Hierarchical Clustering

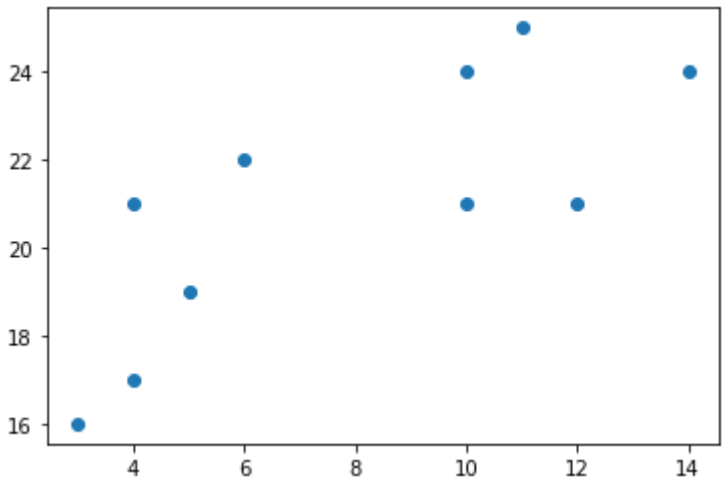
Hierarchical clustering is an **unsupervised learning method** for clustering data points.

The algorithm builds clusters **by measuring the dissimilarities between data**.

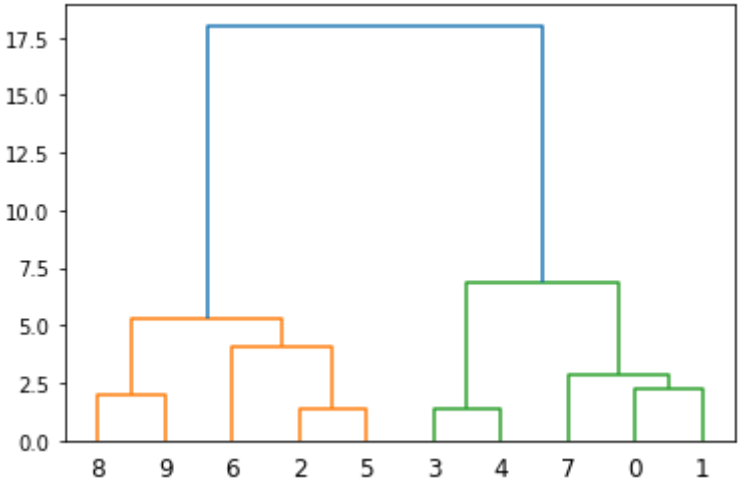
Unsupervised learning means that a model does not have to be trained, and we do not need a "target" variable. This method can be used on any data to visualize and interpret the relationship between individual data points.

Here we will use hierarchical clustering to group data points and visualize the clusters using both a dendrogram and scatter plot.

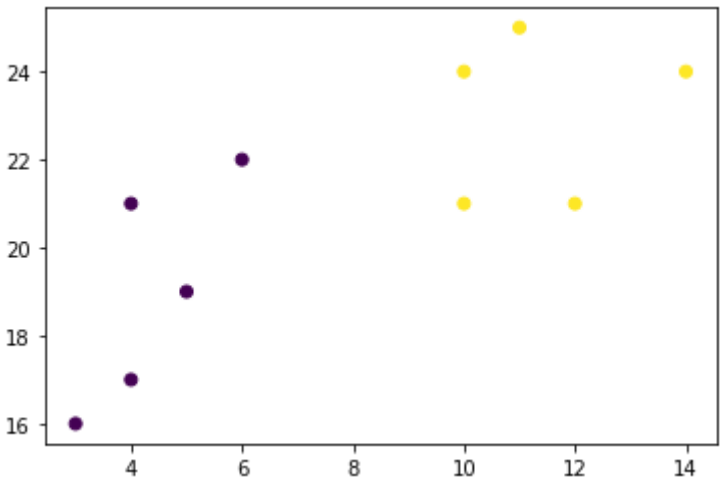
In [1]:
```
## Example 1

import numpy as np
import matplotlib.pyplot as plt

x = [4, 5, 10, 4, 3, 11, 14 , 6, 10, 12]
y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

plt.scatter(x, y)
plt.show()
```



In [2]:
```
## Example 2

import numpy as np
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage

x = [4, 5, 10, 4, 3, 11, 14 , 6, 10, 12]
y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

data = list(zip(x, y))

linkage_data = linkage(data, method='ward', metric='euclidean')
dendrogram(linkage_data)

plt.show()
```



In [3]:
```
## Example 3

import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import AgglomerativeClustering

x = [4, 5, 10, 4, 3, 11, 14 , 6, 10, 12]
y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]

data = list(zip(x, y))

hierarchical_cluster = AgglomerativeClustering(n_clusters=2, affinity='eucl
labels = hierarchical_cluster.fit_predict(data)

plt.scatter(x, y, c=labels)
plt.show()
```



In [4]:
```
## Example 4

import numpy as np
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage
from sklearn.cluster import AgglomerativeClustering
```

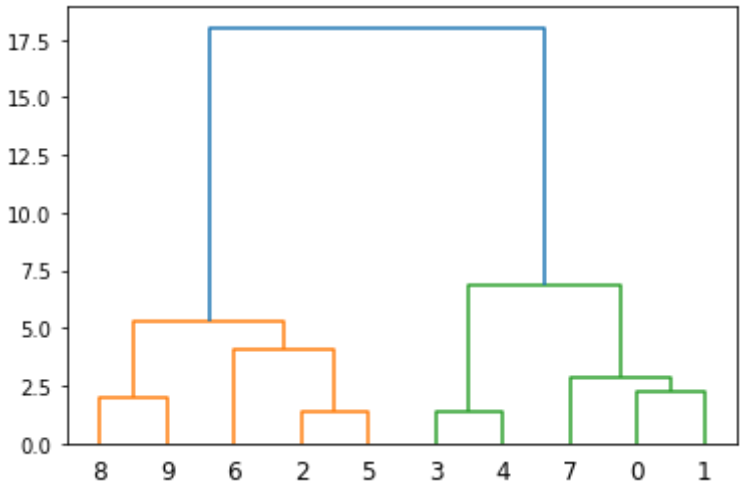
In [5]:
```
x = [4, 5, 10, 4, 3, 11, 14 , 6, 10, 12]
y = [21, 19, 24, 17, 16, 25, 24, 22, 21, 21]
```

In [6]:
```
data = list(zip(x, y))
print(data)
```

```
[(4, 21), (5, 19), (10, 24), (4, 17), (3, 16), (11, 25), (14, 24), (6, 22),
(10, 21), (12, 21)]
```

In [7]:
```
linkage_data = linkage(data, method='ward', metric='euclidean')
```

In [8]:
```
dendrogram(linkage_data)
plt.show()
```

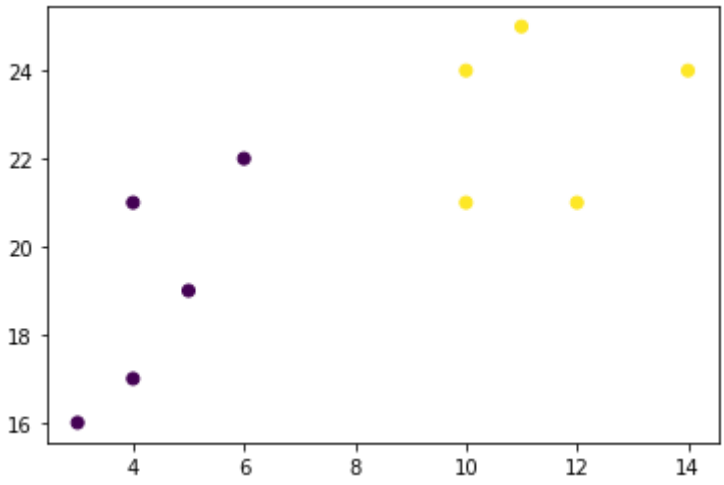


In [9]:
```
hierarchical_cluster = AgglomerativeClustering(n_clusters=2, affinity='eucl
```

In [11]:
```
labels = hierarchical_cluster.fit_predict(data)
print(labels)
```

```
[0 0 1 0 0 1 1 0 1 1]
```

In [12]:
```
plt.scatter(x, y, c=labels)
plt.show()
```



In [ ]:
```

```