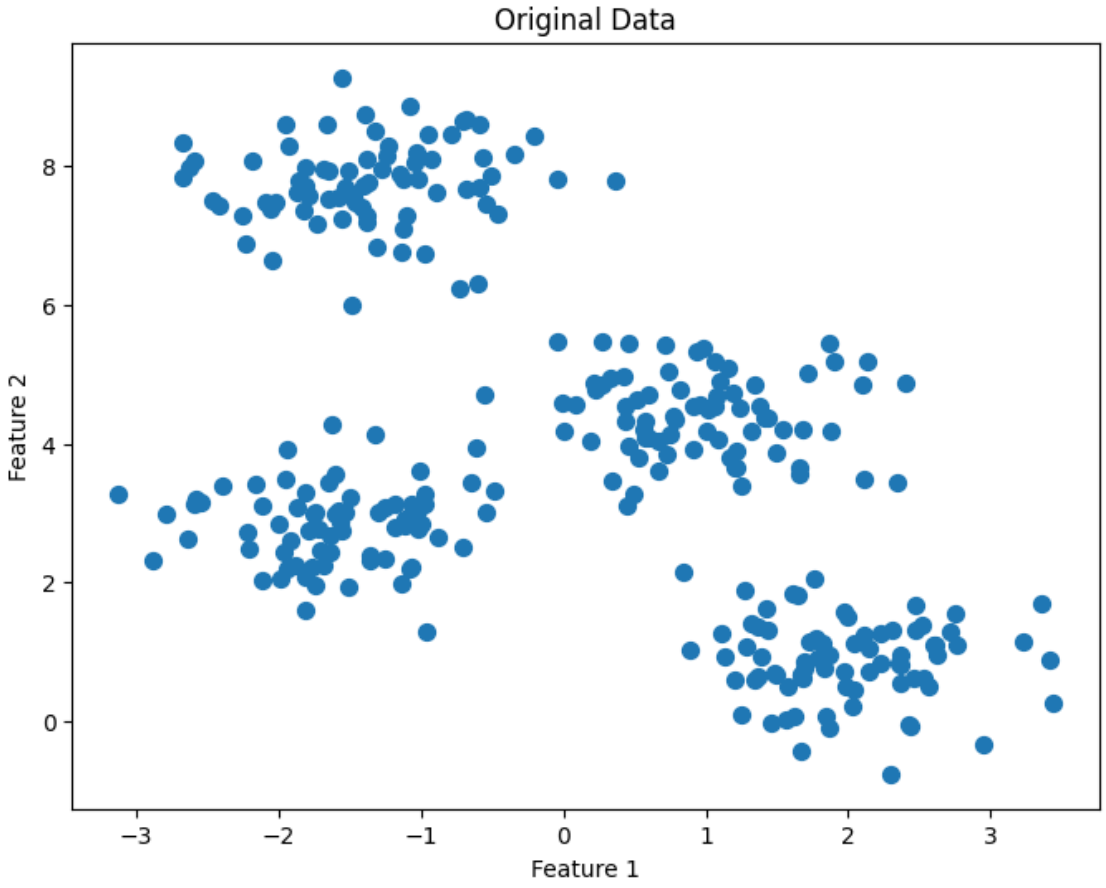
make-blobs

June 20, 2024

[1]:
```
# Import necessary libraries
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import numpy as np

# Generate synthetic dataset
X, _ = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)

# Plot original data
plt.figure(figsize=(8, 6))
plt.scatter(X[:, 0], X[:, 1], s=50, cmap='viridis')
plt.title('Original Data')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()

# Perform Exploratory Data Analysis (EDA)
# Compute basic statistics
print(f"Dataset shape: {X.shape}")
print(f"Min values: {np.min(X, axis=0)}")
print(f"Max values: {np.max(X, axis=0)}")
print(f"Mean values: {np.mean(X, axis=0)}")
print(f"Standard deviation: {np.std(X, axis=0)}")

# Train KMeans model
kmeans = KMeans(n_clusters=4, random_state=42)
kmeans.fit(X)

# Plot clustered data
plt.figure(figsize=(8, 6))
plt.scatter(X[:, 0], X[:, 1], c=kmeans.labels_, s=50, cmap='viridis')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1],␣
 ↪c='red', s=200, alpha=0.75)
plt.title('Clustered Data with KMeans')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
```

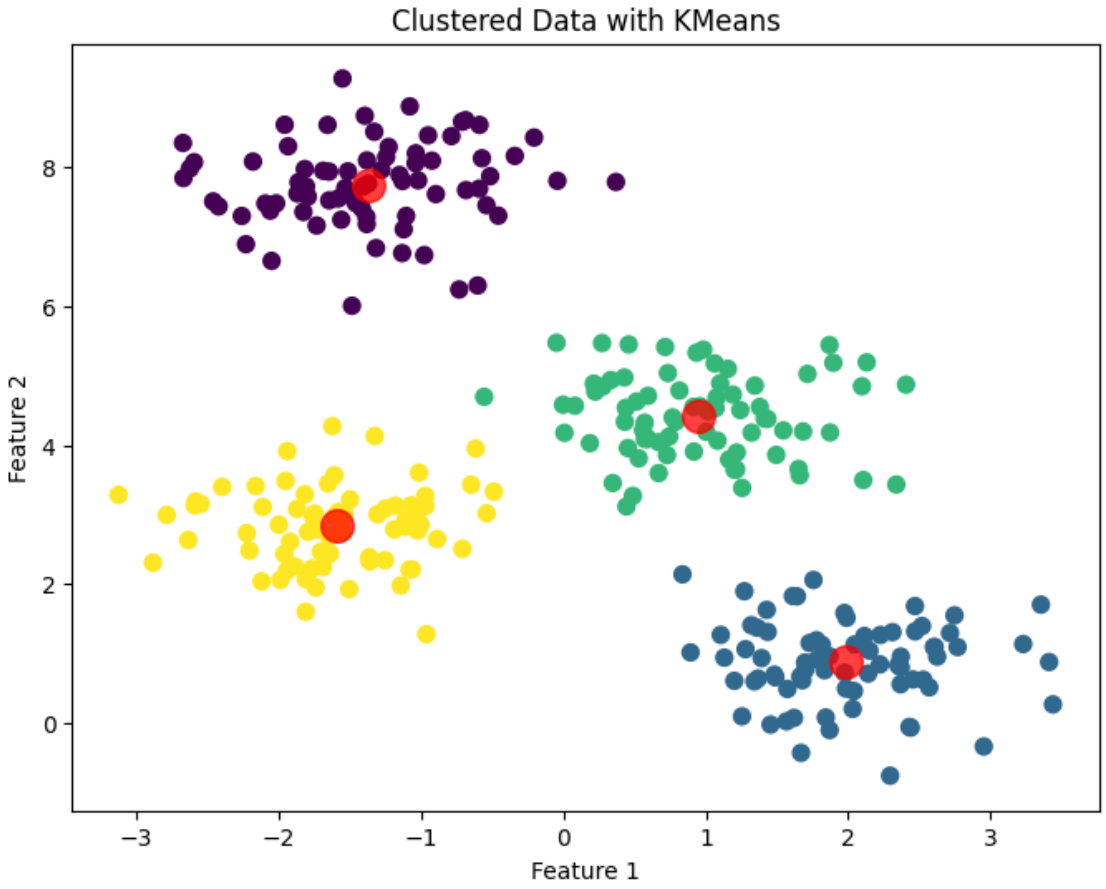
```
plt.show()

# Analyze cluster centers
print("\nCluster Centers (Centroids):")
for i, center in enumerate(kmeans.cluster_centers_):
    print(f"Cluster {i}: {center}")

# Analyze cluster assignments
unique, counts = np.unique(kmeans.labels_, return_counts=True)
print("\nCluster Assignments:")
for cluster, count in zip(unique, counts):
    print(f"Cluster {cluster}: {count} points")

# Evaluate model performance (Inertia)
print(f"\nKMeans Inertia: {kmeans.inertia_}")
```

```
<ipython-input-1-0af09e827a1a>:12: UserWarning: No data for colormapping
provided via 'c'. Parameters 'cmap' will be ignored
  plt.scatter(X[:, 0], X[:, 1], s=50, cmap='viridis')
```



```
Dataset shape: (300, 2)
Min values: [-3.12240736 -0.76589199]
Max values: [3.43761754 9.28293222]
Mean values: [-0.00632763  3.96782089]
Standard deviation: [1.63327104 2.58963659]
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```




```
Cluster Centers (Centroids):
Cluster 0: [-1.37324398  7.75368871]
Cluster 1: [1.98258281 0.86771314]
Cluster 2: [0.94973532 4.41906906]
Cluster 3: [-1.58438467  2.83081263]

Cluster Assignments:
Cluster 0: 75 points
```

```
Cluster 1: 75 points
Cluster 2: 75 points
Cluster 3: 75 points

KMeans Inertia: 212.00599621083484
```