

10.5.2 Hierarchical Clustering

May 18, 2018

1 10.5.2 Hierarchical Clustering

Here we are going to use scipy. First we will do the same as the last exercise. It will all be made in one cell. For details please see 10.5.1

```
In [3]: # conventional way to import pandas
import pandas as pd
# conventional way to import seaborn
import seaborn as sns
```

```
from sklearn.cluster import KMeans
import numpy as np
import matplotlib.pyplot as plt
```

```
np.random.seed(0)
X = np.random.standard_normal(size=(25,2))
```

```
for row in X:
    row[0] = row[0]+3
    row[1] = row[1]-4
```

```
kmeans2 = KMeans(n_clusters=2, random_state=0,n_init=20).fit(X)
kmeans3 = KMeans(n_clusters=3, random_state=0,n_init=20).fit(X)
```

Here comes the interesting part...

```
In [4]: from scipy.cluster import hierarchy
fig, (ax1,ax2,ax3) = plt.subplots(3,1, figsize=(15,18))

for linkage, cluster, ax in zip([hierarchy.complete(X), hierarchy.average(X), hierarchy.single(X)],
                                [ax1,ax2,ax3]):
    cluster = hierarchy.dendrogram(linkage, ax=ax, color_threshold=0)

ax1.set_title('Complete Linkage')
ax2.set_title('Average Linkage')
ax3.set_title('Single Linkage');
```

