**Univariate and Bivariate Analysis in Machine Learning**

Univariate Analysis: It involves the analysis of a single variable. The main purpose is to describe the variable by summarizing its central tendency, dispersion, and shape.

- Summarizes and describes the main features of a single variable.

- Helps in understanding the distribution, central tendency, and dispersion of the variable.

- Useful for detecting outliers and understanding the overall pattern.

**Univariate Analysis Techniques:**

1. Summary Statistics: Mean, median, mode, standard deviation, variance, range, etc.

2. Visualizations: Histograms, box plots, density plots, etc.

**Example:**

python

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load dataset

df = sns.load\_dataset('iris')

# Summary Statistics

print(df['sepal\_length'].describe())

# Histogram

plt.figure(figsize=(10, 5))

sns.histplot(df['sepal\_length'], kde=True)

plt.title('Histogram of Sepal Length')

plt.show()

# Box Plot

plt.figure(figsize=(10, 5))

sns.boxplot(y=df['sepal\_length'])

plt.title('Box Plot of Sepal Length')

plt.show()

**Bivariate Analysis:** It involves the analysis of two variables simultaneously. The main purpose is to understand the relationship between the two variables.

- Explores the relationship between two variables.

- Helps in identifying correlations and potential causal relationships.

- Useful for detecting patterns and trends that involve two variables.

**Bivariate Analysis Techniques:**

1. Summary Statistics: Correlation coefficient, covariance, etc.

2. Visualizations: Scatter plots, heatmaps, pair plots, etc.

*Example:*

python

# Scatter Plot

plt.figure(figsize=(10, 5))

sns.scatterplot(x='sepal\_length', y='sepal\_width', data=df)

plt.title('Scatter Plot of Sepal Length vs Sepal Width')

plt.show()

# Correlation Heatmap

plt.figure(figsize=(10, 5))

sns.heatmap(df.corr(), annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Correlation Heatmap')

plt.show()

# Pair Plot

sns.pairplot(df)

plt.show()

**Task for Students**

1. Load the Iris dataset into a Pandas DataFrame.

2. Perform univariate analysis on all numeric columns (sepal length, sepal width, petal length, petal width).

- Calculate summary statistics.

- Create histograms and box plots for each variable.

3. Perform bivariate analysis to explore relationships between pairs of variables.

- Create scatter plots for pairs of variables.

- Calculate and visualize the correlation matrix.

- Create a pair plot to visualize relationships between all pairs of variables.

**Example Code to Start:**

python

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load dataset

df = sns.load\_dataset('iris')

# Univariate Analysis

# Summary Statistics

print(df.describe())

# Histograms and Box Plots

for column in df.select\_dtypes(include=['float64']).columns:

plt.figure(figsize=(10, 5))

sns.histplot(df[column], kde=True)

plt.title(f'Histogram of {column}')

plt.show()

plt.figure(figsize=(10, 5))

sns.boxplot(y=df[column])

plt.title(f'Box Plot of {column}')

plt.show()

# Bivariate Analysis

# Scatter Plots

sns.pairplot(df)

plt.show()

# Correlation Heatmap

plt.figure(figsize=(10, 5))

sns.heatmap(df.corr(), annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Correlation Heatmap')

plt.show()