clustering-small-exercise

October 24, 2024

1 Small exercise on clustering

[15]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import warnings

warnings.filterwarnings('ignore')

#from sklearn.cluster import KMeans
#from sklearn.metrics import silhouette_samples , silhouette_score
from sklearn.datasets import make_blobs
colormap = ['r', 'g', 'b', 'c', 'm', 'y', 'k']
```

[38]:
```
# Generating the sample data from make_blobs
X, y = make_blobs(n_samples=10,cluster_std=1,centers=3,
                  shuffle=True,center_box=(-2,2),␣
 ↪n_features=2,random_state=54321)
```

[41]:
```
y = y.astype(int) #making y integers.
```

[ ]:
```
dataset = pd.DataFrame(np.column_stack((X,y)),columns=['x1','x2','y'])
```

[44]:
```
dataset[['y']] = dataset[['y']].astype(int)
```

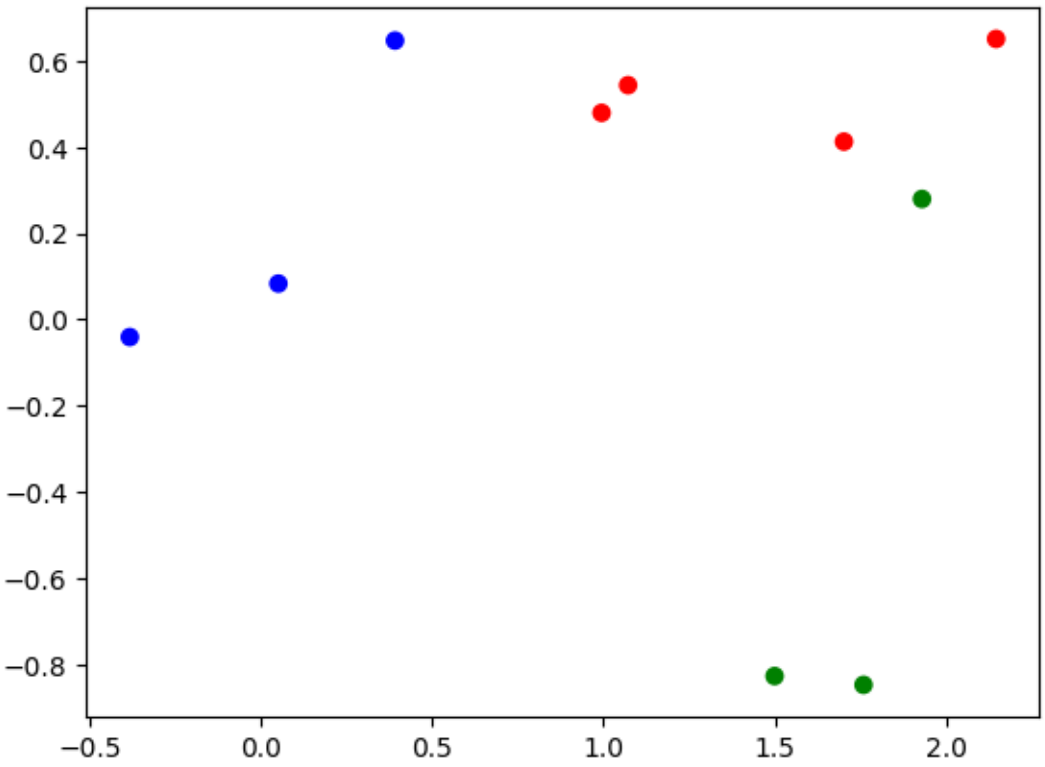
[45]:
```
dataset.head(n=-1)
```

[45]:

	x1	x2	y
0	1.759215	-0.846642	1
1	1.499699	-0.826284	1
2	1.702011	0.412539	0
3	0.051737	0.083544	2
4	2.146060	0.650696	0
5	0.994875	0.479277	0
6	-0.381877	-0.040455	2
7	1.929758	0.279628	1

```
8  1.072098  0.543442  0
```

[46]:
```
plt.scatter(X[:,0],X[:,1],c=[colormap[k] for k in y])
plt.show()
```



[46]:
```
(10, 2)
```

[49]:
```
c0_members = dataset.query('y==0')
```

[50]:
```
c1_members = dataset.query('y==1')
```

[51]:
```
c2_members = dataset.query('y==2')
```

[53]:
```
c0_members
```

[53]:
```
         x1        x2  y
2  1.702011  0.412539  0
4  2.146060  0.650696  0
5  0.994875  0.479277  0
8  1.072098  0.543442  0
```

[54]:
```
np.sum(c0_members[['x1']])
```

[54]:
```
x1    5.915045
dtype: float64
```

[55]:
```
np.sum(c0_members[['x2']])
```

[55]:
```
x2    2.085955
dtype: float64
```

[56]:
```
len(c0_members)
```

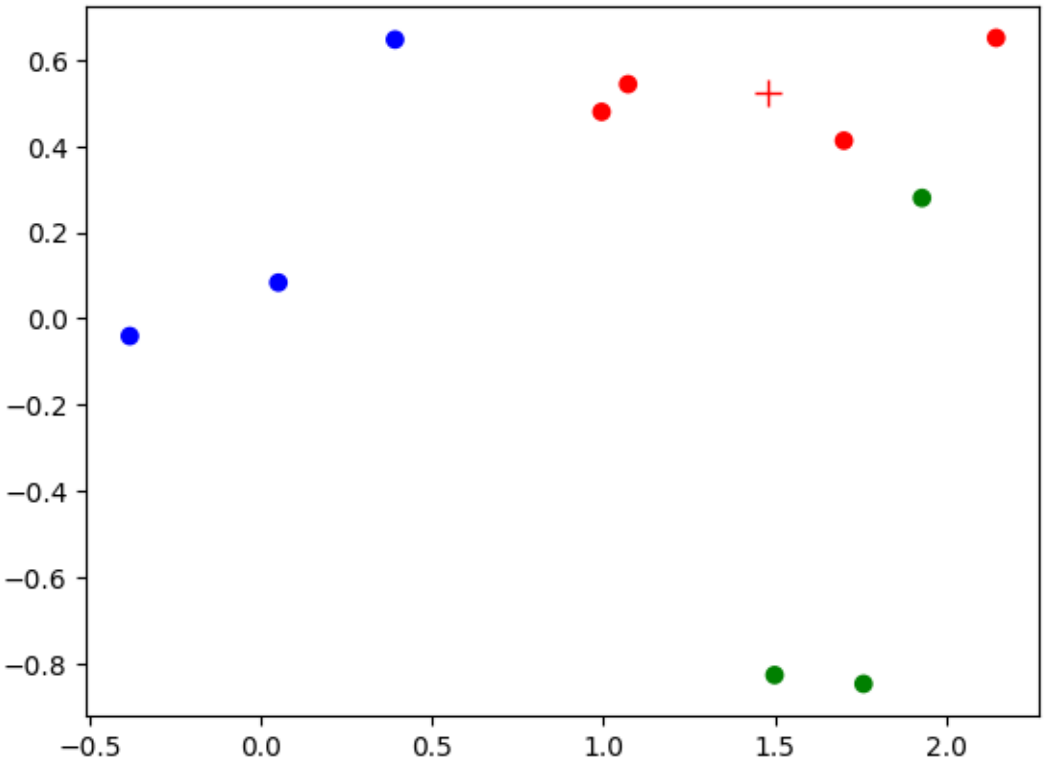
[56]:
```
4
```

[57]:
```
c0_centroid = [ np.sum(c0_members[['x1']])/len(c0_members),
                np.sum(c0_members[['x2']])/len(c0_members),
              ]
```

[58]:
```
c0_centroid
```

[58]:
```
[x1    1.478761
 dtype: float64,
 x2    0.521489
 dtype: float64]
```

[60]:
```
plt.scatter(X[:,0],X[:,1],c=[colormap[k] for k in y])
plt.plot(c0_centroid[0],c0_centroid[1],marker='+', markersize=10, color='r')
plt.show()
```


0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−0.5
0.0
0.5
1.0
1.5
2.0

[ ]: