Forward School

Program Code: J620-002-4:2020

Program Name: FRONT-END SOFTWARE DEVELOPMENT

Title : Exe27 - Hierarchical Cluster of Grains

Name: Ooi Caaron

IC Number: 990701-07-5837

Date :2/8/23

Introduction : Learning cluster of grains

Conclusion :

Hierarchical clustering of the grain data

The exercise is adapted from https://github.com/benjaminwilson/python-hierarchical-clustering-exercises (https://github.com/benjaminwilson/python-hierarchical-clustering-exercises)

In this exercise, we will try to use the SciPy `linkage()` function performs hierarchical clustering on an array of samples. Use the `linkage()` function to obtain a hierarchical clustering of the grain samples, and use `dendrogram()` to visualize the result. A sample of the grain measurements is provided in the array `samples`, while the variety of each grain sample is given by the list `varieties`.

Step 1: Load the dataset *(done for you)*.

In [1]:

```python
import pandas as pd

seeds_df = pd.read_csv('seeds-less-rows.csv')

# remove the grain species from the DataFrame, save for later
varieties = list(seeds_df.pop('grain_variety'))


samples = seeds_df.values
```

Step 2: Create a variable called varieties, and remove grain species column from seeds_df and keep it in varieties. After that, create another variable called samples that stores the measurements (values) as a NumPy array

In [1]:

```python
# remove the grain species from the DataFrame, save for later
varities = varieties = seeds_df.iloc[:, -1].values

# extract the measurements as a NumPy array
samples = seeds_df.values

print(varieties)
print(samples)
```

Step 3: Import:

- `linkage` and `dendrogram` from `scipy.cluster.hierarchy` .
- `matplotlib.pyplot as plt` .

In [ ]:

```python
from scipy.cluster.hierarchy import linkage, dendrogram
import matplotlib.pyplot as plt
```

Step 4: Perform hierarchical clustering on `samples` using the `linkage()` function with the `method='complete'` keyword argument. Assign the result to `mergings` .

In [ ]:

```python
mergings = linkage(samples, method='complete')
```

Step 5: Plot a dendrogram using the `dendrogram()` function on `mergings` , specifying the keyword arguments `labels=varieties` , `leaf_rotation=90` , and `leaf_font_size=6` . Remember to call `plt.show()` afterwards, to display your plot.

In [ ]:

```python
dendrogram(mergings,
          labels=varieties,
          leaf_rotation=90,
          leaf_font_size=6,
)
plt.show()
```