Project 04

April 11, 2020

```
[54]: #Suja Basnet
      #Project 04

      from sklearn import datasets
      import matplotlib.pyplot as plt
      import pandas as pd
      from sklearn.cluster import KMeans
```

```
[55]: #Load the data
      iris = datasets.load_iris()
```

```
[56]: #Defining Target and Predictors
      X = iris.data[:, :2]
      y = iris.target
      print(y)
```

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2]
```

```
[62]: plt.scatter(X[:,0], X[:,1], c=y, cmap='gist_rainbow')
      plt.xlabel('Sepa1 Length', fontsize=18)
      plt.ylabel('Sepal Width', fontsize=18)
```

```
[62]: Text(0, 0.5, 'Sepal Width')
```
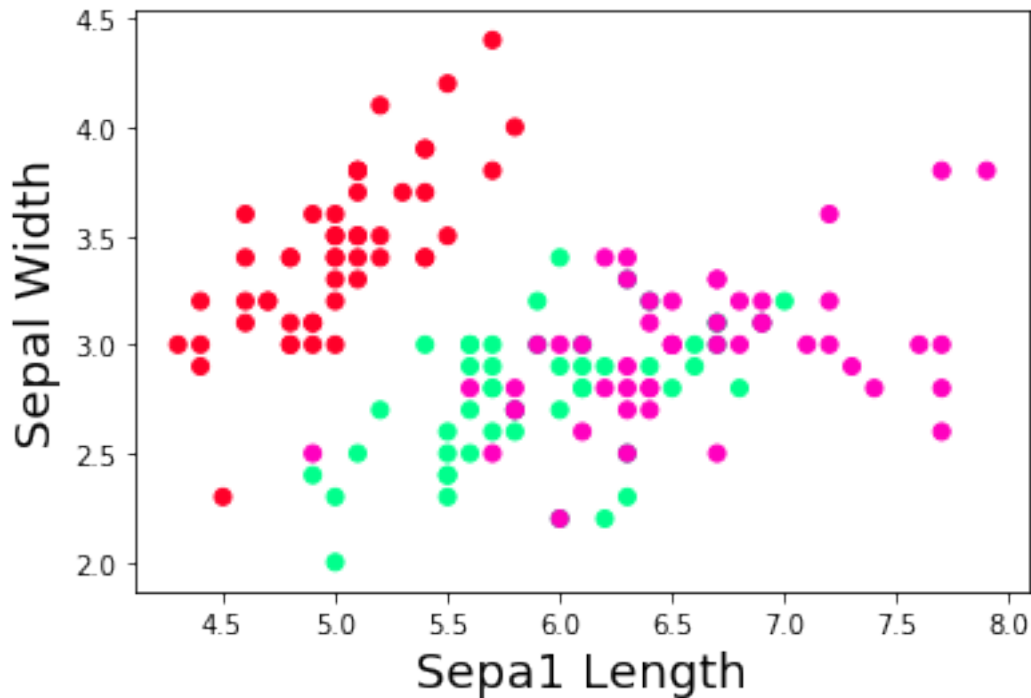
```
[58]: km = KMeans(n_clusters = 3, n_jobs = 4, random_state=21)
      km.fit(X)
```

```
[58]: KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
             n_clusters=3, n_init=10, n_jobs=4, precompute_distances='auto',
             random_state=21, tol=0.0001, verbose=0)
```

```
[59]: #Identify center points of the data
      centers = km.cluster_centers_
      print(centers)
```

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

```
[60]: #this will tell us to which cluster does the data observations belong.
      new_labels = km.labels_
      # 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)
plt.scatter(km.cluster_centers_[:, 0], km.cluster_centers_[:, 1], s = 200, c =␣
 ↪'red', label = 'Centroids')
plt.legend()
```

[60]: <matplotlib.legend.Legend at 0x12387dda0>