MACHINE LEARNING

*Example 2

K-means clusterin using sklearn

In this example we will model data of handball players in attempt to classify players height with scoring capacity

In [2]:

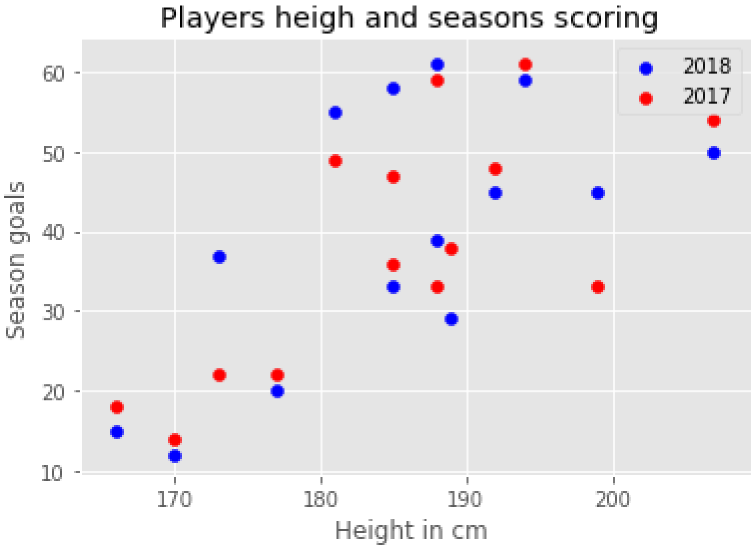
```
import matplotlib.pyplot as plt
from matplotlib import style
style.use("ggplot")

import numpy as np
from sklearn.cluster import KMeans

# datasets for an handball team

Size = [170 , 177, 185, 188, 166, 192, 181, 207, 188, 185, 173, 189, 194, 199 ]
Goals2018 = [12,  20,   33,   39, 15,  45, 55,  50,  61,  58,  37,  29,  59,  45 ]
Goals2017 = [14,  22,   36,   33, 18,  48, 49,  54,  59,   47, 22, 38,   61,  33 ]


plt.xlabel('Height in cm ')
plt.ylabel('Season goals')
plt.title("Players heigh and seasons scoring ")
plt.scatter(Size, Goals2018, color = "blue")
plt.scatter(Size, Goals2017, color = "red" )
plt.legend(('2018', '2017'))
plt.show()
```



In [3]:

```
# generating a numpy array with the data
DATA = []
for i in range(len(Size)):
    datai2017 = [Size[i], Goals2017[i] ]
    datai2018 = [Size[i], Goals2018[i] ]
    DATA.append(datai2017 )
    DATA.append(datai2018)

X = np.array(DATA)
```

In [4]:

```
# Fitting K-means algortithm on data to a predifined number of clusters
# Imagin we want to classify the players in 3 groups of height
kmeans = KMeans(n_clusters = 3)
kmeans.fit(X)
```

Out[4]:

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)
```

In [5]:

```
# geting the centriods and lables for the data fitted

centroids = kmeans.cluster_centers_
labels = kmeans.labels_

print(centroids)
print(labels)
```

```
[[187.          34.75      ]
 [191.76923077  53.15384615]
 [171.28571429  17.57142857]]
[2 2 2 2 0 0 0 0 2 2 1 1 1 1 1 1 1 1 1 1 2 0 0 0 1 1 0 1]
```

In [9]:

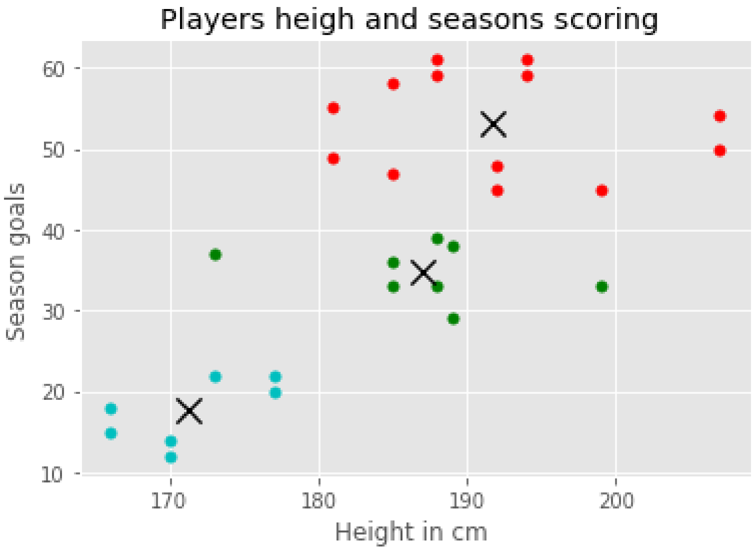
```
# Ploting and visualization of centroids and lables


plt.xlabel('Height in cm ')
plt.ylabel('Season goals')
plt.title("Players heigh and seasons scoring ")

colors = ["g.", "r." , "c.", "y."]
for i in range(len(X)):
    print ("cordinate: ", X[i] ,"label: ", labels[i] )
    plt.plot(X[i][0] , X[i][1] , colors[labels [i]], markersize = 10)

plt.scatter(centroids [:, 0] , centroids[:, 1], marker = "x", s=150, linewidths = 5, zo
rder =10, color = "black")
plt.show()
```

```
cordinate:  [170  14] label:  2
cordinate:  [170  12] label:  2
cordinate:  [177  22] label:  2
cordinate:  [177  20] label:  2
cordinate:  [185  36] label:  0
cordinate:  [185  33] label:  0
cordinate:  [188  33] label:  0
cordinate:  [188  39] label:  0
cordinate:  [166  18] label:  2
cordinate:  [166  15] label:  2
cordinate:  [192  48] label:  1
cordinate:  [192  45] label:  1
cordinate:  [181  49] label:  1
cordinate:  [181  55] label:  1
cordinate:  [207  54] label:  1
cordinate:  [207  50] label:  1
cordinate:  [188  59] label:  1
cordinate:  [188  61] label:  1
cordinate:  [185  47] label:  1
cordinate:  [185  58] label:  1
cordinate:  [173  22] label:  2
cordinate:  [173  37] label:  0
cordinate:  [189  38] label:  0
cordinate:  [189  29] label:  0
cordinate:  [194  61] label:  1
cordinate:  [194  59] label:  1
cordinate:  [199  33] label:  0
cordinate:  [199  45] label:  1
```



In [ ]: