MACHINE LEARNING

ANS.1 A) Least Square Error

ANS.2 A) Linear regression is sensitive to outliers

ANS.3 A) Positive

ANS.4 B) Correlation

ANS.5 C) Low bias and high variance

ANS.6 B) Predictive modal

ANS7 A) Cross validation

ANS.8 D) SMOTE

ANS.9 A) TPR and FPR

ANS.10 B) False

ANS.11 B) Apply PCA to project high dimensional data

ANS.12 D) It does not make use of dependent variable.

ANS.13 1. regularization

Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting. It can balance between overfitting and underfitting. We use many types of techniques to regularization..

ANS.14**particular algorithms are used for regularization**

Regularization techniques can be applied to a wide range of machine learning algorithms to prevent overfitting and improve generalization. Here are some popular algorithms across different machine learning domains that often incorporate regularization techniques:

1.Ridge Regression (L2 Regularization): Imagine you're building a model to predict house prices, but your model is trying too hard to fit every detail of the training data. Ridge regression helps by gently pulling back your model's enthusiasm, so it doesn't get too carried away with those details. It's like putting a leash on a very energetic dog to keep it in control.

2. Lasso Regression (L1 Regularization): Lasso is like a magic wand that not only helps control your model's enthusiasm but also says, "Hey, if some

features are not that important, let's make their coefficients zero." It's like decluttering your model's toolkit and keeping only the essential tools for the job.

Both Ridge and Lasso are ways to make your model more balanced and reliable, preventing it from focusing too much on small things and making sure it doesn't get too complex. They're like guides that help your model stay on the right track for making accurate predictions.

15. the term error present in linear regression equation.

The term "error" refers to the discrepancy between the actual observed values of the dependent variable (also known as the response variable or target variable) and the predicted values generated by the linear regression model. These discrepancies are also commonly referred to as "residuals."

The linear regression equation aims to find the best-fitting line that represents the relationship between the independent variable(s) (also known as features or predictors) and the dependent variable. However, due to factors such as measurement inaccuracies, inherent variability in the data, and the simplifications made by the linear model itself, the predicted values are not likely to match the actual observed values perfectly.

Mathematically, for each data point i, the error or residual e_i is given by

$e_i = y_i - \hat{y}_i$

Where

y_i is the actual observed value of the dependent variable for data point i,
$\hat{y}_i$ is the predicted value of the dependent variable for data point i.

PYTHON – WORKSHEET 1

ANS.1 C) %
ANS.2 B) 0
ANS.3 C) 24
ANS.4 A) 2
ANS.5 D) 6
ANS.6 C) the finally block will be executed no matter if the try block raises an error or not.
ANS.7 A) It is used to raise an exception.
ANS.8 C) in defining a generator
ANS.9 A) _abc C) abc2

ANS.10 A) yield B) raise

ANS.11

```
def factorial(n):
    if n == 0:
        return 1
    else:
        return n * factorial(n - 1)

num = int(input("Enter a number: "))
if num < 0:
    print("Factorial is not defined for negative numbers.")
else:
    result = factorial(num)
    print(f"The factorial of {num} is {result}")
```

ANS.12

```
 def is_prime(number):
    if number <= 1:
        return False
    for i in range(2, int(number**0.5) + 1):
        if number % i == 0:
            return False
    return True

num = int(input("Enter a number: "))
if is_prime(num):
    print(f"{num} is a prime number.")
else:
    print(f"{num} is a composite number.")
```

ANS.13

```
 def is_palindrome(string):
    string = string.lower()  # Convert the string to lowercase
    cleaned_string = ''.join(char for char in string if char.isalnum())  # Remove non-
alphanumeric characters
    return cleaned_string == cleaned_string[::-1]

input_string = input("Enter a string: ")
if is_palindrome(input_string):
    print("The string is a palindrome.")
else:
    print("The string is not a palindrome.")
```

ANS.14

```
def calculate_third_side(side1, side2):
```

```
    third_side = (side1**2 + side2**2)**0.5
    return third_side

side1 = float(input("Enter the length of the first side: "))
side2 = float(input("Enter the length of the second side: "))

third_side = calculate_third_side(side1, side2)
print(f"The length of the third side is: {third_side}")
```

ANS.15

```
def character_frequency(string):
    frequency = {}
    for char in string:
        if char.isalnum():  # Consider only alphanumeric characters
            char = char.lower()  # Convert to lowercase for case-insensitive counting
            frequency[char] = frequency.get(char, 0) + 1
    return frequency

input_string = input("Enter a string: ")
freq_dict = character_frequency(input_string)

print("Character frequencies:")
for char, count in freq_dict.items():
    print(f"'{char}': {count}")
```

STATISTICS WORKSHEET-1

ANS.1 a) True
ANS.2 a) Central Limit Theorem
ANS.3 b) Modeling bounded count data
ANS.4 d) All of the mentioned
ANS.5 c) Poisson
ANS.6 b) False
ANS.7 b) Hypothesis
ANS.8 a) 0

ANS.9 c) Outliers cannot conform to the regression relationship
ANS.10
The normal distribution, also known as the Gaussian distribution, is a continuous probability distribution that is symmetric and bell-shaped. It is characterized by two parameters: the mean (μ) and the standard deviation (σ). The distribution is centered around the mean, and its shape is determined by the standard deviation. Many natural phenomena and statistical processes tend to follow a normal distribution due to the Central Limit Theorem, making it a fundamental concept in statistics and data analysis.

ANS.11

Handling missing data is an important step in data analysis because incomplete information can lead to inaccurate results. There are various ways to deal with missing data, and one common approach is to use imputation techniques. Imputation means filling in the missing values with estimated or predicted values.

Here are a couple of simple imputation techniques:

1. Mean/Median Imputation: This involves replacing missing values with the mean (average) or median (middle value) of the available data. It's a straightforward approach that maintains the overall data distribution.

2. Mode Imputation: This is used for categorical data. It involves replacing missing values with the most common category (mode) in the dataset.

3. Forward Fill/Backward Fill: In time-series data, you can use the last observed value (forward fill) or the next observed value (backward fill) to fill in missing data.

4. Predictive Modeling: For more advanced cases, you can use machine learning algorithms to predict missing values based on other variables in the dataset. This can provide more accurate imputations, but it requires more complexity.

ANS.12

A/B testing is like trying two different things to see which one works better. Imagine you have a new idea for a website or an app, but you're not sure if it's better than what you already have. A/B testing helps you decide.

Here's how it works: You show one version of your idea to some people (let's call it A), and you show a slightly different version to another group of people (let's call it B). Then, you see which version makes people do what you want more. It could be clicking on a button, signing up, buying something, or anything you care about.

By comparing the results from both versions, you can figure out which idea is more successful and use that one. It's like trying out two options and picking the one that's more popular or effective. A/B testing helps you make smart choices based on what actually works best for your audience.

ANS.13

Imagine you have a list of people's ages, but some ages are missing. Mean imputation means you fill in the missing ages with the average age of all the people you do have ages for.

In simple language, it's like if you have a group of friends and you know their ages, but one friend didn't tell you their age. Instead of leaving it blank, you decide to use the average age of all your friends as a guess for that missing age.

Using mean imputation is okay in some cases, especially if you have just a few missing values. It helps you keep your data complete and makes calculations easier. However, it's important to remember that it can make your data less accurate if there are a lot of missing values or if the missing values have some special meaning.

So, it's like making an educated guess, but you should be careful and think about whether it makes sense for your specific situation.

ANS.14

Linear regression is like drawing a straight line through a cloud of points on a graph. It helps us understand how two things are related and predicts what might happen next.

Think of it this way: You have some data points that show how one thing changes when another thing changes. For example, how the amount of studying affects test scores. Linear regression finds the best-fitting line that goes through these points. This line helps you see if there's a pattern and how strong the relationship is.

Once you have this line, you can use it to make predictions. If you know how much someone studied, you can use the line to guess their test score. Or if you have data from the past, you can use the line to make predictions about the future.

So, linear regression is a tool that helps us understand and predict relationships between things by drawing a line that fits the data points as closely as possible.

ANS.15

Here's a brief explanation of some main branches of statistics:

1. Descriptive Statistics: Summarizes data using numbers like averages and graphs to give a clear picture.

2. Inferential Statistics: Makes predictions about a whole group based on a smaller sample.

3. Probability Theory: Studies the chance or likelihood of events happening.

4. Biostatistics: Deals with health and medical data to find patterns and draw conclusions.

5. Econometrics: Uses stats to understand economic data and make predictions.

6. Social Statistics: Analyzes data about people and societies to find trends and insights.

7. Psychometrics: Creates and analyzes tests used in psychology and education.

8. Actuarial Science: Deals with risks and uncertainties in fields like insurance and finance.

9. Time Series Analysis: Studies data collected over time to spot patterns.

10. Spatial Statistics: Examines data with a geographical aspect to find spatial patterns.

11. Multivariate Statistics: Looks at data with many variables to see how they're related.

12. Nonparametric Statistics: Uses methods that don't need specific assumptions about data.

13. Quality Control: Uses stats to make sure products and processes meet high standards.

14. Bayesian Statistics: Makes decisions using prior knowledge and updated beliefs.

15. Statistical Computing: Develops tools and software for data analysis.