Working with Unlabeled Data – Clustering Analysis

Overview

```
from IPython.display import Image
%matplotlib inline
```

Grouping objects by similarity using k-means

1. Randomly pick k centroids from the sample points as initial cluster centers.

$$(j)$$

2. Assign each sample to the nearest centroid μ , $j \in \{ 1, \dots , k \}$.
3. Move the centroids to the center of the samples that were assigned to it.
4. Repeat the steps 2 and 3 until the cluster assignment do not change or a user-defined tolerance or a maximum number of iterations is reached.

```
from sklearn.cluster import KMeans
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style
style.use("ggplot")
%matplotlib inline
```

```
data = pd.DataFrame([[1, 2],
              [5, 8],
              [1.5, 1.8],
              [8, 8],
              [1, 0.6],
              [9, 11]], columns=['x','y'])
print( data )
```

```
          x     y
     0  1.0   2.0
     1  5.0   8.0
     2  1.5   1.8
     3  8.0   8.0
```

```
4  1.0   0.6
5  9.0  11.0
```

```
kmeans = KMeans(n_clusters=2).fit(data)

centroids = kmeans.cluster_centers_
labels = kmeans.labels_

print(centroids)
print(labels)
```

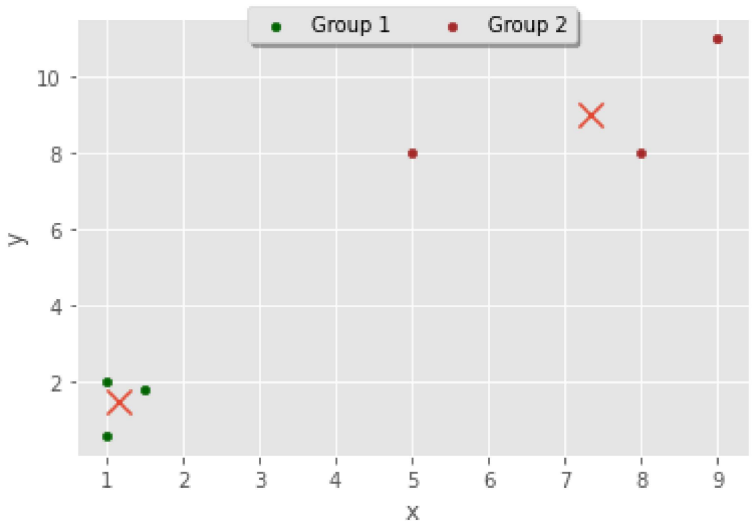
```
[[7.33333333 9.        ]
 [1.16666667 1.46666667]]
[1 0 1 0 1 0]
```

```
data['labels'] = labels

#plt.plot(data, colors[data['labels'], markersize = 10)

group1 = data[data['labels']==1].plot( kind='scatter', x='x', y='y', color='DarkGreen', label
group2 = data[data['labels']==0].plot( kind='scatter', x='x', y='y', color='Brown', ax=group1
group1.legend(loc='upper center', bbox_to_anchor=(0.5, 1.05),
          ncol=3, fancybox=True, shadow=True)
plt.scatter(centroids[:, 0],centroids[:, 1], marker = "x", s=150, linewidths = 5, zorder = 10

plt.show()
```



Colab paid products - Cancel contracts here

0s completed at 3:28 PM