
Name: Tufan Kundu

Registration no: 24MDY0184

Experiment: 11

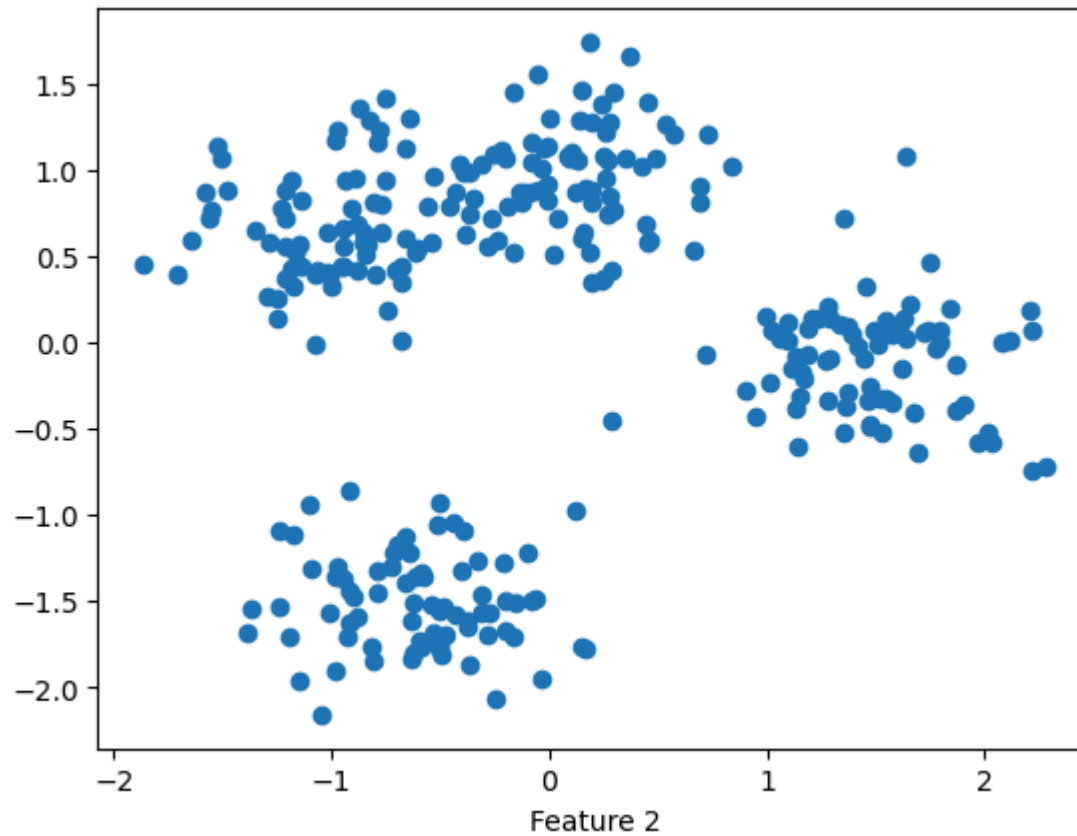
Clustering

Creating synthetic dataset

```
In [34]: import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans, AgglomerativeClustering, DBSCAN
from sklearn.datasets import make_blobs
from sklearn.preprocessing import StandardScaler
from scipy.cluster.hierarchy import dendrogram, linkage
```

```
In [4]: x, _ = make_blobs(n_samples = 300, centers = 4, cluster_std = 2.0, random_state=42)
x = StandardScaler().fit_transform(x)
```

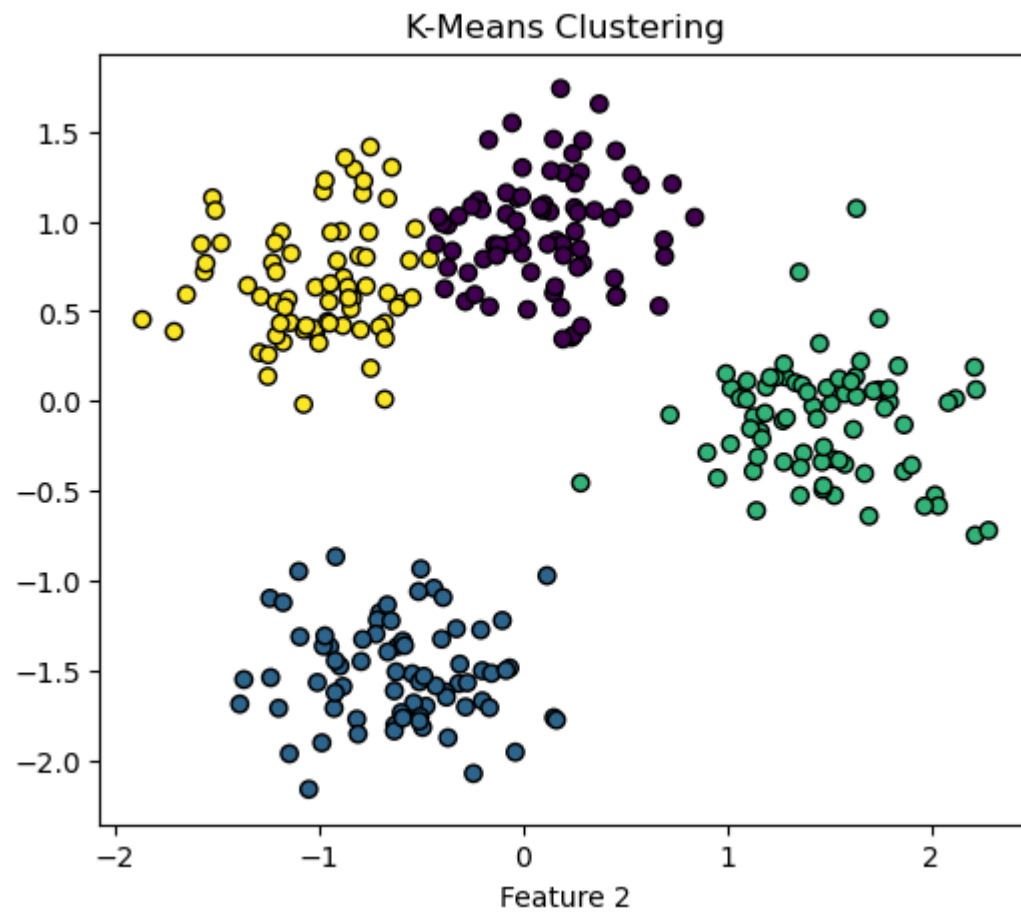
```
In [9]: plt.scatter(x[:,0],x[:,1], label = "Data points")  
plt.xlabel("Feature 1")  
plt.xlabel("Feature 2")  
plt.show()
```



```
In [12]: k = 4
kmeans = KMeans(n_clusters = k, random_state=42)
kmeans_labels = kmeans.fit_predict(x)

plt.figure(figsize = (6,5))
plt.scatter(x[:,0],x[:,1], c = kmeans_labels, cmap='viridis', edgecolor = 'k')
plt.title("K-Means Clustering")
plt.xlabel("Feature 1")
plt.xlabel("Feature 2")
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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
  super()._check_params_vs_input(X, default_n_init=10)
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1436: UserWarning: KMeans is known to have a memo
ry leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environ
ment variable OMP_NUM_THREADS=2.
  warnings.warn(
```



```
In [14]: # Reducing k to see the difference
k = 3
kmeans = KMeans(n_clusters = k, random_state=42)
kmeans_labels = kmeans.fit_predict(x)

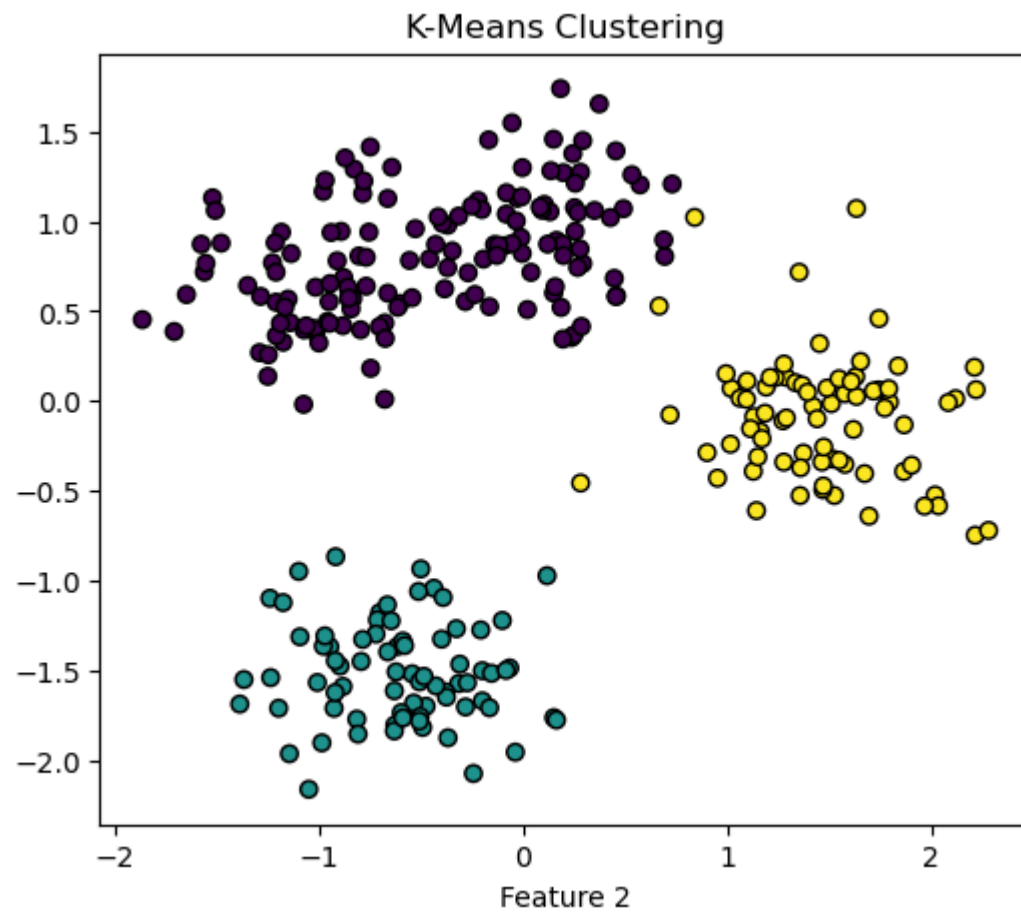
plt.figure(figsize = (6,5))
plt.scatter(x[:,0],x[:,1], c = kmeans_labels, cmap='viridis', edgecolor = 'k')
plt.title("K-Means Clustering")
plt.xlabel("Feature 1")
plt.xlabel("Feature 2")
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: 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
```

```
    super()._check_params_vs_input(X, default_n_init=10)
```

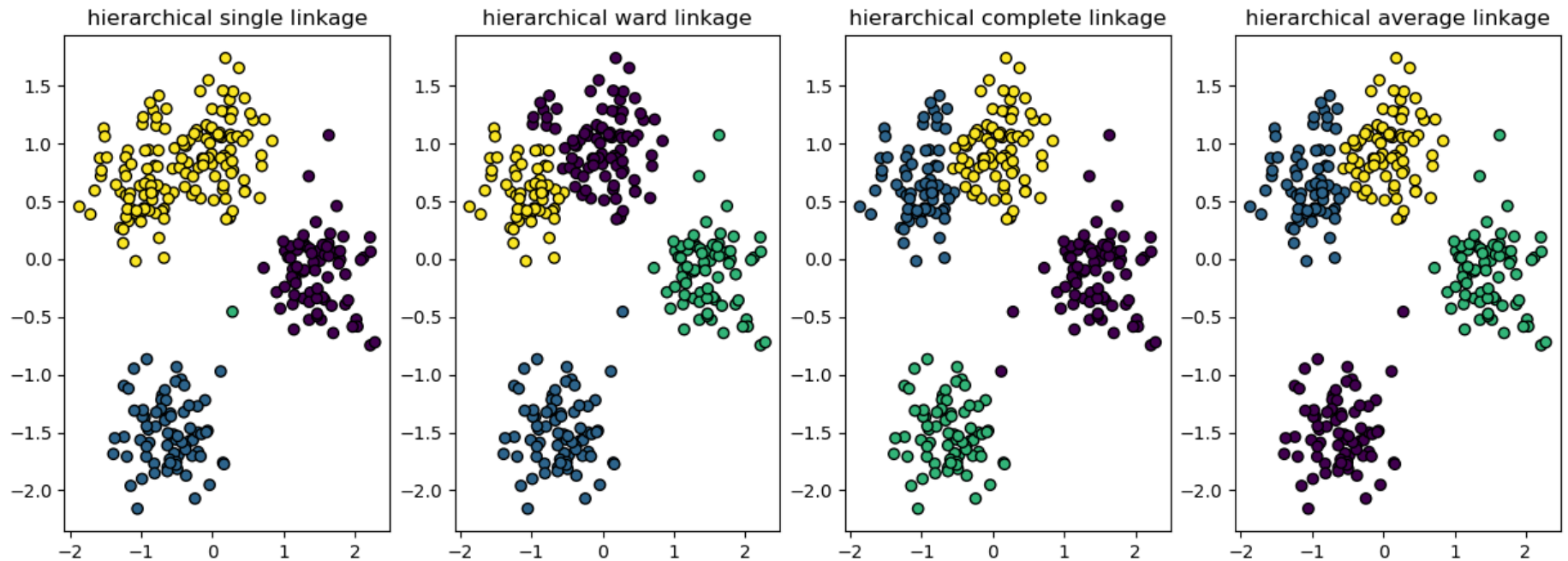
```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1436: UserWarning: KMeans is known to have a memo
ry leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environ
ment variable OMP_NUM_THREADS=2.
```

```
    warnings.warn(
```



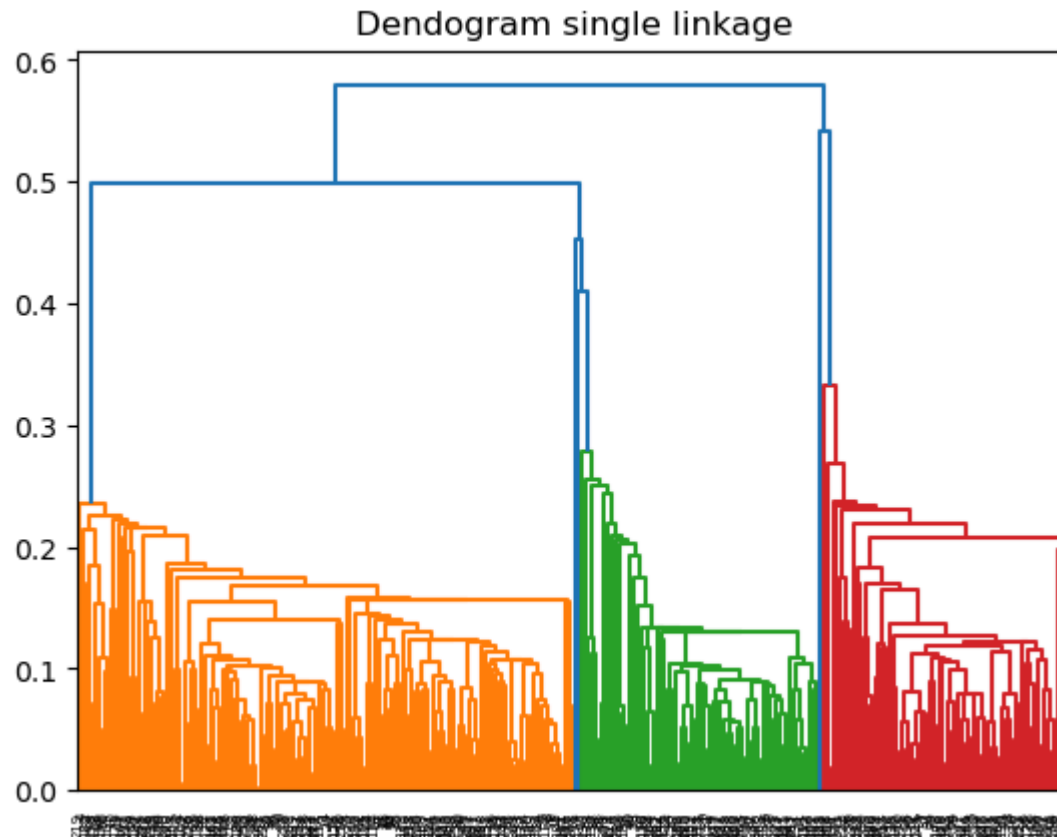
Agglomerative hierarchical clustering

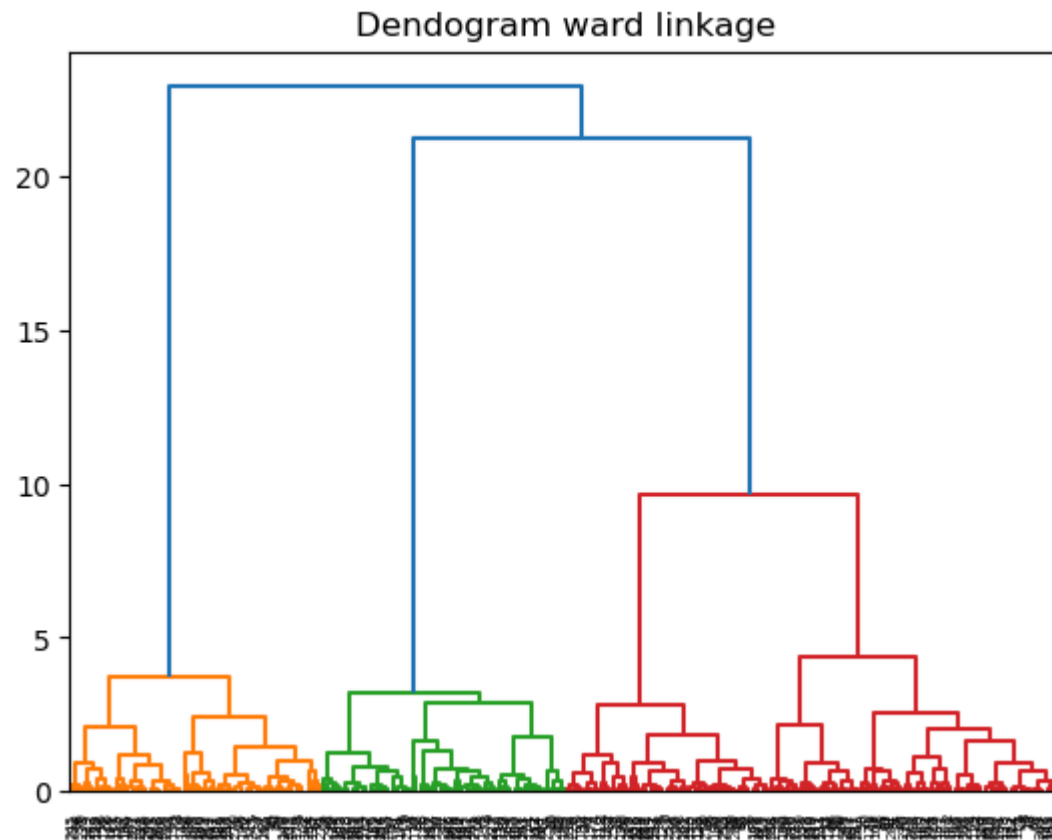
```
In [25]: linkage_methods = ['single', 'ward', 'complete', 'average']
fig, axes = plt.subplots(1, 4, figsize = (15, 5))
for i, method in enumerate(linkage_methods):
    hierarchical = AgglomerativeClustering(n_clusters=4, linkage = method)
    labels = hierarchical.fit_predict(x)
    axes[i].scatter(x[:,0], x[:,1], c = labels, cmap = 'viridis', edgecolor = 'k')
    axes[i].set_title(f"hierarchical {method} linkage")
plt.show()
```

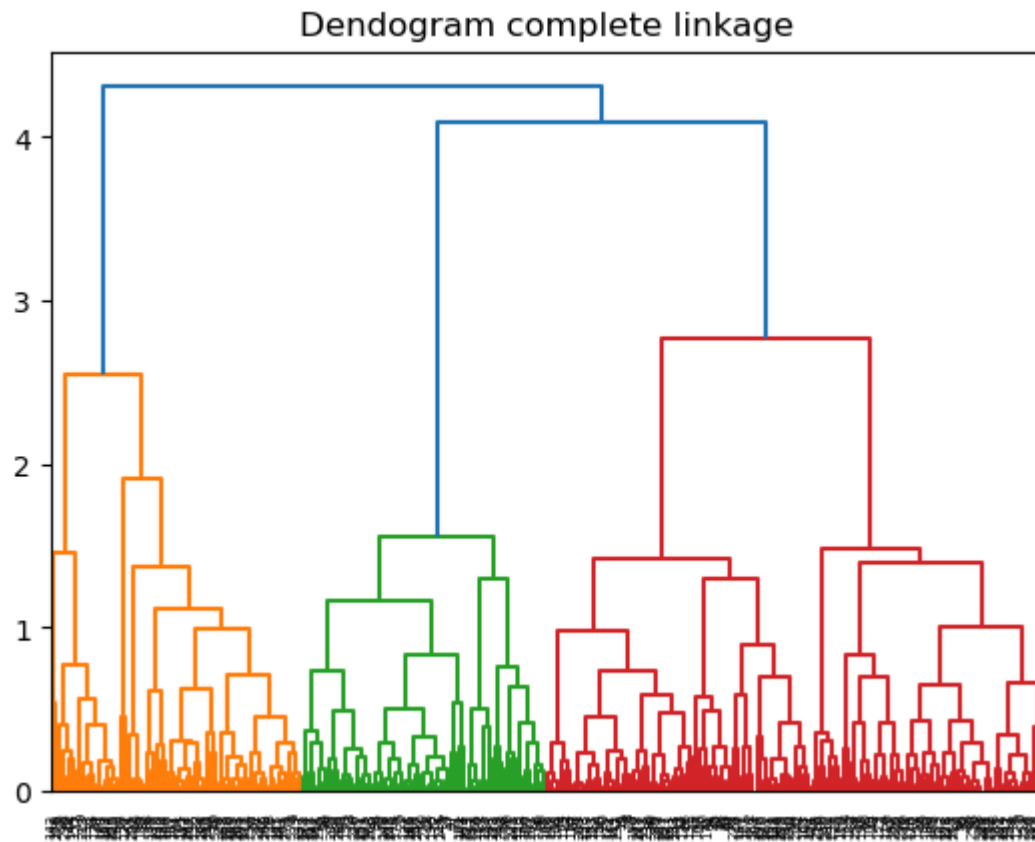


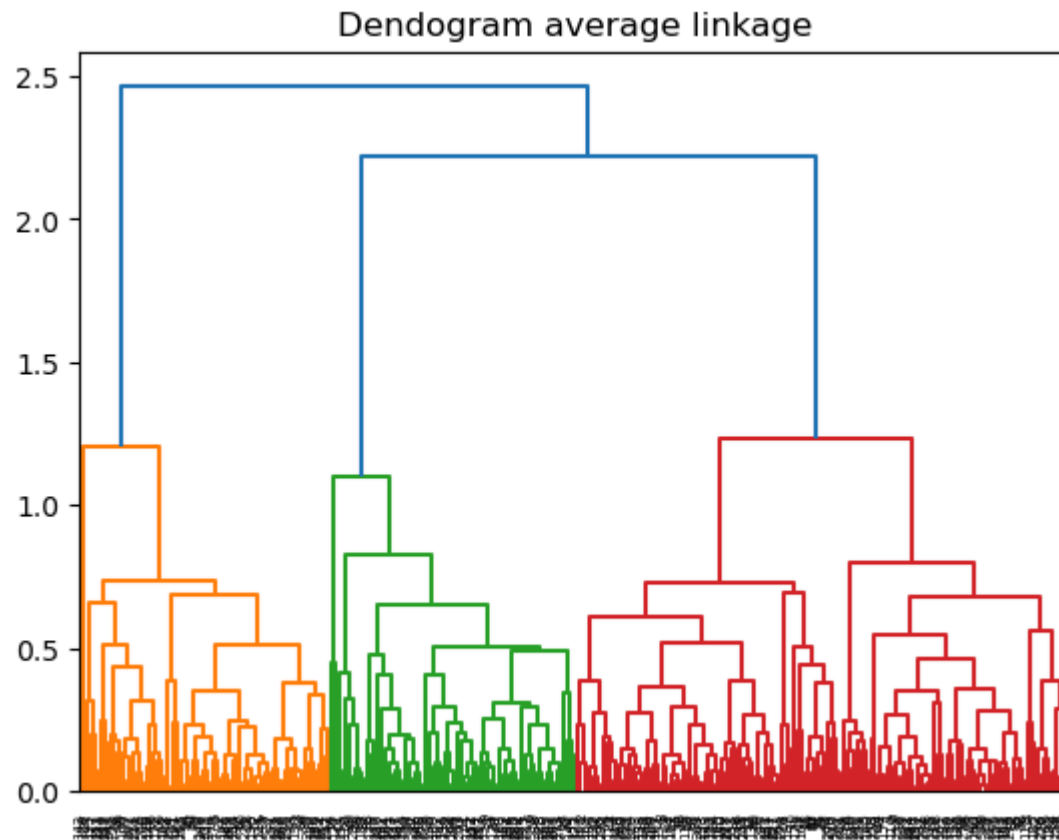
Dendrograms

```
In [40]: # fig, axes = plt.subplots(1,4,figsize = (20,5))
for i, method in enumerate(linkage_methods):
    linkage_matrix = linkage(x,method = method)
    plt.title(f"Dendrogram {method} linkage")
    dendrogram(linkage_matrix)
    plt.show()
```









DBSCAN

```
In [43]: fig, axes = plt.subplots(1, 3, figsize=(15, 5))
eps_values = [0.3, 0.5, 0.7]
for i, eps in enumerate(eps_values):
    dbscan = DBSCAN(eps=eps, min_samples=5)
    labels = dbscan.fit_predict(x)
    axes[i].scatter(x[:, 0], x[:, 1], c=labels, cmap='viridis', edgecolors='k')
    axes[i].set_title(f'DBSCAN (eps={eps})')
plt.show()
```

