yk3xtsrrq

February 12, 2025

[2]:
```
from sklearn.cluster import AgglomerativeClustering
import pandas as pd
import numpy as np
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
```

[3]:
```
iris = load_iris()
data = pd.DataFrame(data=iris.data, columns=iris.feature_names)
data['target'] = iris.target
```

[3]:
```
data.head()
```

[3]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

[4]:
```
data.tail()
```

[4]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

```
146       2
147       2
148       2
149       2
```

[5]:
```
data.describe()
```

[5]:
```
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.057333           3.758000
std             0.828066          0.435866           1.765298
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)      target
count        150.000000  150.000000
mean           1.199333    1.000000
std            0.762238    0.819232
min            0.100000    0.000000
25%            0.300000    0.000000
50%            1.300000    1.000000
75%            1.800000    2.000000
max            2.500000    2.000000
```

[6]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
 4   target             150 non-null    int64
dtypes: float64(4), int64(1)
memory usage: 6.0 KB
```

[7]:
```
data.shape
```

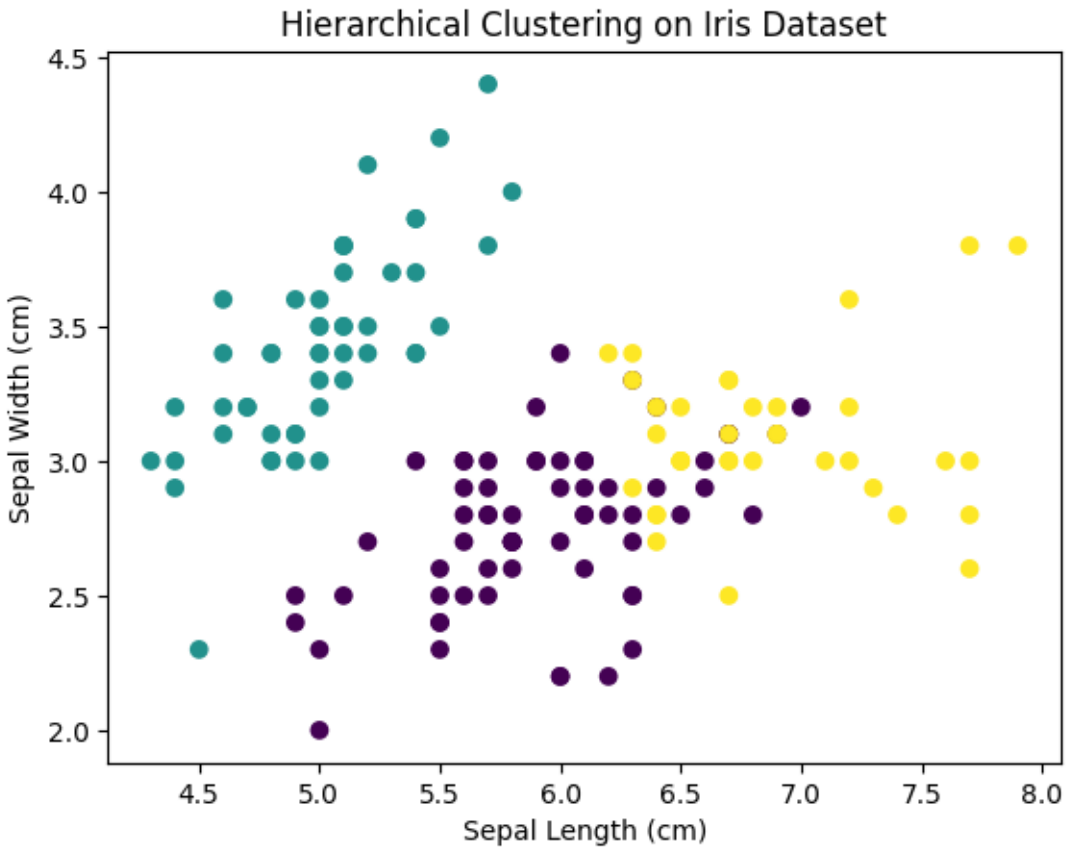
[7]:
```
(150, 5)
```

[8]:
```
data.isnull().sum()
```

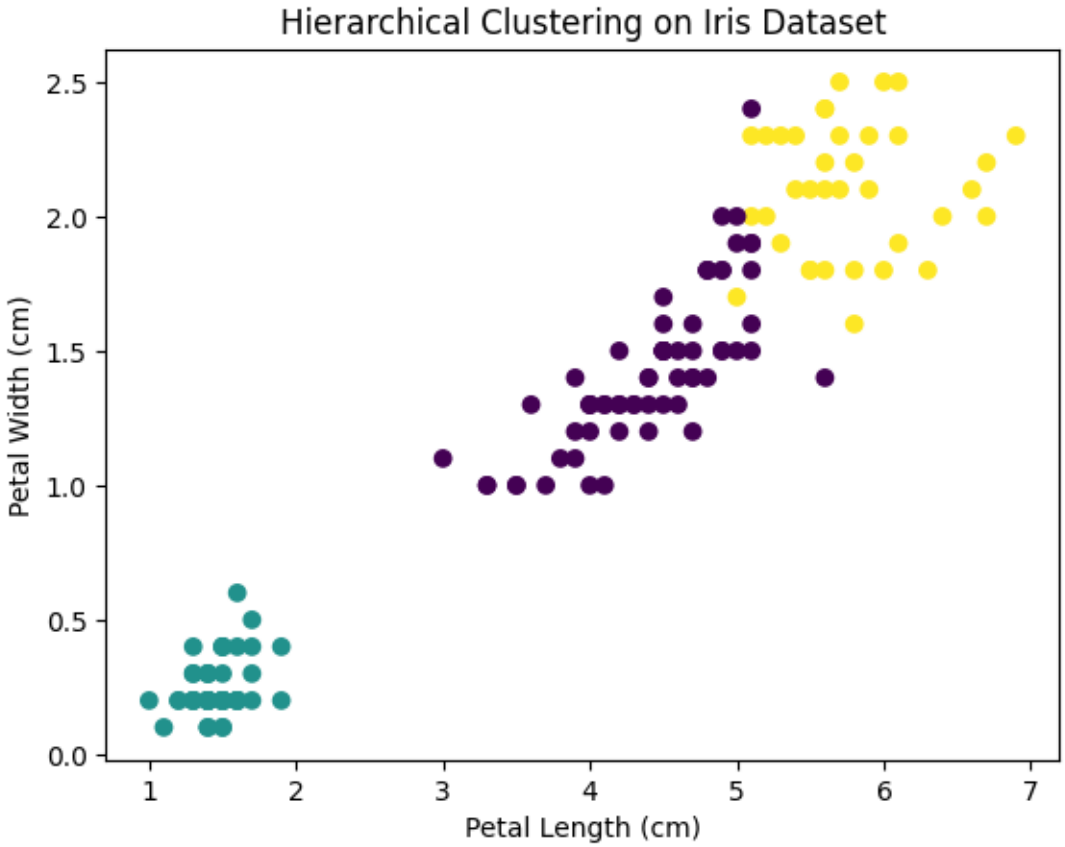
[8]:
```
sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
target               0
dtype: int64
```

[9]:
```
#Applying Hierarchical Clustering:
agglo_cluster = AgglomerativeClustering(n_clusters=3)
agglo_cluster.fit(data[['sepal length (cm)', 'sepal width (cm)', 'petal length␣
 ↪(cm)', 'petal width (cm)']])
data['cluster'] = agglo_cluster.labels_
```

[10]:
```
#Visualizing clusters:
plt.scatter(data['sepal length (cm)'], data['sepal width (cm)'],␣
 ↪c=data['cluster'])
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.title('Hierarchical Clustering on Iris Dataset')
plt.show()
```



[11]:
```
plt.scatter(data['petal length (cm)'], data['petal width (cm)'],␣
 ↪c=data['cluster'])
plt.xlabel('Petal Length (cm)')
plt.ylabel('Petal Width (cm)')
plt.title('Hierarchical Clustering on Iris Dataset')
plt.show()
```



[13]:
```
from sklearn.metrics import silhouette_score, calinski_harabasz_score,␣
 ↪davies_bouldin_score
#Calculate metrics
silhouette = silhouette_score(data[['sepal length (cm)', 'sepal width (cm)',␣
 ↪'petal length (cm)', 'petal width (cm)']], data['cluster'])
```

[14]:
```
calinski = calinski_harabasz_score(data[['sepal length (cm)', 'sepal width␣
 ↪(cm)', 'petal length (cm)', 'petal width (cm)']], data['cluster'])
davies_bouldin = davies_bouldin_score(data[['sepal length (cm)', 'sepal width␣
 ↪(cm)', 'petal length (cm)', 'petal width (cm)']], data['cluster'])
```

[16]:
```python
#Print results
print("Hierarchical Clustering Metrics on Iris Dataset:")
print(f"Silhouette Score: {silhouette}")
print(f"Calinski-Harabasz Index: {calinski}")
print(f"Davies-Bouldin Index: {davies_bouldin}")
```

```
Hierarchical Clustering Metrics on Iris Dataset:
Silhouette Score: 0.5543236611296419
Calinski-Harabasz Index: 558.0580408128307
Davies-Bouldin Index: 0.6562564540642021
```