

```
In [ ] : import pandas as pd
import sklearn.cluster import KMeans
import numpy as np

In [ ] : df = pd.read_csv("normalized-diabetes.csv", header=None)
df.head()

Out[ ] :
   0      1      2      3      4      5      6      7      8
0  0.352941  0.743719  0.590164  0.353535  0.000000  0.500745  0.259091  0.617284  1.0
1  0.058824  0.427136  0.540984  0.292929  0.000000  0.396423  0.145041  0.382716  0.0
2  0.470588  0.919598  0.524590  0.000000  0.000000  0.347243  0.277686  0.395062  1.0
3  0.058824  0.447236  0.540984  0.232323  0.111111  0.418778  0.069008  0.259259  0.0
4  0.000000  0.688442  0.327869  0.353535  0.198582  0.642325  0.945455  0.407407  1.0

In [ ] : df_with_diabetes = df[df[8] == 1.0].iloc[:, :-1]
df_with_diabetes.iloc[:, :8].head()

Out[ ] :
   0      1      2      3      4      5      6      7
0  0.352941  0.743719  0.590164  0.353535  0.000000  0.500745  0.259091  0.617284
2  0.470588  0.919598  0.524590  0.000000  0.000000  0.347243  0.277686  0.395062
4  0.000000  0.688442  0.327869  0.353535  0.198582  0.642325  0.945455  0.407407
6  0.176471  0.391960  0.409836  0.323232  0.104019  0.461997  0.102479  0.320988
8  0.117647  0.989950  0.573770  0.454545  0.641844  0.454545  0.065289  0.654321

In [ ] : k_error = []
X = df_with_diabetes.iloc[:, :8]

for k in range(1,10):
    total = 0
    print('k = {}'.format(k))
    model = KMeans(n_clusters=k).fit(X)
    model.cluster_centers_

    for i in range(k):
        print("Cluster {} Center: {}".format(i, model.cluster_centers_[i]))
        print("=====")

    print('\n')

    df_with_diabetes['Cluster'] = model.labels_
    groupings = []

    for i in range(k):
        num_members = len(df_with_diabetes[df_with_diabetes['Cluster'] == i])
        print("Cluster {} Num Members: {}".format(i, num_members))

        groupings.append(df_with_diabetes[df_with_diabetes['Cluster'] == i])

    print('\n')

    for i in range(len(groupings)):
        d = 0
        centers = model.cluster_centers_[i]
        for j in range(len(groupings[i])):
            d += np.linalg.norm(groupings[i].iloc[j, :8].values - centers)

        print("Cluster {} Error: {}".format(i, d))
        total += d
        k_error.append([k, total])

    print("Final Error: {}".format(total))
    print("=====")
print(k_error)

k = 1
Cluster 0 Center: [0.28621598 0.7098365 0.58052973 0.2238806 0.11860026 0.52373379
0.22747934 0.45761931]
=====

Cluster 0 Num Members: 268

Cluster 0 Error: 118.40145767446779
Cluster 1 Error: 118.40145767446779
=====

k = 2
Cluster 0 Center: [0.10886743 0.72268057 0.54679471 0.26330469 0.14803641 0.53963788
0.25093746 0.37884651]
=====
Cluster 1 Center: [0.46356453 0.69699242 0.61426474 0.18445651 0.08916411 0.5078297
0.20402122 0.53639211]
=====

Cluster 0 Num Members: 134
Cluster 1 Num Members: 134

Cluster 0 Error: 53.40782021763029
Cluster 1 Error: 51.88207404401892
Final Error: 105.28989426164921
=====

k = 3
Cluster 0 Center: [0.09945421 0.73024918 0.60081122 0.35166094 0.19495016 0.54977184
0.26188123 0.37800687]
=====
Cluster 1 Center: [0.51506456 0.68911631 0.62914834 0.30574033 0.15389494 0.52524445
0.24890647 0.53914484]
=====
Cluster 2 Center: [0.27891606 0.70667946 0.5136305 0.00919305 0.00286875 0.4939634
0.17024329 0.46927452]
=====

Cluster 0 Num Members: 97
Cluster 1 Num Members: 82
Cluster 2 Num Members: 89

Cluster 0 Error: 34.40053353265967
Cluster 1 Error: 29.092909528865647
Cluster 2 Error: 32.98378665690488
Final Error: 96.47722971843021
=====

k = 4
Cluster 0 Center: [0.09803922 0.73131281 0.6010929 0.35343013 0.19547872 0.55101217
0.25968492 0.37471741]
=====
Cluster 1 Center: [2.31617647e-01 6.97864322e-01 0.00000000e+00 1.89393939e-02
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====
Cluster 2 Center: [0.29779412 0.69007538 0.61936475 0.02638889 0.00499409 0.49258569
0.17991736 0.48903211]
=====
Cluster 3 Center: [0.52321981 0.70603015 0.63589301 0.3112706 0.16604454 0.52543337
0.2472923 0.54191033]
=====

Cluster 0 Num Members: 96
Cluster 1 Num Members: 16
Cluster 2 Num Members: 80
Cluster 3 Num Members: 76

Cluster 0 Error: 34.00938882176532
Cluster 1 Error: 5.219265071142205
Cluster 2 Error: 25.185107299832527
Cluster 3 Error: 26.028506462827828
Final Error: 90.44226765556789
=====

k = 5
Cluster 0 Center: [0.29895246 0.71246644 0.6230631 0.00608828 0.0027203 0.49568218
0.17041775 0.49315068]
=====
Cluster 1 Center: [0.24246772 0.87020468 0.59936026 0.3481153 0.39554864 0.52055541
0.23709938 0.4849211]
=====
Cluster 2 Center: [0.54341737 0.65183058 0.6362217 0.31104698 0.10150475 0.52877723
0.24994753 0.52361356]
=====
Cluster 3 Center: [0.09333333 0.67088777 0.60590164 0.33845118 0.11965327 0.55195231
0.26945455 0.3665636]
=====
Cluster 4 Center: [2.31617647e-01 6.97864322e-01 0.00000000e+00 1.89393939e-02
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====

Cluster 0 Num Members: 73
Cluster 1 Num Members: 41
Cluster 2 Num Members: 63
Cluster 3 Num Members: 75
Cluster 4 Num Members: 16

Cluster 0 Error: 22.065569023100032
Cluster 1 Error: 14.539655883202332
Cluster 2 Error: 20.64829466963895
Cluster 3 Error: 22.38300407670221
Cluster 4 Error: 5.219265071142205
Final Error: 86.66219746209772
=====

k = 6
Cluster 0 Center: [0.21691176 0.85427136 0.58145492 0.31786616 0.42941046 0.51383197
0.19622934 0.45023148]
=====
Cluster 1 Center: [ 3.11516156e-01 7.06773303e-01 6.21911799e-01 1.15236876e-02
-1.11022302e-16 4.92265066e-01 1.64276569e-01 5.01999652e-01]
=====
Cluster 2 Center: [0.56078431 0.70619765 0.65163934 0.31178451 0.12304965 0.52771982
0.24900138 0.54279835]
=====
Cluster 3 Center: [0.11586453 0.63194762 0.6010929 0.31833486 0.10752919 0.54696744
0.19403331 0.36120464]
=====
Cluster 4 Center: [0.11508951 0.7596679 0.61083393 0.39086517 0.15494912 0.56871639
0.54035214 0.4331723 ]
=====
Cluster 5 Center: [2.31617647e-01 6.97864322e-01 0.00000000e+00 1.89393939e-02
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====

Cluster 0 Num Members: 71
Cluster 1 Num Members: 32
Cluster 2 Num Members: 60
Cluster 3 Num Members: 66
Cluster 4 Num Members: 23
Cluster 5 Num Members: 16

Cluster 0 Error: 10.285127873464464
Cluster 1 Error: 21.24095785396159
Cluster 2 Error: 18.878512177370748
Cluster 3 Error: 19.74504245947153
Cluster 4 Error: 8.769516814682317
Cluster 5 Error: 5.219265071142205
Final Error: 84.13842225009286
=====

k = 7
Cluster 0 Center: [0.16408669 0.68447501 0.59987058 0.03110048 0.01984571 0.50917719
0.17928447 0.41033138]
=====
Cluster 1 Center: [0.18147685 0.54207206 0.57778165 0.31270148 0.10389316 0.51035292
0.23526464 0.39374836]
=====
Cluster 2 Center: [0.05882353 0.8005142 0.62619138 0.39346958 0.1668316 0.60160122
0.34583894 0.36577663]
=====
Cluster 3 Center: [ 4.28180575e-01 7.40913872e-01 6.49256576e-01 4.22832981e-03
-6.93889390e-17 4.80192701e-01 1.59465693e-01 5.54694229e-01]
=====
Cluster 4 Center: [0.29411765 0.85694095 0.60655738 0.32859848 0.42205969 0.51760432
0.1803719 0.51003086]
=====
Cluster 5 Center: [2.31617647e-01 6.97864322e-01 0.00000000e+00 1.89393939e-02
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====
Cluster 6 Center: [0.56902761 0.69141626 0.64034794 0.33065347 0.11159357 0.5269625
0.26014505 0.53917863]
=====

Cluster 0 Num Members: 38
Cluster 1 Num Members: 47
Cluster 2 Num Members: 43
Cluster 3 Num Members: 43
Cluster 4 Num Members: 32
Cluster 5 Num Members: 16
Cluster 6 Num Members: 49

Cluster 0 Error: 9.859933448651136
Cluster 1 Error: 13.44375659782635
Cluster 2 Error: 15.421786147767262
Cluster 3 Error: 12.38300407670221
Cluster 4 Error: 10.434651697487615
Cluster 5 Error: 5.219265071142205
Cluster 6 Error: 14.534613716832158
Final Error: 81.29701075640894
=====

k = 8
Cluster 0 Center: [0.09907121 0.78021687 0.6350302 0.42477406 0.17027498 0.56529924
0.5516094 0.42040286]
=====
Cluster 1 Center: [ 4.27450980e-01 7.30541597e-01 6.42622951e-01 4.40404040e-03
-8.32667268e-17 4.81304852e-01 1.52782369e-01 5.41563786e-01]
=====
Cluster 2 Center: [0.21882353 0.85085427 0.56459016 0.32525253 0.47621749 0.51868852
0.17616529 0.4617284 ]
=====
Cluster 3 Center: [0.08936652 0.65297642 0.59836066 0.35878011 0.13079651 0.5726241
0.19820407 0.34924027]
=====
Cluster 4 Center: [0.4743833 0.5091587 0.60629297 0.31150212 0.06630824 0.51314841
0.25489203 0.49422541]
=====
Cluster 5 Center: [2.31617647e-01 6.97864322e-01 0.00000000e+00 1.89393939e-02
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====
Cluster 6 Center: [0.5396983 0.81162221 0.65468684 0.33022533 0.16718191 0.52986358
0.25657978 0.56220323]
=====
Cluster 7 Center: [0.14634146 0.70020836 0.6107557 0.05173688 0.03090584 0.50052706
0.19859907 0.41734417]
=====

Cluster 0 Num Members: 19
Cluster 1 Num Members: 45
Cluster 2 Num Members: 25
Cluster 3 Num Members: 52
Cluster 4 Num Members: 31
Cluster 5 Num Members: 16
Cluster 6 Num Members: 39
Cluster 7 Num Members: 41

Cluster 0 Error: 7.044941401259049
Cluster 1 Error: 12.605349830173596
Cluster 2 Error: 8.065599264522419
Cluster 3 Error: 14.321218343154126
Cluster 4 Error: 9.167759834798206
Cluster 5 Error: 5.219265071142205
Cluster 6 Error: 11.186560715839516
Cluster 7 Error: 11.4905460551997
Final Error: 79.10124051608882
=====

k = 9
Cluster 0 Center: [0.23202614 0.83528755 0.56921676 0.33501684 0.55227213 0.51432356
0.17825987 0.45816187]
=====
Cluster 1 Center: [0.09033613 0.72819454 0.61358314 0.31908369 0.14460908 0.54654567
0.17677834 0.36397707]
=====
Cluster 2 Center: [ 2.09653092e-01 6.47854658e-01 6.11391341e-01 3.10800311e-03
-5.55111512e-17 5.18323207e-01 1.84424666e-01 4.15321304e-01]
=====
Cluster 3 Center: [ 4.06417112e-01 7.84071875e-01 6.38350720e-01 5.50964187e-03
-6.93889390e-17 4.70532448e-01 1.54507889e-01 5.87729143e-01]
=====
Cluster 4 Center: [0.67420814 0.6132586 0.6276797 0.28049728 0.07069467 0.53141121
0.21710108 0.536522 ]
=====
Cluster 5 Center: [0.08359133 0.79317641 0.62122519 0.44019139 0.16728879 0.58153581
0.54243149 0.41910331]
=====
Cluster 6 Center: [1.96078431e-01 7.01172529e-01 0.00000000e+00 0.00000000e+00
2.77555756e-17 5.07730999e-01 1.77944215e-01 3.88888889e-01]
=====
Cluster 7 Center: [0.47404844 0.82707656 0.65260366 0.33333333 0.18797803 0.5317349
0.262421 0.56499637]
=====
Cluster 8 Center: [0.27521008 0.48671931 0.56264637 0.31457431 0.0835022 0.51000639
0.28856257 0.41490 ]
=====

Cluster 0 Num Members: 18
Cluster 1 Num Members: 56
Cluster 2 Num Members: 29
Cluster 3 Num Members: 33
Cluster 4 Num Members: 26
Cluster 5 Num Members: 19
Cluster 6 Num Members: 15
Cluster 7 Num Members: 34
Cluster 8 Num Members: 28

Cluster 0 Error: 5.927895195815207
Cluster 1 Error: 15.245429542148461
Cluster 2 Error: 10.161446374940372
Cluster 3 Error: 8.793950447322083
Cluster 4 Error: 8.046270096244562
Cluster 5 Error: 7.074360585118303
Cluster 6 Error: 4.46775968039878
Cluster 7 Error: 9.268257296379998
Cluster 8 Error: 8.24439459771119
Final Error: 77.22916381598895
=====

[[1, 118.40145767446779], [2, 105.28989426164921], [3, 96.47722971843021], [4, 90.44226765556789], [5, 86.66219746209772], [6, 84.13842225009286], [7, 81.297010756408
94], [8, 79.10124051608882], [9, 77.22916381598895]]

In [ ] : import matplotlib.pyplot as plt
k_error = np.array(k_error)
plt.plot(k_error[:,0], k_error[:,1], 'bx-')
plt.xlabel('Values of K')
plt.ylabel('Total Error')
plt.title('The Elbow Method using Distortion')
plt.show()
```

