Data Visualization Techniques Lab

Pulla Manoj Kumar
Assistant Professor, CSD, ACEEC Hyderabad

Statistical Analysis
such as Multivariate Analysis, PCA, LDA, Correlation, Regression, and Analysis of Variance.

STEP 1: Import Necessary Libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.datasets import load_wine
from sklearn.decomposition import PCA
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.linear_model import LinearRegression
from scipy.stats import f_oneway

sns.set(style="whitegrid")
```

STEP 2: Load the Wine Dataset

```
wine = load_wine()
df = pd.DataFrame(wine.data, columns=wine.feature_names)
```

STEP 3: Display Dataset Information

```
print("Shape of dataset:", df.shape)
print(df.info())
df.describe()
```

STEP 4: Check for Missing Values

```
df.isnull().sum()
```

STEP 5: Correlation Analysis (Heatmap)

```
plt.figure(figsize=(12,8))
sns.heatmap(df.corr(), annot=True, cmap="coolwarm")
plt.title("Wine Dataset Correlation Heatmap")
plt.show()
```

STEP 6: Apply PCA (Principal Component Analysis)

```
pca = PCA(n_components=2)
pca_components = pca.fit_transform(df)
```

STEP 7: PCA Visualization

```
plt.scatter(
    pca_components[:, 0],
    pca_components[:, 1],
    c=wine.target,
    cmap='rainbow'
)
plt.xlabel("Principal Component 1")
plt.ylabel("Principal Component 2")
plt.title("PCA Visualization of Wine Dataset")
plt.show()
```

Linear Discriminant Analysis (LDA)

STEP 8: Prepare Data for LDA

```
features = wine.data
target = wine.target
```

STEP 9: Apply LDA

```
lda = LinearDiscriminantAnalysis(n_components=2)
lda_transformed = lda.fit_transform(features, target)
```

STEP 10: LDA Visualization

```
plt.scatter(
    lda_transformed[:, 0],
    lda_transformed[:, 1],
    c=target,
    cmap='rainbow'
)
plt.xlabel("LDA Component 1")
plt.ylabel("LDA Component 2")
plt.title("LDA Visualization of Wine Dataset")
plt.show()
```

Regression Analysis

STEP 11: Simple Linear Regression

- Alcohol vs Malic Acid relationship.

```
X = df[['alcohol']]
y = df['malic_acid']

model = LinearRegression()
model.fit(X, y)

y_pred = model.predict(X)
```

STEP 12: Regression Visualization

```
plt.scatter(X, y, label='Actual Data')
plt.plot(X, y_pred, label='Regression Line')
plt.xlabel("Alcohol")
plt.ylabel("Malic Acid")
plt.title("Linear Regression Analysis")
plt.legend()
plt.show()
```

ANOVA – Analysis of Variance

STEP 13: Add Target Column

```
df['target'] = wine.target
```

STEP 14: Perform One-Way ANOVA Test

```
class0 = df[df['target'] == 0]['alcohol']
class1 = df[df['target'] == 1]['alcohol']
class2 = df[df['target'] == 2]['alcohol']

f_stat, p_value = f_oneway(class0, class1, class2)

print("F-statistic:", f_stat)
print("P-value:", p_value)
```

Interpretation

- If $p < 0.05$: Significant difference exists
- If $p \geq 0.05$: No significant difference

STEP 15: ANOVA Visualization

```
sns.boxplot(x='target', y='alcohol', data=df)
plt.xlabel("Wine Class")
plt.ylabel("Alcohol Content")
plt.title("ANOVA: Alcohol Content Across Wine Classes")
plt.show()
```