**Assignment No. 3**

**Name: Aditya Wanave**

**PRN No: 122B1F135**

**Problem Statement:**

1. **Find the Correlation Matrix on the Iris Dataset.**  
   **b. Plot the Correlation Plot on the Dataset and Visualize Relationships Among the Features of the Iris Dataset.**

**Objective:**  
The objective of this assignment is to calculate the **correlation matrix** of the Iris dataset and visualize it using a **correlation plot**. The correlation matrix will help us understand the relationships between the features (sepal length, sepal width, petal length, petal width) in the dataset, and the plot will provide a visual representation of these relationships.

**Prerequisites:**

1. **Python environment** with the following libraries:
   * **pandas** for loading and processing the dataset.
   * **numpy** for numerical operations.
   * **seaborn** and **matplotlib** for plotting the correlation matrix.
2. A **text editor** or **IDE** to write and execute Python code.
3. **Basic understanding** of correlation and data visualization concepts.

**Theory:**

* **Correlation** measures the strength and direction of a linear relationship between two features. The correlation coefficient varies between -1 and 1:
  + **+1**: Perfect positive correlation.
  + **-1**: Perfect negative correlation.
  + **0**: No linear relationship.
* The **correlation matrix** is a table that shows the correlation coefficients between all possible pairs of features in the dataset.
* A **correlation plot** (heatmap) visually represents the correlation matrix, making it easier to identify patterns, positive or negative relationships, and the magnitude of these relationships.

**Algorithm for Calculating Correlation Matrix and Plotting Correlation Plot:**

**Step 1: Import necessary libraries**

* Load the **Iris dataset** into a pandas DataFrame.

**Step 2: Calculate the correlation matrix**

* Use pandas' corr() function to compute the correlation matrix.

**Step 3: Visualize the correlation matrix using a heatmap**

* Use the seaborn library to create a heatmap of the correlation matrix.

**Python Script Example:**

*# Step 1: Import necessary libraries*

*import pandas as pd*

*import seaborn as sns*

*import matplotlib.pyplot as plt*

*from sklearn.datasets import load\_iris*

*# Load the Iris dataset into a pandas DataFrame*

*iris = load\_iris()*

*df\_iris = pd.DataFrame(data=iris.data, columns=iris.feature\_names)*

*# Step 2: Calculate the correlation matrix*

*correlation\_matrix = df\_iris.corr()*

*print("Correlation Matrix:")*

*print(correlation\_matrix)*

*# Step 3: Plot the correlation matrix as a heatmap*

*plt.figure(figsize=(8, 6))*

*sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', linewidths=0.5)*

*plt.title("Correlation Heatmap of Iris Dataset")*

*plt.show()*

**Step-by-step explanation:**

1. **Loading the Dataset:**
   * The **Iris dataset** is loaded using sklearn.datasets.load\_iris() and converted into a pandas DataFrame for easy manipulation.
2. **Calculating the Correlation Matrix:**
   * The corr() function computes the **correlation matrix**, which returns the pairwise correlation coefficients between all features in the dataset.
   * The correlation values will range from -1 to 1, indicating the strength and direction of the linear relationships between features.
3. **Plotting the Correlation Matrix:**
   * Using the seaborn library, a **heatmap** is created to visualize the correlation matrix. The annot=True parameter displays the correlation values on the heatmap, and the cmap='coolwarm' argument is used for coloring the matrix based on the magnitude of the correlation values.
   * The heatmap provides a quick, visual representation of which features are positively or negatively correlated.

**Output Example:**

**1. Correlation Matrix:**

|  | **Sepal Length** | **Sepal Width** | **Petal Length** | **Petal Width** |
| --- | --- | --- | --- | --- |
| Sepal Length | 1.00 | -0.12 | 0.87 | 0.82 |
| Sepal Width | -0.12 | 1.00 | -0.37 | -0.37 |
| Petal Length | 0.87 | -0.37 | 1.00 | 0.96 |
| Petal Width | 0.82 | -0.37 | 0.96 | 1.00 |

**2. Correlation Plot (Heatmap):**

The heatmap will show strong positive correlations in red and negative correlations in blue, helping us easily visualize the relationships between the features.

**References:**

* Pandas Documentation
* Seaborn Documentation
* Matplotlib Documentation
* Scikit-learn Documentation

**Conclusion:**

In this assignment, we successfully calculated the correlation matrix for the Iris dataset and visualized it using a heatmap. This analysis provides valuable insights into the relationships between different features of the dataset, such as the strong positive correlation between petal length and petal width. The correlation plot enables us to quickly grasp these relationships, which can be crucial for further data analysis, feature selection, or modeling tasks.