

▼ Anvesh Khode

PRN: 22070521021 | Practical 3: Principal Component Analysis (PCA)

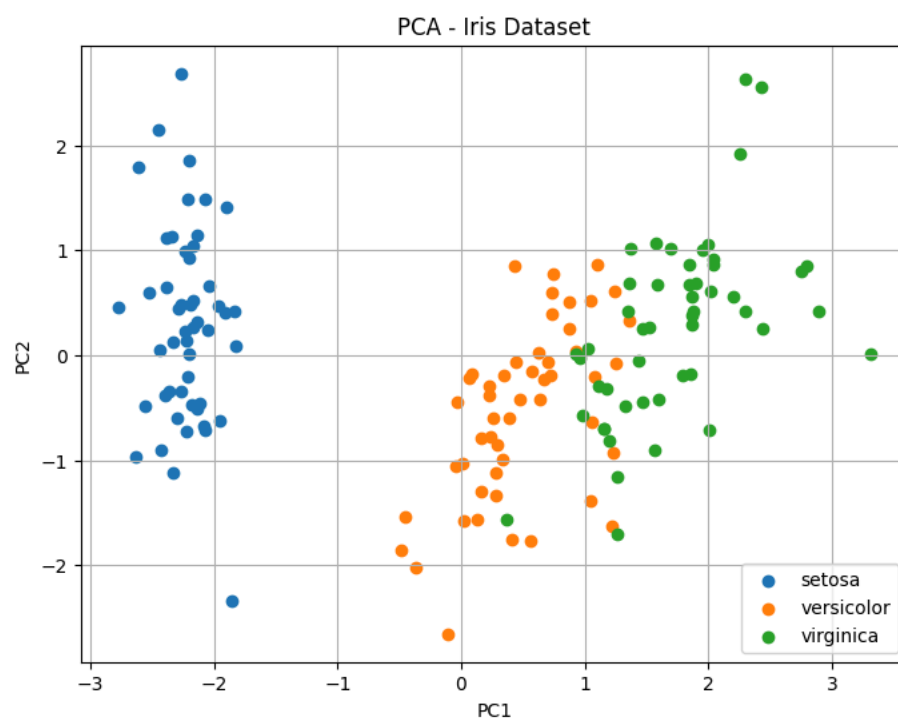
```
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.datasets import load_iris
from sklearn.preprocessing import StandardScaler
```

```
# Load the Iris dataset
iris = load_iris()
X = iris.data
y = iris.target
target_names = iris.target_names
```

```
# Standardize the features
X_std = StandardScaler().fit_transform(X)
```

```
# Perform PCA
pca = PCA()
X_pca = pca.fit_transform(X_std)
```

```
# Scatterplot for first two principal components
plt.figure(figsize=(8,6))
for i, target_name in enumerate(target_names):
    plt.scatter(X_pca[y == i, 0], X_pca[y == i, 1], label=target_name)
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.title("PCA - Iris Dataset")
plt.legend()
plt.grid(True)
plt.show()
```



```
# Plot explained variance with cumulative variance as a straight line (not staircase)
plt.figure(figsize=(8,6))
explained_variance = pca.explained_variance_ratio_
cumulative_variance = explained_variance.cumsum()

# Bar plot
plt.bar(range(1, len(explained_variance)+1), explained_variance, alpha=0.5, align="center", label="Individual explained variance")

# Cumulative line as straight connections between points
plt.plot(range(1, len(cumulative_variance)+1), cumulative_variance, marker='o', label="Cumulative explained variance")
```

```
# Annotate cumulative variance values on the line
for i, val in enumerate(cumulative_variance, start=1):
    plt.text(i, val, f"{val:.2f}", ha='center', va='bottom')

plt.xlabel("Principal Components")
plt.ylabel("Explained Variance Ratio")
plt.title("Explained Variance by Principal Components - Iris Dataset")
plt.legend()
plt.show()
```

