PAT-4

Shylesh Kumar. S

23MCA0346


```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
from sklearn.decomposition import PCA



iris = load_iris()
df = pd.DataFrame(iris.data, columns=iris.feature_names)



# Check for missing values
print("Missing values:\n", df.isnull().sum())
```

```
Missing values:
 sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
dtype: int64
```

```
#KMean
k = 3
initial_centroids = df.sample(n=k, random_state=42)
max_iterations = 100
prev_centroids = None
iteration = 0

while iteration < max_iterations:
    distances = []
    for i in range(len(df)):
        distances.append([((df.iloc[i] - c) ** 2).sum() for c in initial_centroids.values])
    cluster_labels = pd.DataFrame(distances).idxmin(axis=1)
    prev_centroids = initial_centroids.copy()
    for i in range(k):
        cluster_points = df[cluster_labels == i]
        initial_centroids.iloc[i] = cluster_points.mean()
    if prev_centroids.equals(initial_centroids):
        break
    iteration += 1

cluster_labels = pd.DataFrame(distances).idxmin(axis=1)
centroids = initial_centroids.values



#performance:
true_labels = iris.target
ari = adjusted_rand_score(true_labels, cluster_labels)
print("Adjusted Rand Index (ARI):", ari)
```

```
Adjusted Rand Index (ARI): 0.7302382722834697
```

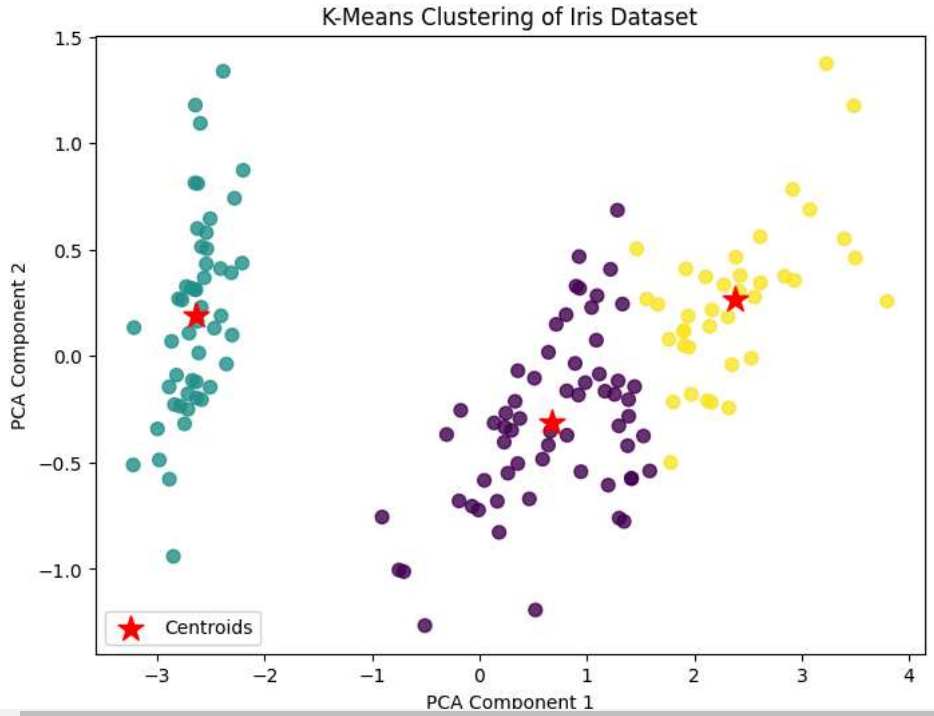
```
# Apply PCA for dimensionality reduction
pca = PCA(n_components=2)
iris_2d = pca.fit_transform(df)
reduced=pca.transform(centroids)

#Plotting:
plt.figure(figsize=(8, 6))
plt.scatter(iris_2d[:, 0], iris_2d[:, 1], c=cluster_labels, cmap='viridis', s=50, alpha=0.8)

plt.scatter(reduced[:, 0], reduced[:, 1], c='red', marker='*', s=200, label='Centroids')
```

```
plt.title('K-Means Clustering of Iris Dataset')
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
plt.legend()
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but PCA was fitted with f
  warnings.warn(
```



Start coding or generate with AI.