CE066 | Jaydeep Mahajan | ML | LAB 8

```
1 # import all libraries
2 from sklearn.datasets import load_breast_cancer
3 from sklearn.cluster import KMeans
4 import numpy as np
5 from scipy.stats import mode
6 from sklearn.metrics import accuracy_score
7 import matplotlib.pyplot as plt
8 from sklearn.metrics import confusion_matrix
9 import seaborn as sns
```

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use
  import pandas.util.testing as tm
```

```
1 # load dataset
2 data = load_breast_cancer()
3 list(data.target_names)
```

```
['malignant', 'benign']
```

```
1 # check data size and target size
2 print("Data size : ",data.data.shape)
3 print("Target size : ",data.target.shape)
```

```
Data size :  (569, 30)
Target size :  (569,)
```

```
1 # define kMeans and fit data into model
2 kmeans = KMeans(n_clusters=2,random_state=74)
3 prediction = kmeans.fit_predict(data.data)
```

```
# Cluster shape 2 - class 30 -features
kmeans.cluster_centers_.shape
```

```
(2, 30)
```

```
# find accuracy score
labels = np.zeros_like(prediction)
for i in range(2):
  mask = (prediction==i)
  labels[mask] = mode(data.target[mask])[0]
accuracy = accuracy_score(data.target,labels)
print("Accuracy is :  ",accuracy)
```

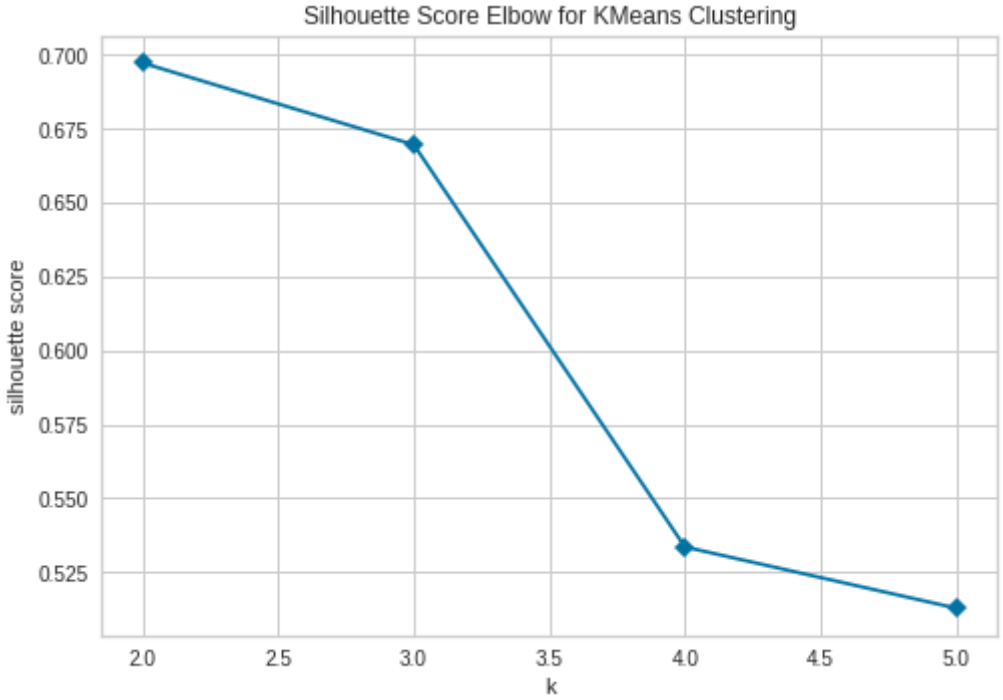
```
Accuracy is :   0.8541300527240774
```

```
# create confusion matrix using heatmap
mat = confusion_matrix(data.target, labels)
ax = sns.heatmap(mat.T, square=True, annot=True, fmt='d',cbar=False,
                 xticklabels=data.target_names,
                 yticklabels=data.target_names)
ax.set_ylim(2,0,0)
plt.xlabel('true label')
plt.ylabel('predicted label')
```

```
Text(91.68, 0.5, 'predicted label')
```

```
1 # Find the best number of cluster for clustering using KElbowVisualizer
2 from yellowbrick.cluster import KElbowVisualizer
3 model = KMeans(random_state=0)
4 visualizer = KElbowVisualizer(model,k=(2,6),metric='silhouette',timings=False)
5 visualizer.fit(data.data)
6 visualizer.poof()
```

```
/usr/local/lib/python3.6/dist-packages/sklearn/utils/deprecation.py:144: FutureWarning: The sklearn.metrics.classification modu
  warnings.warn(message, FutureWarning)
```




```
1
```

```
1 # Check siihouette_score which is same as above graph for no.cluster=2
```

```
from sklearn.metrics import silhouette_score
print(silhouette_score(data.data,labels))
```

```
0.6972646156059464
```

```
1
```