song

March 24, 2025

1 import libraries

[4]:
```
import pandas as pd
import numpy as np
```

2 import dataset and construct adjacency matrix A

[5]:
```
def cal_A(songs):
    A = np.zeros((songs.shape[0], songs.shape[0]))
    for i in range(songs.shape[0]):
        for j in range(songs.shape[0]):
            distance = np.linalg.norm(songs[i] - songs[j])
            if distance > 1:
                A[i][j] = 1

    return A

file_path = 'songs.csv'
df = pd.read_csv(file_path)
song = df.to_numpy() # convert the dataframe to numpy array
A = cal_A(song)
```

3 perform spectral clustering on normalized L

[6]:
```
D = np.diag(np.sum(A, axis=1))
L = D - A

D_inv_sqrt = np.linalg.inv(np.sqrt(D))
L_norm = D_inv_sqrt @ L @ D_inv_sqrt
eigenvalues, eigenvectors = np.linalg.eig(L_norm)

sorted_indices = np.argsort(eigenvalues)
eigenvectors = eigenvectors[:, sorted_indices]
eigenvalues = eigenvalues[sorted_indices]

print(eigenvalues[:5]) # here we print out the first 5 eigenvalues

v = eigenvectors[:, 1]
x = D_inv_sqrt @ v
clusters = np.where(x >= 0, 1, 2) # do the clustering based on the second␣
 ↪smallest eigenvector
print("The first 10 clusters are:")
print(clusters[:10])
```

```
[-3.79980777e-16  9.89839228e-01  9.89945494e-01  9.90020082e-01
  9.90087966e-01]
The first 10 clusters are:
[2 1 2 1 2 1 1 1 2 1]
```

3.1 interpret the output

- The first 5 eigenvalues are [**-3.79980777e-16, 9.89839228e-01, 9.89945494e-01, 9.90020082e-01, 9.90087966e-01**]
- There is one eigenvalue that is very close to 0, which means the graph has 1 connected component

4 explore the systematic differences between the two clusters

[7]:
```
cluster_2_mean = song[clusters == 2].mean(axis=0)
cluster_1_mean = song[clusters == 1].mean(axis=0)

mean_differences = cluster_2_mean - cluster_1_mean
top_3_features = np.argsort(np.abs(mean_differences))[-3:][::-1]
print(top_3_features)
```

```
[1 6 9]
```

4.1 interpret the output

- This means that there is a significant difference between the two clusters in terms of feature 1, 6 and 9, which can be used to distinguish between them.