

clustering_by_silhouette - implementation

created by Etzion Harari - etzionhar@gmail.com

Introduction

import & define essential modules + variables

```
In [1]: from clustering_by_silhouette import silhouette_clustering

import pandas as pd
from create_3d_gif import pd_to_gif
from sklearn.decomposition import PCA

MPL_Colors=['blue', 'green','red', 'gold', 'purple', 'lime', 'tomato', 'navy', 'teal', 'maroon',
            'olive', 'orange', 'sienna', 'indigo','yellow', 'darkgreen', 'darkblue', 'chocolate',
            'black']

def pca(df,dim):
    return PCA(n_components=dim).fit(df.T).components_
```

import data

```
In [2]: df1 = pd.read_excel('data_1.xlsx')
df2 = pd.read_excel('data_2.xlsx')
fields1 = ['a','b','c','d','e','f','g','h','i','j','k','l','m','n']
fields2 = ['f' + str(i) for i in range(1, 15)] # fields names
```

Analysis

run silhouette_clustering for hdbscan & kmeans

```
In [3]: m_label = silhouette_clustering(df1[fields1], typ='meanshift', org=9, lim=13) # MeanShift
h_label = silhouette_clustering(df2[fields2], typ='hdbscan') # HDBSCAN
k_label = silhouette_clustering(df2[fields2], org=3 ,lim=10) # Kmeans (default)

cluster kind: meanshift, input value = 9, silhouette = 31.2%
cluster kind: meanshift, input value = 10, silhouette = 37.7%
cluster kind: meanshift, input value = 11, silhouette = 28.0%
cluster kind: meanshift, input value = 12, silhouette = 33.9%
cluster kind: meanshift, input value = 13, silhouette = 46.2%
cluster kind: hdbscan, input value = 2, silhouette = 7.9%
cluster kind: hdbscan, input value = 3, silhouette = -34.3%
cluster kind: hdbscan, input value = 4, silhouette = 18.3%
cluster kind: hdbscan, input value = 5, silhouette = -14.0%
cluster kind: hdbscan, input value = 6, silhouette = 4.5%
cluster kind: hdbscan, input value = 7, silhouette = 5.7%
cluster kind: hdbscan, input value = 8, silhouette = 5.4%
cluster kind: hdbscan, input value = 9, silhouette = 8.2%
cluster kind: hdbscan, input value = 10, silhouette = 36.1%
cluster kind: hdbscan, input value = 11, silhouette = 35.5%
cluster kind: hdbscan, input value = 12, silhouette = 35.4%
cluster kind: hdbscan, input value = 13, silhouette = 35.7%
cluster kind: hdbscan, input value = 14, silhouette = 34.6%
cluster kind: hdbscan, input value = 15, silhouette = 34.7%
cluster kind: hdbscan, input value = 16, silhouette = 35.6%
cluster kind: hdbscan, input value = 17, silhouette = 34.9%
cluster kind: hdbscan, input value = 18, silhouette = 35.4%
cluster kind: hdbscan, input value = 19, silhouette = 53.2%
cluster kind: hdbscan, input value = 20, silhouette = 52.6%
cluster kind: kmeans, input value = 3, silhouette = 37.3%
cluster kind: kmeans, input value = 4, silhouette = 44.7%
cluster kind: kmeans, input value = 5, silhouette = 46.1%
cluster kind: kmeans, input value = 6, silhouette = 46.0%
cluster kind: kmeans, input value = 7, silhouette = 36.2%
cluster kind: kmeans, input value = 8, silhouette = 36.2%
cluster kind: kmeans, input value = 9, silhouette = 36.0%
cluster kind: kmeans, input value = 10, silhouette = 35.8%
```

```
In [4]: print(f'meanshift number of clusters: {len(set(m_label))}\nkmeans number of clusters: {len(set(k_label))}\nhdbscan number of clusters: {len(set(h_label))-1}')

meanshift number of clusters: 5
kmeans number of clusters: 5
hdbscan number of clusters: 4
```

Plot Results

Arrange Data

```
In [5]: df1['x'], df1['y'], df1['z'] = pca(df1[fields1],3)
df2['x'], df2['y'], df2['z'] = pca(df2[fields2],3)

df1['m_color'] = [MPL_Colors[i] for i in m_label]
df2['h_color'] = [MPL_Colors[i] for i in h_label]
df2['k_color'] = [MPL_Colors[i] for i in k_label]
```

Plot Data

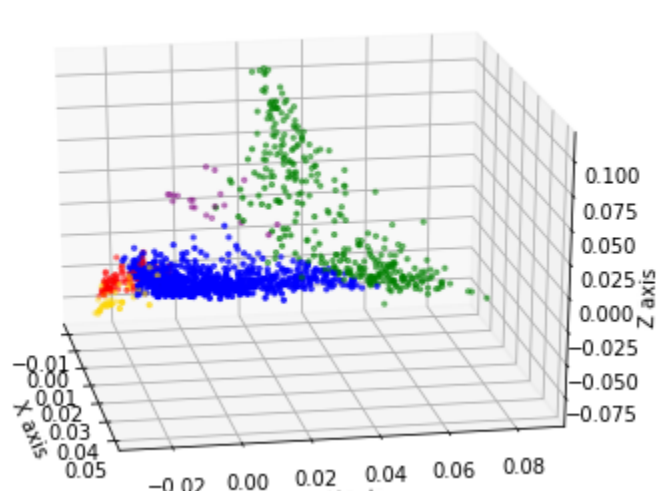
```
In [6]: pd_to_gif(df1,['x','y','z'],'meanshift_plot',clrs='m_color')
pd_to_gif(df2,['x','y','z'],'hdbscan_plot',clrs='h_color')
pd_to_gif(df2,['x','y','z'],'kmeans_plot',clrs='k_color')
```

ggggggg
img number 1 from 36 created
img number 2 from 36 created
img number 3 from 36 created
img number 4 from 36 created
img number 5 from 36 created
img number 6 from 36 created
img number 7 from 36 created
img number 8 from 36 created
img number 9 from 36 created
img number 10 from 36 created
img number 11 from 36 created
img number 12 from 36 created
img number 13 from 36 created
img number 14 from 36 created
img number 15 from 36 created
img number 16 from 36 created
img number 17 from 36 created
img number 18 from 36 created
img number 19 from 36 created
img number 20 from 36 created
img number 21 from 36 created
img number 22 from 36 created
img number 23 from 36 created
img number 24 from 36 created
img number 25 from 36 created
img number 26 from 36 created
img number 27 from 36 created
img number 28 from 36 created
img number 29 from 36 created
img number 30 from 36 created
img number 31 from 36 created
img number 32 from 36 created
img number 33 from 36 created
img number 34 from 36 created
img number 35 from 36 created
img number 36 from 36 created
the gif ' meanshift_plot ' created

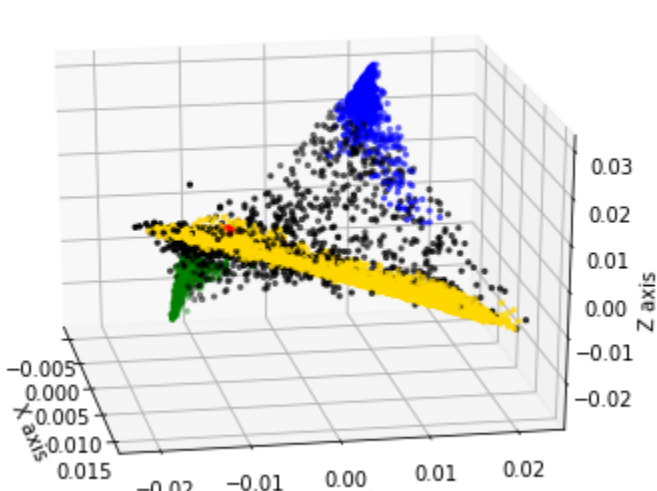
ggggggg
img number 1 from 36 created
img number 2 from 36 created
img number 3 from 36 created
img number 4 from 36 created
img number 5 from 36 created
img number 6 from 36 created
img number 7 from 36 created
img number 8 from 36 created
img number 9 from 36 created
img number 10 from 36 created
img number 11 from 36 created
img number 12 from 36 created
img number 13 from 36 created
img number 14 from 36 created
img number 15 from 36 created
img number 16 from 36 created
img number 17 from 36 created
img number 18 from 36 created
img number 19 from 36 created
img number 20 from 36 created
img number 21 from 36 created
img number 22 from 36 created
img number 23 from 36 created
img number 24 from 36 created
img number 25 from 36 created
img number 26 from 36 created
img number 27 from 36 created
img number 28 from 36 created
img number 29 from 36 created
img number 30 from 36 created
img number 31 from 36 created
img number 32 from 36 created
img number 33 from 36 created
img number 34 from 36 created
img number 35 from 36 created
img number 36 from 36 created
the gif ' hdbscan_plot ' created

ggggggg
img number 1 from 36 created
img number 2 from 36 created
img number 3 from 36 created
img number 4 from 36 created
img number 5 from 36 created
img number 6 from 36 created
img number 7 from 36 created
img number 8 from 36 created
img number 9 from 36 created
img number 10 from 36 created
img number 11 from 36 created
img number 12 from 36 created
img number 13 from 36 created
img number 14 from 36 created
img number 15 from 36 created
img number 16 from 36 created
img number 17 from 36 created
img number 18 from 36 created
img number 19 from 36 created
img number 20 from 36 created
img number 21 from 36 created
img number 22 from 36 created
img number 23 from 36 created
img number 24 from 36 created
img number 25 from 36 created
img number 26 from 36 created
img number 27 from 36 created
img number 28 from 36 created
img number 29 from 36 created
img number 30 from 36 created
img number 31 from 36 created
img number 32 from 36 created
img number 33 from 36 created
img number 34 from 36 created
img number 35 from 36 created
img number 36 from 36 created
the gif ' kmeans_plot ' created

meanshift_plot



hdbscan_plot



kmeans_plot

