and soft margin SVM

Hard margin SVM:

In a hard margin SVM, the goal is to find the hyperplane that can perfectly separate the data into two classes without any misclassification. However, this is not always possible when the data is not linearly separable or contains outliers. In such cases, the hard margin SVM will fail to find a hyperplane that can perfectly separate the data, and the optimization problem will have no solution.

Soft margin SVM:

In a soft margin SVM, we allow some misclassification by introducing slack variables that allow some data points to be on the wrong side of the margin. The optimization problem in a soft margin SVM is modified as follows:

minimize $1/2 \; ||w||^2 + C * \Sigma \xi i$

subject to $yi(w^T xi + b) \geq 1 - \xi i$

Answer

Decision Tree

A decision tree is a flowchart-like structure used to make decisions or predictions. It consists of nodes representing decisions or tests on attributes, branches representing the outcome of these decisions, and leaf nodes representing final outcomes or predictions. Each internal node corresponds to a test on an attribute, each branch corresponds to the result of the test, and each leaf node corresponds to a class label or a continuous value.

Structure of a Decision Tree

Root Node: Represents the entire dataset and the initial decision to be made.
Internal Nodes: Represent decisions or tests on attributes. Each internal node has

one or more branches.
Branches: Represent the outcome of a decision or test, leading to another node.
Leaf Nodes: Represent the final decision or prediction. No further splits occur at these nodes.

Answer

Working principle of Decision Tree

Even in only this simple form, a decision tree is useful to show the possibilities for a decision. However, a decision tree becomes especially useful when numerical data is added.

First, each decision usually involves costs. If a company decides to produce a product, engage in market research, advertise, or any other number of activities, the predicted costs for those decisions are written on the appropriate branch of the decision tree. Also, each pathway eventually leads to an outcome that usually results in income. The predicted amount of income provided by each outcome is added to that branch of the decision tree.

The other numerical data that needs to be provided is the probability of each outcome from the uncertainty nodes. If an uncertainty node has two branches that are both equally likely, each should be labeled with a 50 percent, or 0.5, probability. Alternatively, an uncertainty node might have three branches with respective probabilities of 60 percent, 30 percent, and 10 percent. In each case, the total of the percentages at each uncertainty node will be 100 percent, representing all possibilities for that node.

Answer

Information Gain

Information Gain is the difference between the entropy of the parent node and the weighted average entropy of the child nodes: Information Gain=Entropy(Parent)−Weighted Average Entropy(Children)Information Gain=Entropy(Parent)−Weighted Average Entropy(Children)

Information gain quantifies the effectiveness of an attribute in splitting the dataset and is used to select the best attribute for decision tree node splits. By calculating information gain for each attribute, decision trees can efficiently partition the dataset based on the most informative features, leading to effective classification or regression outcomes.

Answer

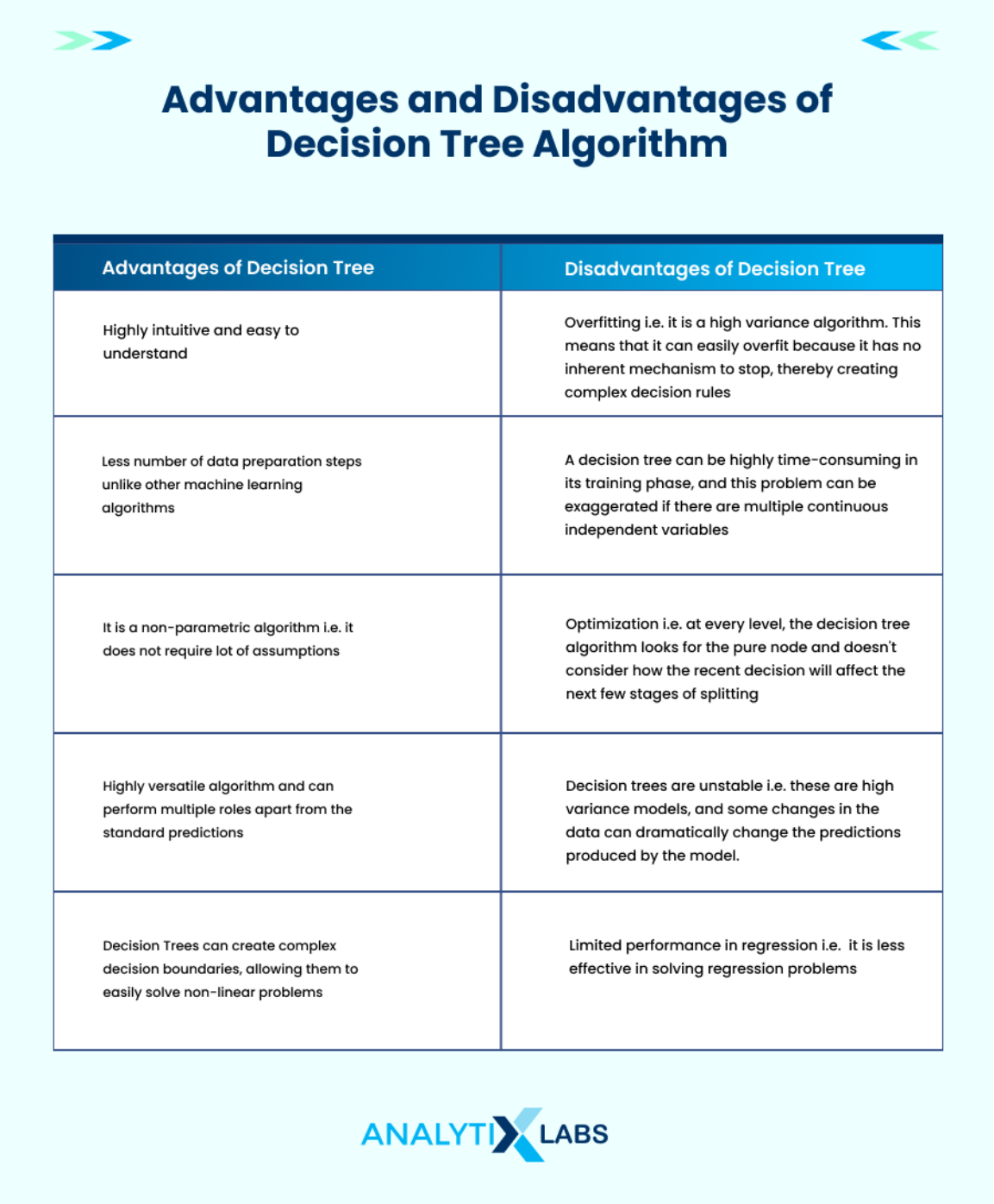
Gini Impurity: The internal working of Gini impurity is also somewhat similar to the working of entropy in the Decision Tree. In the Decision Tree algorithm, both are used for building the tree by splitting as per the appropriate features but there is quite a difference in the computation of both methods.

Working in Decision Trees.

Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

Answer

Advantages and Disadvantages of Decision Tree



Answer

Randon Forests outperform Decision Tree

Random forests are an extension of decision trees that work by constructing multiple decision trees and combining their results. Instead of relying on a single decision tree, random forests generate a collection of decision trees, each trained on a random subset of the input data.

The idea behind this approach is that by combining multiple models, the overall prediction accuracy will be improved. Random forests also help to reduce overfitting by using a technique called “bootstrap aggregating” or “bagging”. This involves sampling the training data with replacement so that each decision tree is trained on a slightly different set of data.

Random forests can be thought of as a group of experts, each with their own opinion on the best way to predict the outcome. By combining their opinions, we can arrive at a more accurate prediction than any one expert alone.

Answer

Random Forests working Principle

Random Forest algorithm is a powerful tree learning technique in Machine Learning. It works by creating a number of Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance.

In prediction, the algorithm aggregates the results of all trees, either by voting (for classification tasks) or by averaging (for regression tasks) This collaborative decision-making process, supported by multiple trees with their insights, provides an example stable and precise results. Random forests are widely used for classification and regression functions, which are known for their ability to handle complex data, reduce overfitting, and provide reliable forecasts in different environments.