

EX: 7  
16/9/25

## CLUSTERING

AIM:

To perform kmeans on customer segmentation dataset. Apply kmeans clustering on a customer group

CODE:

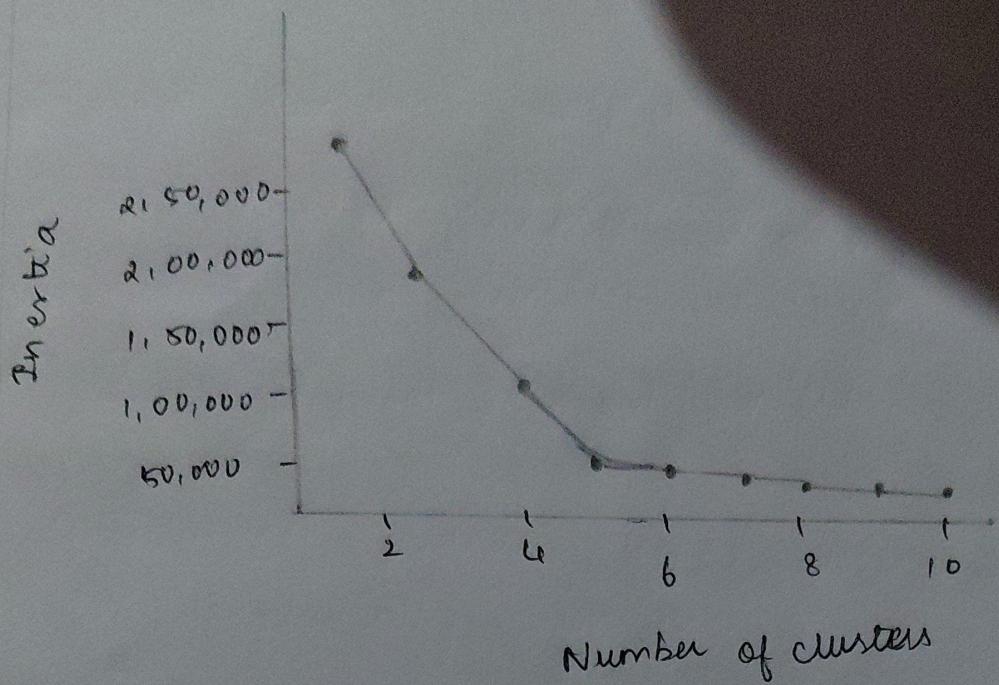
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import seaborn as sns

df = pd.read_csv('Mall_Customers.csv')
kmeans = KMeans(n_clusters=5, random_state=42)
df['Cluster'] = kmeans.fit_predict(df[['Annual Income (k$)', 'Spending Score (1-100)']])

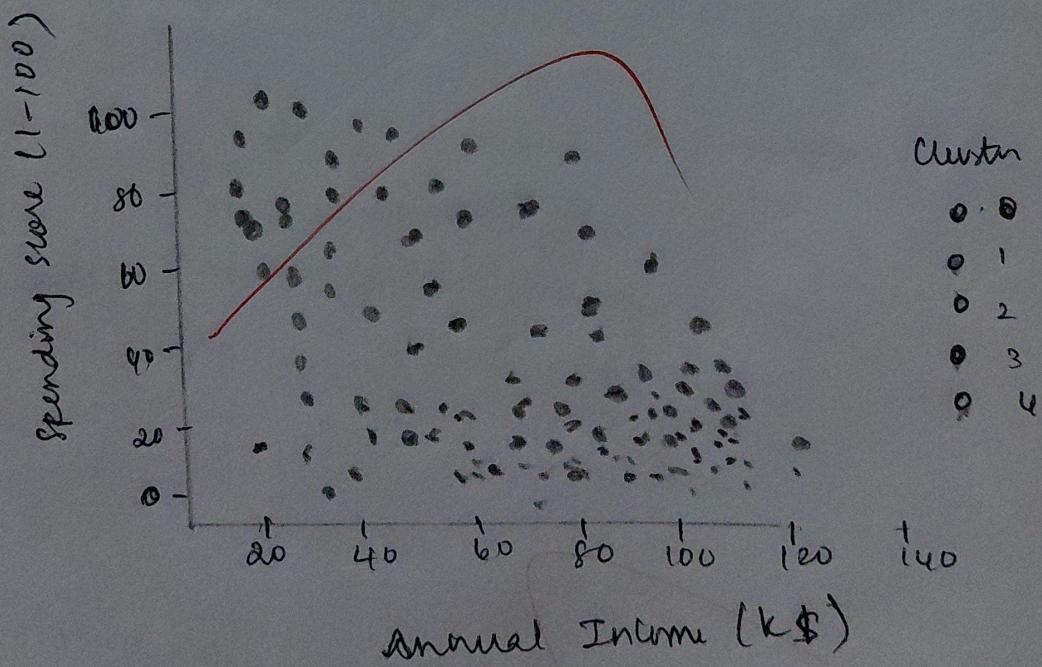
distortions = []
for i in range(1,11):
    km = KMeans(n_clusters=i)
    km.fit(df[['Annual Income (k$)', 'Spending Score (1-100)']])
    distortions.append(km.inertia_)

plt.plot(range(1,11), distortions, marker='o')
plt.title("Elbow Method")
plt.show()
```

## ELBOW METHOD



K-Means clustering on Mall Customers



```
sns.scatterplot(data=df, x='Annual Income ($)',  
y='Spending Score (1-100)', hue='Color',  
palette='Set2')
```

```
from sklearn.datasets import load_wine  
from sklearn.preprocessing import  
wine = load_wine()  
X = pd.DataFrame(wine.data, columns=  
wine.feature_names)
```

```
X_scaled = StandardScaler().fit_transform(X)
```

```
base_clustering = []
```

```
for k in [3, 4, 5]:
```

```
kmeans = KMeans(n_clusters=k, random_  
state=42)
```

for clustering in clusterings:

```
for i in range(n_samples):
```

```
for j in range(n_samples):
```

```
if clustering[i] == clustering[j]:
```

```
similarity_matrix[i][j] += 1
```

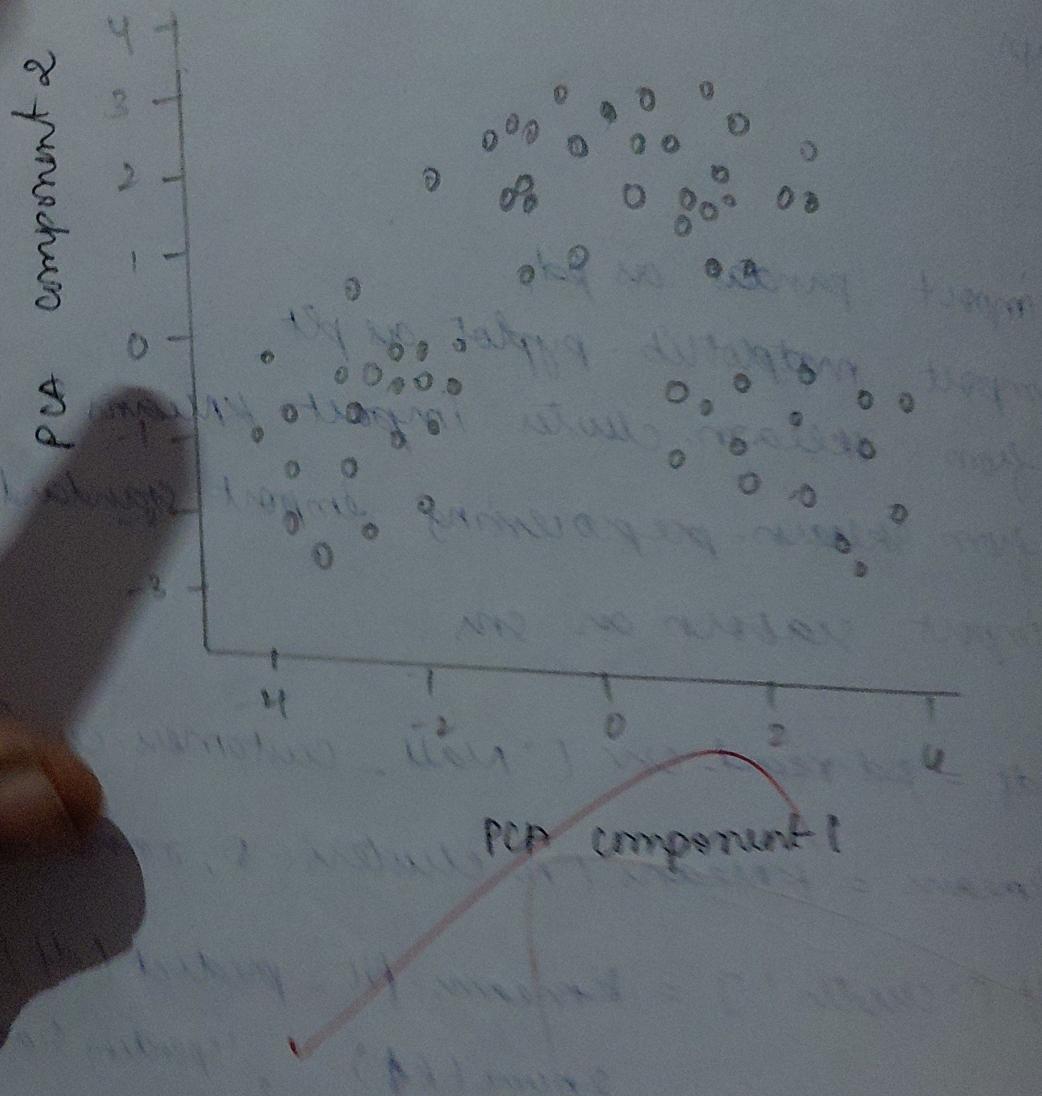
```
return ensemble_labels
```

```
pca = PCA(n_components=2)
```

```
X_pca = pca.fit_transform(X_scaled)
```

Silhouette Score : 0.884280322

PCA Ensemble Clustering on Wine dataset



```
plt.figure(figsize=(10, 6))
```

```
plt.scatter(x_pca[:, 0], x_pca[:, 1], c=ensemble_labels,  
cmap='viridis', s=50, edgecolor='k')
```

```
plt.title("PCA Ensemble clustering on Wine dataset")
```

```
plt.xlabel("PCA Component 1")
```

```
plt.ylabel("PCA Component 2")
```

```
plt.colorbar(label="Cluster Label")
```

```
plt.grid(True)
```

```
plt.show()
```



Result:

\* ~~disclaimer~~ the clustering is applied in  
mall - customer data and executed it  
successfully.