گزارش تمرین دوم داده کاوی

۳-۱:

In [17]:
```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np
import pandas as pd
%matplotlib inline
from sklearn import datasets
#Iris Dataset
iris = datasets.load_iris()
X = iris.data
```

۳-۲: ----- ۳-۳: باهم

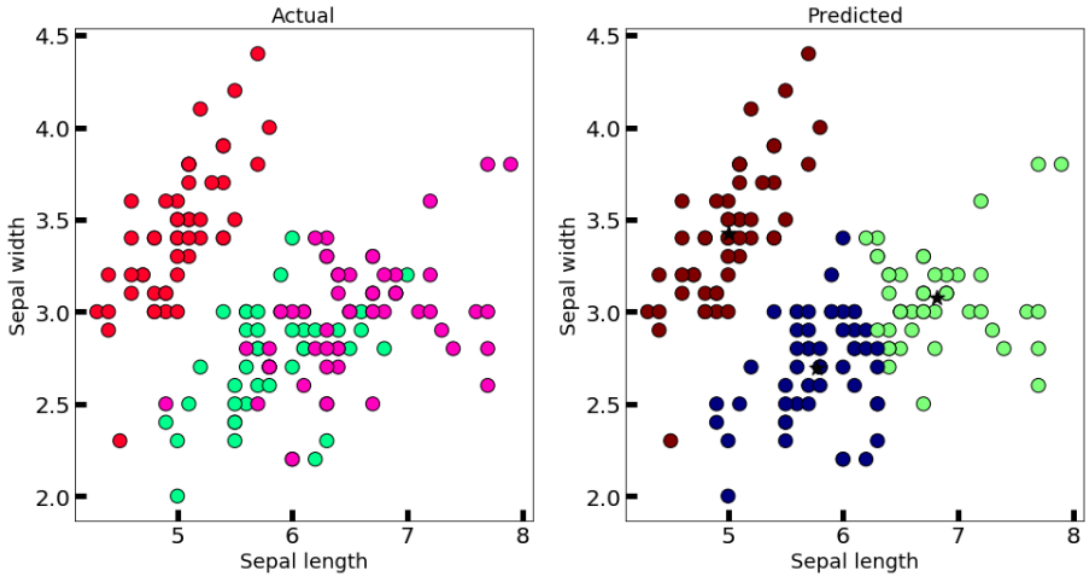
In [45]:
```
#KMeans
km = KMeans(n_clusters=3)
km.fit(X)
km.predict(X)
labels = km.labels_
centroids = km.cluster_centers_
print(centroids)
```

```
[[5.77358491 2.69245283]
 [6.81276596 3.07446809]
 [5.006      3.428     ]]
```

۳-۴:

In [51]:
```
from matplotlib import pyplot as plt
new_labels = km.labels_
X = iris.data[:,:3]
y = iris.target
# Plot the identified clusters and compare with the answers
fig, axes = plt.subplots(1, 2, figsize=(16,8))
axes[0].scatter(X[:, 0], X[:, 1], c=y, cmap='gist_rainbow',
edgecolor='k', s=150)
axes[1].scatter(X[:, 0], X[:, 1], c=new_labels, cmap='jet',
edgecolor='k', s=150)
axes[0].set_xlabel('Sepal length', fontsize=18)
axes[0].set_ylabel('Sepal width', fontsize=18)
axes[1].set_xlabel('Sepal length', fontsize=18)
axes[1].set_ylabel('Sepal width', fontsize=18)
axes[0].tick_params(direction='in', length=10, width=5, colors='k', labelsize=20)
axes[1].tick_params(direction='in', length=10, width=5, colors='k', labelsize=20)
axes[0].set_title('Actual', fontsize=18)
axes[1].set_title('Predicted', fontsize=18)
axes[1].scatter(centroids[:, 0], centroids[:, 1], marker='*', s=200, c='#050505')
```

Out[51]: <matplotlib.collections.PathCollection at 0x1edb90c8d68>



۳-۵:

In [52]:
```
A=km.inertia_
print (A)
```

```
37.05070212765958
```

۳-۶:

In [75]:
```
for x in [1,2,3,4,5]:

    km = KMeans(n_clusters=x)
    km.fit(X)
    km.predict(X)
    labels = km.labels_
    centroids = km.cluster_centers_

    V=km.inertia_
    print ("n : "  )
    print(x)
    print (V)
```

```
n :
1
594.8006666666666
n :
2
133.46431822602608
n :
3
69.42973924466338
n :
4
49.43781558441559
n :
5
40.37201367521368
```

۳-۷: