Variance :

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (xi - \mu)^2$$

$xi \rightarrow$ individual data point

$\bar{x}$ (or) $\mu \rightarrow$ ~~Median~~ Mean

$n \rightarrow$ Total no. of observation in the dataset.

Standar deviation:

$$\sigma = \sqrt{\sigma^2}$$

Eg :

[70, 75, 80, 85, 90]

1. Mean :

$$\text{Mean} = \frac{(70 + 75 + 80 + 85 + 90)}{5}$$

$$= 80.$$

2. variance :

$$\text{variance} = [\underbrace{(70-80)^2}_{x_i - \bar{x}} + (75-80)^2 + (80-80)^2 + (85-80)^2 + (90-80)^2] / 5$$

$$= [100 + 25 + 0 + 25 + 100] / 5$$

$$= 250/5$$

$$= 50.$$

3. SD :

$$SD = \sqrt{50}$$

$$= 7.07.$$

Interquartile Range (IQ):

$$IQR = Q3 - Q1.$$

$Q1 \rightarrow$ 25% of the Data falls

$Q3 \rightarrow$ 75% of the Data falls

Eg:

Dataset = [12, 15, 18, 20, 22, 25, 30, 35, 40, 50].

1. Arrange in Ascending order:

[12, 15, 18, 20, 22, 25, 30, 35, 40, 50].

2. calculate Q1 & Q3:

$Q1 \rightarrow$ is the median of first half.

$Q1 = 18.$

$Q3 \rightarrow$ is the median of the second half.

$Q3 = 35.$

3. calculate IQR.

$$IQR = Q3 - Q1$$
$$= 35 - 18$$
$$= 17.$$

Coefficient of Variation (CV):

$$CV = \frac{SD}{Mean} \times 100\%$$

Eg:

Consider two datasets:

Dataset A : Mean = 50, standard Deviation = 10.

Dataset B : " = 100, " = 20.

For Dataset A:

$$CV = \frac{10}{50} \times 100\% = 20\%.$$

For Dataset B:

$$CV = \frac{20}{100} \times 100\% = 20\%.$$

univariate and multivariate analyses are two distinct approaches to analyzing data, each focusing on different aspects of the data and using different methods to derive insights.

Aspect	univariate Analysis	Multivariate analysis.
Definition:	Analyzing a single variable at a time	simultaneously ananlysing Multiple variables.
Focus	Examines the distribution and characterstics of one variable.	studies relationships and interactions between multiple variables.
Purpose	Descriptive statistics, summarizing data.	Identifying patterns, associations, and dependencies among variables.
Examples	Mean, Median, Mode, standard deviation, histograms.	correlation analysis, regression analysis, principle component analysis cluster analysis.
Techniques	Measures of central Tendancy, Measures of dispersion, frequency distribution.	Regression models, factor analysis, principles component, analysis (PCA), discriminant analysis, Cluster analysis.
Inter pretation.	Provides insights into the behavior and characteristics of individual data variables.	Provides insights into the relationships and interactions among variables and their impact on outcomes.

Data Requirements	Requires data on a single variable	Requires data on multiple variables.
Visualization.	Typically represented using histograms, box plots, bar charts.	Represented using scatter plots, heatmaps, correlation matrix.

summary:

> While univariate analysis explores the characteristics of individual variables.

> Multivariate analysis delves into the relationships and interactions between multiple variables.

> Each approach serves distinct purposes and employs different techniques and visualization to derive insights from the data.

> Both univariate and multivariate analyses play crucial roles in data analysis and interpretation.

Scaling :

- Scaling Transforms the feature of a data set. so that they all fall within a specified range.
- Scaling adjusts the range of features without changing their distribution.

Standardization :

- It transforms the features of a dataset so that they have a mean of 0 and a standar deviation of 1.
- It centre the data around the mean and scales it to have a consistent standard deviation.

Normalization :

- Normalization scales the values of features to a range between 0 and 1.
- It adjusts the values of features based on their minimum and maximum values, making them fall within a specified range.

1. Scaling :

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import MinMaxScaler
                                 standard Scaler

# sample data
data = np.array([[1,2],[3,4],[5,6]])

# create a MinMaxScaler object:
scaler = MinMaxScaler() -> StandardScaler()

# Fit and Transform the data :
Scaled-data = scaler.fit.transform(data)
standardized

# plot orginal and scaled data.
plt.figure(figsize = (8,5)).
plt.scatter(data[:,0], data[:,1], color = 'blue',
            label = 'original Data')
                     Standardized-data
plt.scatter(Scaled-data[:,0], Scaled-data[:,1],
            color = 'red', label = 'scaled Data')
                                      Standardized
plt.title('scaling')
```

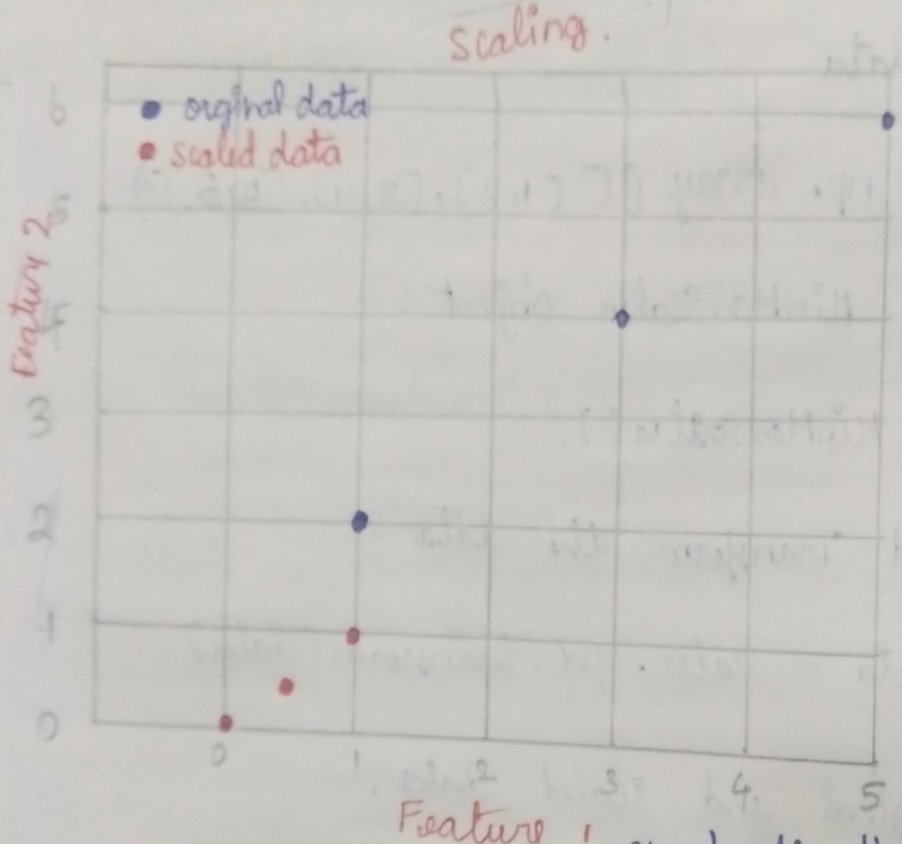
plt.xlabel ('Feature 1')

plt.ylabel ('Feature 2')

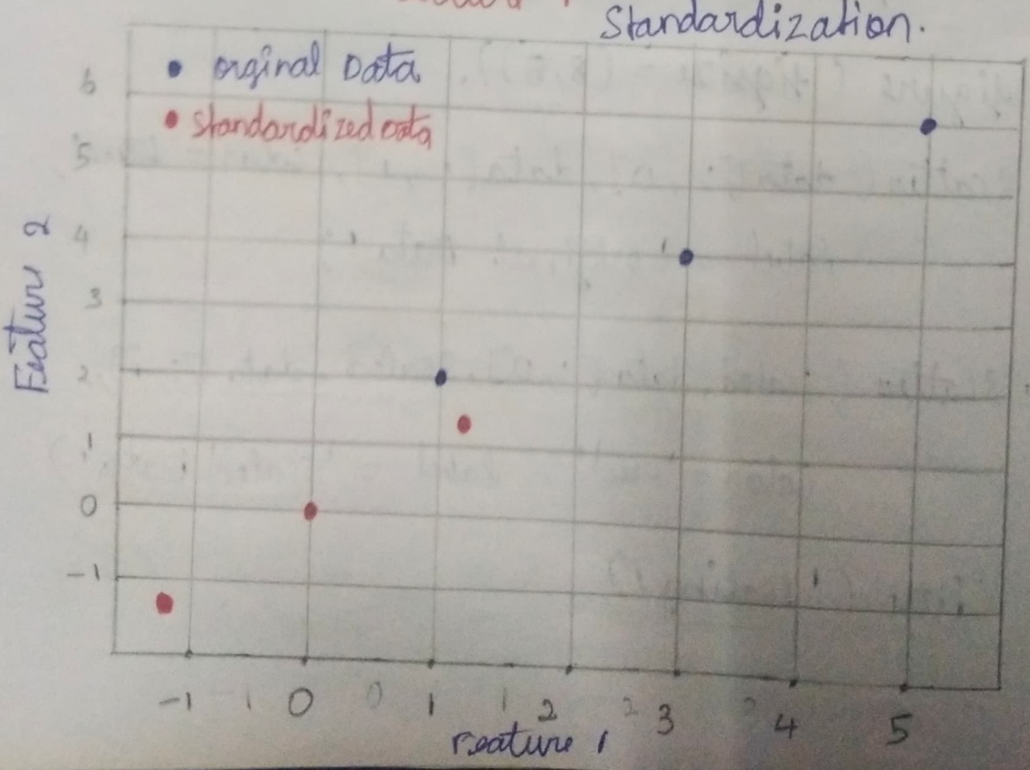
plt.legend ()

plt.grid (True)

plt.show ().



⇒ Same for Normalization



Normalization :

```
import matplotlib.pyplot as plt
import numpy as np.
from sklearn.preprocessing import MinMaxScaler

data = np.array([[1,2],[3,4],[5,6]])

Scaler = MinMaxScaler()

Normalized_data = scaler.fit_transform(data)

plt.figure (figsize = (8,5)).

plt.scatter (data [:,0], data [:,1], color = 'blue',
                label = 'orginal Data').

plt.Scatter (normalized_data [:,0], normalized_
            data , color = 'red' , label = 'Normalized
                                          data').

plt.title ('Normalization')
plt.xlabel ('Feature 1')
plt.ylable ('Feature 2')
plt.legend ()
plt.grid (True)
plt.show ().
```