Lab04

October 31, 2016

1 AKARI ISHIKAWA 163282

In [65]:
```
%matplotlib inline
import numpy as np
import pandas as pandas
import sklearn
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
```

In [52]:
```
data = np.genfromtxt('cluster-data.csv', delimiter=',')
data = data[1:, :]
data.reshape(1,-1)
print(data.shape)
output = np.genfromtxt('cluster-data-class.csv', delimiter =',')
output = output[1:]
print(output.shape)
for i in range(len(output)):
    if output[i] == 1:
        output[i] = 0
print(output)
```

```
(5417, 3)
(5417,)
[ 0.  0.  0. ...,  7.  7.  7.]
```

In [23]:
```
kmeans_model = KMeans(n_clusters=5, n_init = 5).fit(data)
```

In [39]:
```
sklearn.metrics.calinski_harabaz_score(data,kmeans_model.labels_)
```

Out[39]:
```
8013.0529636691754
```

2 Fazendo as métricas internas

In [73]:
```
data_interno = []
K_interno = []
for i in range(2,10):
```

```
            kmeans_model = KMeans(n_clusters=i, n_init = 5).fit(data)
            X = sklearn.metrics.calinski_harabaz_score(data,kmeans_model.labels_)
            print("K =" ,i, " --" , X)
            data_interno.append(kmeans_model.labels_)
            K_interno.append(X)
```

```
K = 2  -- 10844.820429
K = 3  -- 11782.805117
K = 4  -- 9656.37296522
K = 5  -- 8617.0197791
K = 6  -- 8892.08613882
K = 7  -- 8699.78512693
K = 8  -- 8291.01328455
K = 9  -- 8010.52789008
```

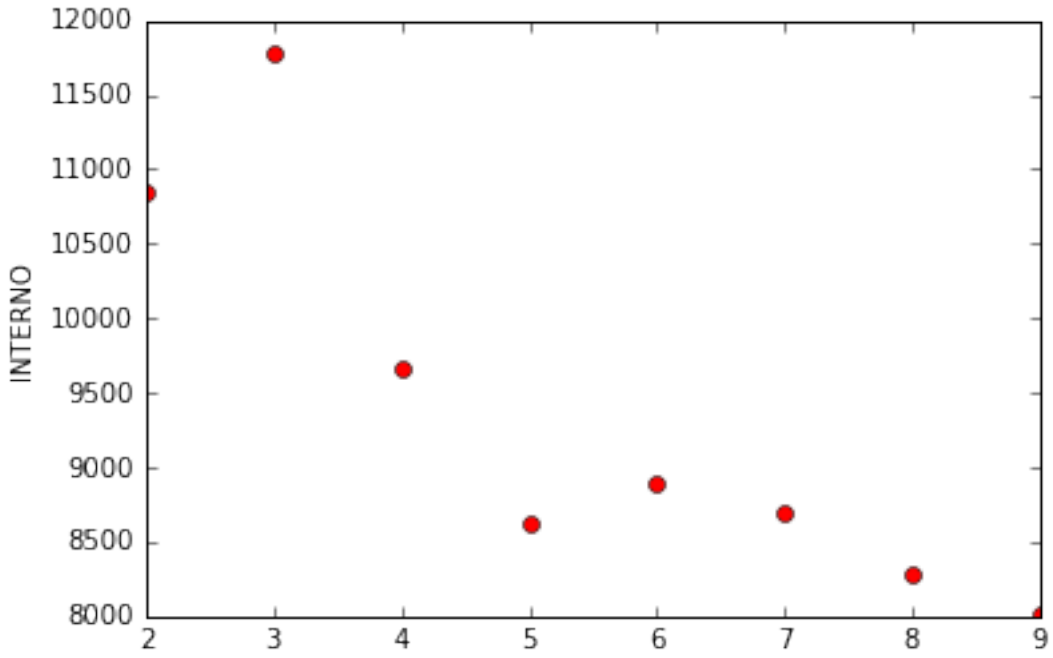
3 Agora vamos fazer as métricas externas.

```
In [62]: K_externo = []
         for i in range(2,10):
             kmeans_model = KMeans(n_clusters=i, n_init = 5).fit(data)
             Y =  sklearn.metrics.adjusted_mutual_info_score(output, data_interno[i
             K_externo.append(Y)
             print("K =" ,i, " --" , Y)
```

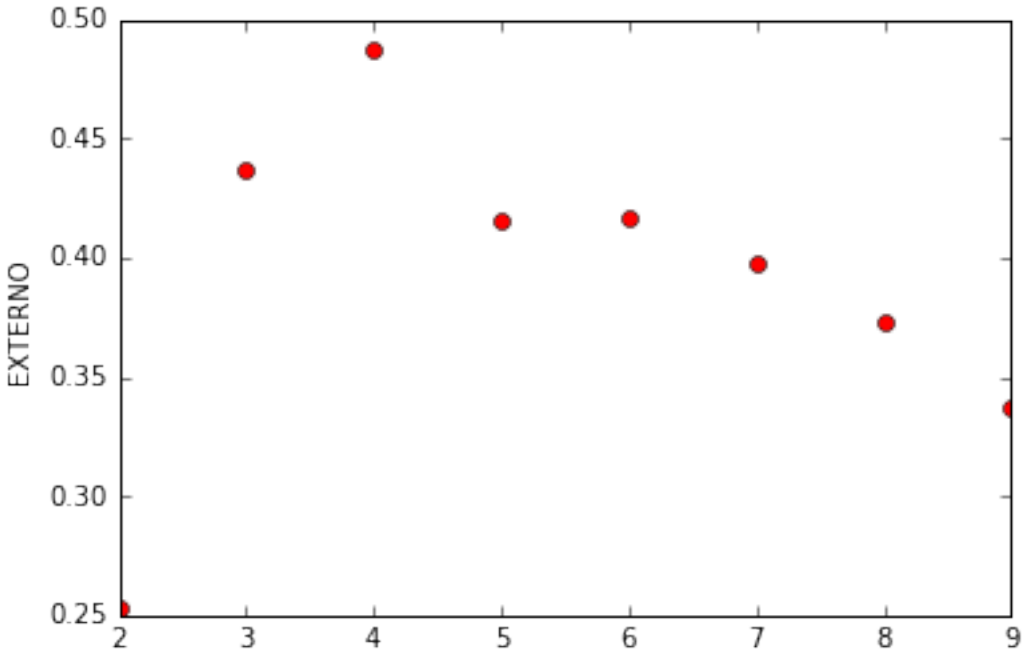
```
K = 2  -- 0.252702262904
K = 3  -- 0.436861342571
K = 4  -- 0.487142254615
K = 5  -- 0.415699196028
K = 6  -- 0.417216150815
K = 7  -- 0.397269958753
K = 8  -- 0.372651736938
K = 9  -- 0.337349506942
```

4 Plotando os valores de K interno e externo

```
In [76]: plt.plot([2,3,4,5,6,7,8,9], K_interno, 'ro')
         plt.ylabel('INTERNO')
         plt.show()
```




```
In [77]: plt.plot([2,3,4,5,6,7,8,9], K_externo, 'ro')
         plt.ylabel('EXTERNO')
         plt.show()
```



5 Podemos então escolher o K=3 para o interno e o K=4 para o externo.