**SECTION-1(Theory session)**

1.**Implement Linear Algebra using python**.

Ans: Linear Algebra is the mathematical foundation that solves the problem of representing data as well as computations in machine learning models.

**Scalar :**

Scalars are **numeric values that are logged during a run**. Scalars are used to log metrics like accuracy and loss. A scalar value is associated with a key and a step. A scalar step denotes the training step or epoch associated with a value.

**Scalars** are single numbers and are an example of a 0th-order tensor. In mathematics it is necessary to describe the set of values to which a scalar belongs. The notation x∈R states that the (lowercase) scalar value x is an element of (or member of) the set of real-valued numbers, R.

**Example:**



**Vectors :**

**Vectors** are ordered arrays of single numbers and are an example of 1st-order tensor. Vectors are members of objects known as **vector spaces**. A vector space can be thought of as the entire collection of *all* possible vectors of a particular length (or dimension). The three-dimensional real-valued vector space, denoted by R3 is often used to represent our real-world notion of three-dimensional space mathematically. Sometimes it is necessary to identify the *components* of a vector explicitly. The ith scalar element of a vector is written as xi. Notice that this is non-bold lowercase since the element is a scalar. An n-dimensional vector itself can be explicitly written using the following notation: x=[x1x2⋮xn]

**Example:**



**Matrices:**

**Matrices** are rectangular arrays consisting of numbers and are an example of 2nd-order tensors. If m and n are positive integers, that is m,nmn∈N then the m×n matrix contains mn numbers, with m rows and n columns.

Its components are now identified by two indices i and j. i represents the index to the matrix row, while j represents the index to the matrix column. Each component of A is identified by aij.

The full m×n matrix can be written as:

A=[a11a12a13…a1na21a22a23…a2na31a32a33…a3n⋮⋮⋮⋱⋮am1am2am3…amn]

It is often useful to abbreviate the full matrix component display into the following expression:

A=[aij]m×n

Example:



**Tensors:** The more general entity of a tensor encapsulates the scalar, vector and the matrix. It is sometimes necessary—both in the physical sciences and machine learning—to make use of tensors with order that exceeds two.

In theoretical physics, and general relativity in particular, the Riemann curvature tensor is a 4th-order tensor that describes the local curvature of spacetime. In machine learning, and deep learning in particular, a 3rd-order tensor can be used to describe the intensity values of multiple channels (red, green and blue) from a two-dimensional image.

**Example:**



**Gradient :**

Gradient descent is an optimization algorithm that’s used when training a machine learning model. It’s based on a convex function and tweaks its parameters iteratively to minimize a given function to its local minimum.

(OR)

b)2. **Describe univariant, bivariant and multi variant analysis with suitable example**

Ans: **Univariate data:**

Univariate data refers to a type of data in which each observation or data point corresponds to a single variable. In other words, it involves the measurement or observation of a single characteristic or attribute for each individual or item in the dataset. Analyzing univariate data is the simplest form of analysis in statistics.

**Example:**

| **Heights (in cm)** | **164** | **167.3** | **170** | **174.2** | **178** | **180** | **186** |
| --- | --- | --- | --- | --- | --- | --- | --- |

**Bivariate data:**

Bivariate data involves two different variables, and the analysis of this type of data focuses on understanding the relationship or association between these two variables. **Example**of bivariate data can be temperature and ice cream sales in summer season.

| **Temperature** | **Ice Cream Sales** |
| --- | --- |
| 20 | 2000 |
| 25 | 2500 |
| 35 | 5000 |

**Multivariate data:**

Multivariate data refers to datasets where each observation or sample point consists of multiple variables or features. These variables can represent different aspects, characteristics, or measurements related to the observed phenomenon. When dealing with three or more variables, the data is specifically categorized as multivariate.

**Example:**

| **Advertisement** | **Gender** | **Click rate** |
| --- | --- | --- |
| **Ad1** | **Male** | **80** |
| **Ad3** | **Female** | **55** |
| **Ad2** | **Female** | **123** |
| **Ad1** | **Male** | **66** |
| **Ad3** | **Male** | **35** |

**3.Give two fair dices through, what is the probability that two dices through sum is 8? When the first dices is 3.**

**Ans:**

Create a table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1,1 | 2,1 | 3,1 | 4,1 | 5,1 | 6,1 |
| 1,2 | 2,2 | 3,2 | 4,2 | 5,2 | 6,2 |
| 1,3 | 2,3 | 3,3 | 4,3 | 5,3 | 6,3 |
| 1,4 | 2,4 | 3,4 | 4,4 | 5,4 | 6,4 |
| 1,5 | 2,5 | 3,5 | 4,5 | 5,5 | 6,5 |
| 1,6 | 2,6 | 3,6 | 4,6 | 5,6 | 6,6 |

When rolling two fair dices ,the possible sums range from 2to12.to find the probability that the sum of the two dice is 8,we first list all the combinations of dice rolls that give a sum of 8 :

\*(2,6)

\*(3,5)

\*(4,4)

\*(5,3)

\*(6,2)

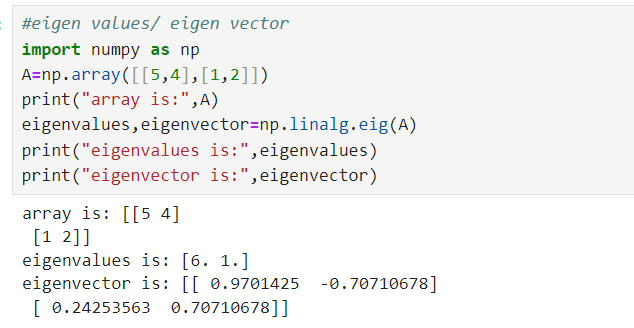
These are 5 possible combinations out of a total of 36 possible outcomes when rolling two dice (since each die has 6 faces ,so 6\*6=36

The first die be 3 to get a total sum of 8,the second die must roll a 5 because 3+5=8.

**SECTION-2(Theory session/practical session)**

**1.Write a python program to find eigenvalues and eigenvectors and how it works.**

**Ans:**



**Logics**

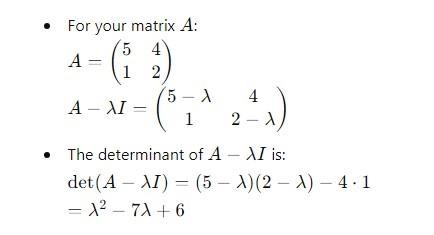
**I.Eigen values**

An eigenvalue is a scalar λ associated with a linear system of equations (represented by a matrix)

eigenvalues,eigenvectors = np.linalg.eig(A)

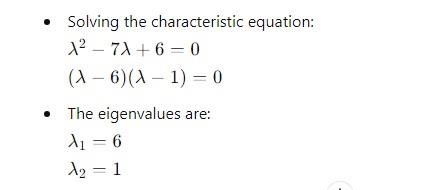
Step 1: First finds the characteristic equation

Ex:



Step 2: Solve the equation and find the eigen values

Ex:



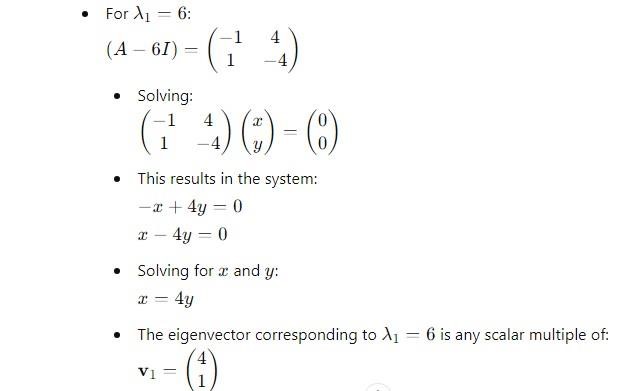
# II. Eigen vectors

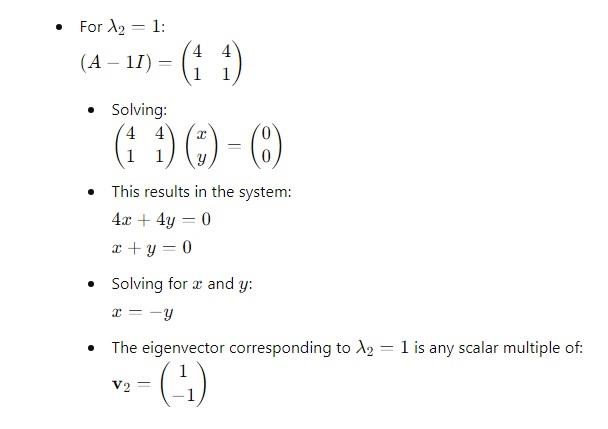
An eigenvector 𝑣 of a matrix 𝐴 is a non-zero vector that only changes by a scalar factor when that matrix is applied to it. In other words, it satisfies the equation

eigenvalues,eigenvectors = np.linalg.eig(A)

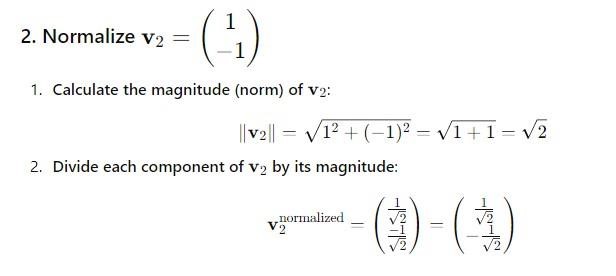
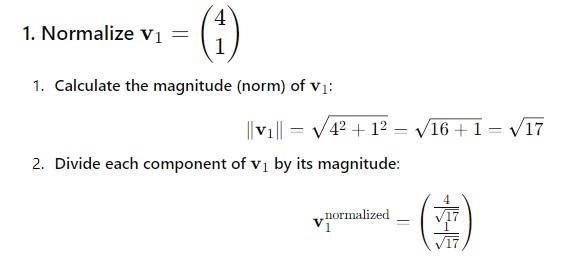
step 1:Find eigen vectors of each eigen value

Ex:



Step 2: now prints the given vectors in normalized form

Ex



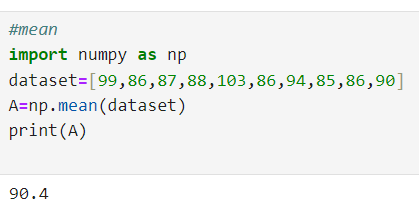
Output

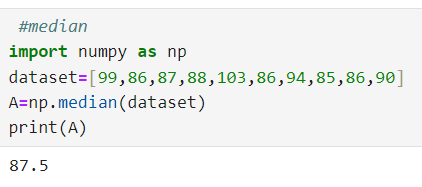
[[0.9701425 ,-0.70710678]

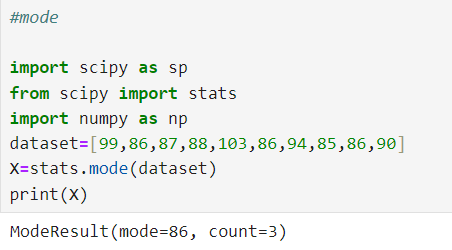
[0.24253563 ,0.70710678]]

**4b. Measure central tendency of Mean ,Median,and Mode use relevant python packages to compute central tendency for the parameter of given dataset=[99,86,87,88,103,86,94,85,86,90]**

**Ans:**

****

****

****

**(OR)**

**5a.Range,variance,standard deviation ,quantities and interquantities range use relevant python package to compute measure of dispution using numpy for the given data set values of [32,46,75,79,88,91,4,91,4,93,46]**

**Ans:**

#Write a program to find Variance, Standard Deviation and Zscore

import pandas as pd

import scipy.stats as sp

# Create a sample DataFrame

data = {

'A': [1, 2, 3, 4, 5, 5]

}

df = pd.DataFrame(data)

# Display the DataFrame print("DataFrame:")

print(df)

# Variance

print("\nVariance of each column:")

print(df.var())

# Standard Deviation

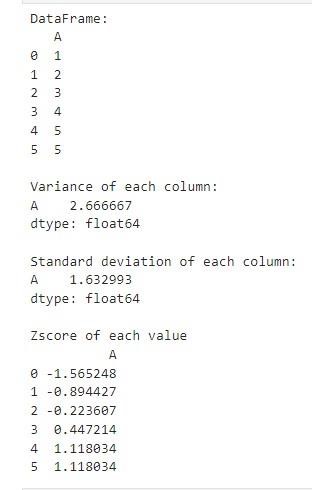
print("\nStandard deviation of each column:")

print(df.std())

#Zscore

print(“\nZscore of each value”) print(stats.zscore(df))

**OUTPUT**



**Logics**

**Variance** (Sample)

Variance measures the average of the squared differences between each data point and the mean of the dataset

print("\nVariance of each column:")

print(df.var())

Step 1: Finds the mean value for the data values

Ex: [1,2,3,4,5,5]=3.33

Step 2: Subtract the mean with each individual values and get squared difference

Ex: (1-3.33)^2=5.4289

(2-3.33)^2=1.7689

(3-3.33)^2=0.1089

(4-3.33)^2=0.4489

(5-3.33)^2=2.7889

(5-3.33)^2=2.7889

Step 3: Sum the squared differences and divide it with the number of elements minus one

Ex:

1. 5.4289+1.7689+0.1089+0.4489+2.7889+2.7889=13.3334
2. 13.3334/(6-1)
3. 13.3334/5=2.666667

**II.Standard Diviation**

Standard deviation is the square root of the variance and provides a measure of the average distance of each data point from the mean print("\nStandard deviation of each column:") print(df.std())

Step 1: It finds square root of variance and prints it

Ex: √2.666667=1.632993

**III. Zscore**

A Z-score, also known as a standard score, is a statistical measurement that describes a value's position relative to the mean of a group of values.

print(“\nZscore of each value”) print(stats.zscore(df))

Step 1: Calculate the mean

Ex: (1+2+3+4+5+5)/6=3.3333

Step 2: Calculate the Standard Deviation

Ex: Subtract the mean with each individual values and get squared difference

i. Ex: (1-3.3333)^2=5.4289

(2-3.3333)^2=1.7689

(3-3.3333)^2=0.1089

(4-3.3333)^2=0.4489

(5-3.3333)^2=2.7889

(5-3.3333)^2=2.7889

Sum the squared differences and divide it with the number of element to find Variance (sigma squared)

* + 1. 5.4289+1.7689+0.1089+0.4489+2.7889+2.7889=13.3334
    2. 13.3334/6=2.2222

Find square root of variance

* + 1. √2.2222=1.4907

Step 3: Now calculate the z-score

Ex: Formula (xi – mean)/std

* + 1. (1-3.3333)/1.4907= -1.5652

(2-3.3333)/1.4907= -0.8944

(3-3.3333)/1.4907= -0.2236

(4-3.3333)/1.4907= 0.4472

(5-3.3333)/1.4907= 1.1180

(5-3.3333)/1.4907= 1.1180

**5b.Write a python program and import scikit learn and load your data and handling missing values and finalize find handling categorical data.**

**ANS: **

OUTPUT:

