

ex4

March 14, 2017

```
In [168]: import numpy as np
          from sklearn.cluster import KMeans
          import matplotlib.pyplot as plt
          from sklearn.decomposition import PCA
          from sets import Set

#Muscle-skeletal - 0
#Lung - 1
#Thyroid - 2
#Adipose-subcutaneous - 3
#Whole blood - 4

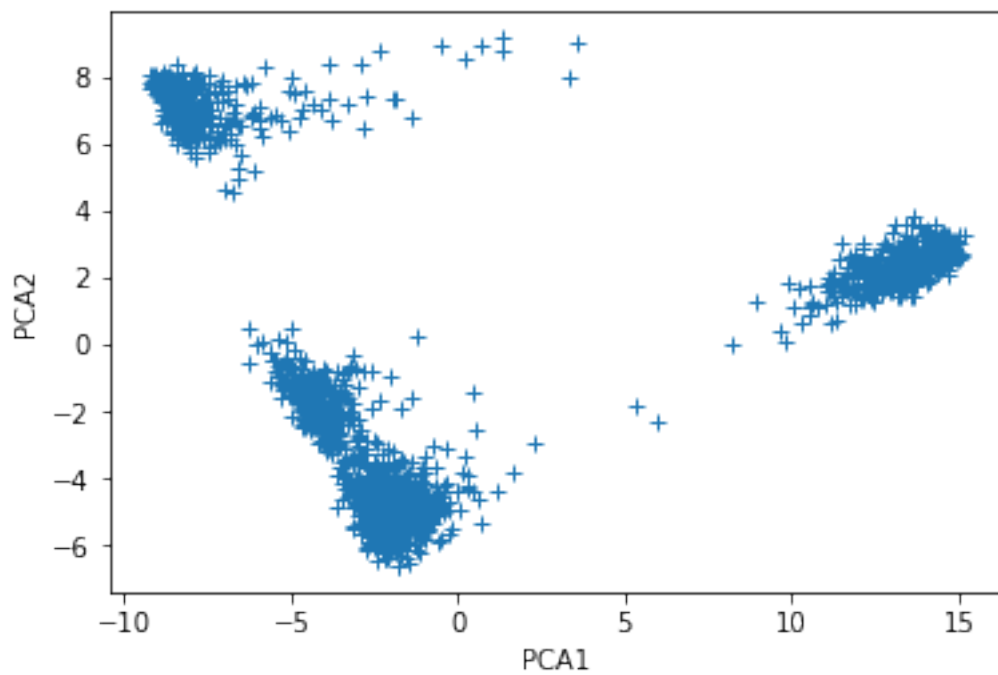
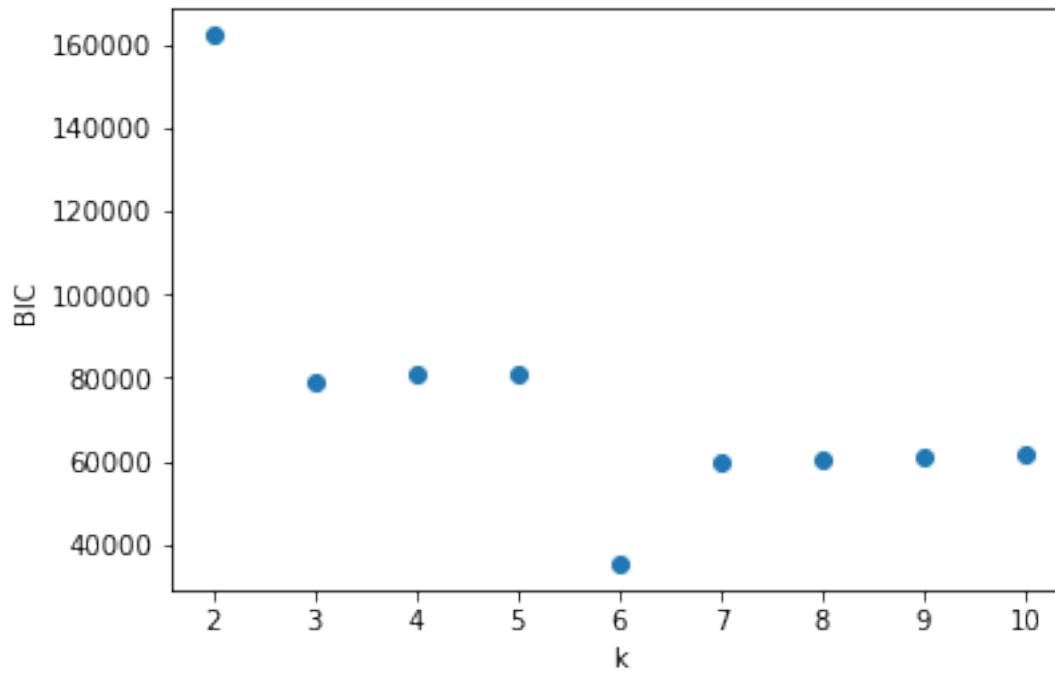
sample_labels = np.genfromtxt('class_labels.txt', delimiter='\t', skip_header=1)
data = np.genfromtxt('expr.txt', delimiter='\t', skip_header=1, unpack=True)
init_clusters = np.array(data[:5])
kmeans_5 = KMeans(n_clusters=5, max_iter=10, n_init=1, init=init_clusters)
cluster_sizes = np.bincount(kmeans_5.labels_)
cluster_centers_5means = kmeans_5.cluster_centers_
bic_values = []

for k in range(2, 11):
    k_init_clusters = np.array(data[:k])
    kmeans = KMeans(n_clusters=k, max_iter=10, n_init=1, init=k_init_clusters)
    bic_k = 2*(kmeans.inertia_) + k*100*np.log(1816)
    bic_values.append(bic_k)

plt.plot(range(2,11), bic_values, "o")
plt.ylabel("BIC")
plt.xlabel("k")
plt.show()

kmeans_6 = KMeans(n_clusters=6, max_iter=10, n_init=1, init=np.array(data[:6]))
pca_2 = PCA(n_components=2)
pca_2.fit(data)
pca_2_transform = pca_2.transform(data)
```

```
plt.figure()
plt.plot(pca_2_transform[:, :1], pca_2_transform[:, 1:], "+")
plt.xlabel("PCA1")
plt.ylabel("PCA2")
plt.show()
```



```
In [138]: kmeans_6.labels_
```

```
Out[138]: array([0, 4, 0, ..., 2, 2, 2])
```

```
In [91]: data.shape
```

```
Out[91]: (1816L, 100L)
```

```
In [124]: bic_values
```

```
Out[124]: [162305.77464085186,  
          78937.837882456311,  
          80992.051168330872,  
          80813.509995936955,  
          35347.057784773169,  
          60018.177057170884,  
          60551.217906929654,  
          61080.751143011839,  
          61700.35071524122]
```

```
In [147]:
```

```
Out[147]: array([[ 0.12134886,  0.12180025, -0.08331284, -0.08313873, -0.08184658,  
                  0.12052692,  0.12274785,  0.12325224,  0.12287417,  0.12034047,  
                  0.11824607,  0.12234112,  0.12089478, -0.02027089,  0.12157094,  
                  0.12381356, -0.03909889,  0.12208834,  0.12227762,  0.11924253,  
                  0.11942682,  0.11798485,  0.10919755,  0.1214834 ,  0.1228597 ,  
                 -0.09607811,  0.12066949, -0.03405653,  0.12520849,  0.12281748,  
                  0.1224944 ,  0.12244409, -0.03196414, -0.08907213,  0.12255935,  
                 -0.0320245 ,  0.12127185, -0.07836238,  0.12189056,  0.12348128,  
                  0.11900591,  0.12134052,  0.11529161, -0.09652548,  0.12001621,  
                  0.12197577,  0.02001903, -0.08913706,  0.12334259, -0.08426228,  
                  0.1227725 ,  0.12060867,  0.01907004,  0.10119018, -0.01890543,  
                 -0.093102 ,  0.12481836, -0.08120722, -0.0960269 , -0.09306007,  
                  0.12393294, -0.1041129 , -0.07247603,  0.00029912, -0.08774367,  
                 -0.10525293,  0.1196841 , -0.09935697, -0.08588656,  0.00510166,  
                 -0.04002951,  0.10014642,  0.01728455,  0.12338048, -0.02258138,  
                 -0.02989612,  0.06034685,  0.11807078, -0.00808817,  0.12331616,  
                 -0.09383576, -0.08792935,  0.12074567, -0.08933808,  0.10051448,  
                  0.0145862 ,  0.11654433,  0.11659074,  0.03434505,  0.12347569,  
                  0.11386584, -0.00030305,  0.00550983,  0.12196163, -0.10498737,  
                 -0.08831611, -0.02673118,  0.10262517,  0.12370184,  0.11819933],  
                [[ 0.06385711,  0.05810253,  0.14284659,  0.14347036,  0.14637192,  
                  0.06183489,  0.059761 ,  0.05498464,  0.0611402 ,  0.06251278,  
                  0.07796605,  0.06313581,  0.06209585, -0.10542416,  0.06500557,  
                  0.04933084, -0.07400317,  0.06251231,  0.06275786,  0.03321918,
```

```
0.06018497, -0.01174222, 0.10406751, 0.05495373, 0.0609463 ,
0.12404039, 0.0619021 , -0.0448944 , 0.03569257, 0.05996636,
0.0539805 , 0.05330312, -0.03543502, 0.13797568, 0.05643779,
-0.03466137, 0.0606403 , 0.13245 , 0.05832639, 0.05430351,
-0.00099506, 0.04857287, 0.03539521, 0.13098184, 0.05143166,
0.05805921, -0.17852028, 0.15011002, 0.0479997 , 0.15460726,
0.05773597, 0.06168317, -0.198153 , -0.03384489, -0.10494334,
0.13862327, 0.02723802, 0.16045695, 0.13527564, 0.1169413 ,
0.04740311, 0.09682003, 0.15887853, -0.10695599, 0.148855 ,
0.09301107, 0.05925815, 0.11073393, 0.02414009, -0.17922327,
-0.13155434, -0.11310886, -0.19830532, 0.05534858, -0.12209196,
-0.16177021, -0.17226115, -0.04217488, -0.13550983, 0.00481103,
0.13445568, 0.03192739, 0.06430065, 0.14357594, -0.11191452,
-0.19806216, 0.03821515, 0.06011795, -0.02247541, 0.05371512,
0.08969912, -0.17702052, -0.11868906, 0.059396 , 0.11439887,
0.14607259, -0.10640089, -0.11488446, 0.05224461, -0.05612505]
```

```
In [ ]:
```

```
In [ ]:
```