IE 345 - K “Introduction to Deep Learning: Fundamentals Concepts”

Prof. Yuzo

Clustering

pg. 73 - 77

In [1]:
```
# import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
dataset = pd.read_csv('Iris.csv')
dataset.head(5)
```

Out[2]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

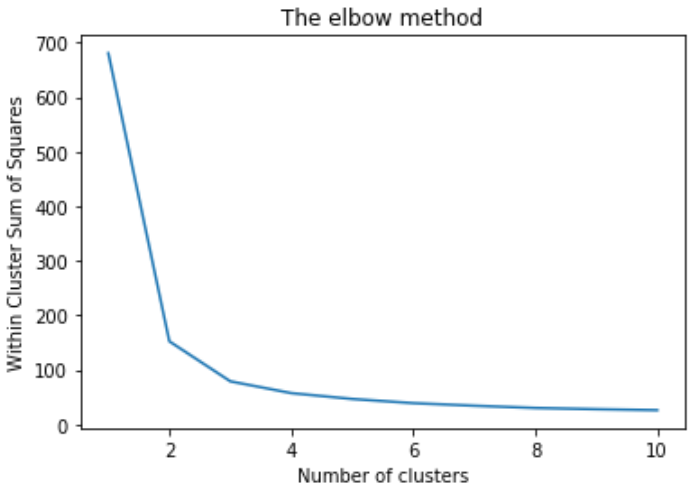
In [3]:
```
x = dataset.drop(['Id', 'Species'], axis=1)
x = x.values # select values and convert dataframe to numpy array
```

In [4]:
```
# finding the optimum number of clusters for k-means classification
from sklearn.cluster import KMeans

wcss = [] # array to hold sum of squared distances within clusters

for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state
=0)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)
```

In [5]:
```
# Plotting the results into a line graph, allowing us to observe "The elbow"
plt.plot(range(1, 11), wcss)
plt.title('The elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('Within Cluster Sum of Squares') #within cluster sum of squares
plt.show()
```



In [6]:
```
# creating the kmeans object
kmeans = KMeans(n_clusters=3, init='k-means++', max_iter=300, n_init=10, random_state=0)
y_kmeans = kmeans.fit_predict(x)
```

In [7]:
```
# visualising the clusters
plt.figure(figsize=(8, 6))
plt.scatter(x[y_kmeans == 0, 0], x[y_kmeans == 0, 1], s=100, c='red', label='Iris-setosa'
)
plt.scatter(x[y_kmeans == 1, 0], x[y_kmeans == 1, 1], s=100, c='blue', label='Iris-versic
olour')
plt.scatter(x[y_kmeans == 2, 0], x[y_kmeans == 2, 1], s=100, c='green', label='Iris-virgi
nica')

# plotting the centroids of the clusters
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=100, c='yello
w', marker='^', label='Centroids')
plt.legend(loc=0)
```

Out[7]:
```
<matplotlib.legend.Legend at 0x23715025630>
```