# -\*- coding: utf-8 -\*-

"""

Created on Tue Apr 24 15:54:40 2018

@author: Arpit

"""

import numpy as np

from scipy.cluster.hierarchy import linkage, dendrogram

import matplotlib.pyplot as plt

import pandas as pd

#x=np.array([[15,30],[60,25],[15,78],[10,40]])

mydata=pd.read\_csv('DM\_PR8\_Full.csv')

X=mydata.iloc[:,[1,6,7,8]].values # loading category and age columns

# finding best number of clusters

import scipy.cluster.hierarchy as sc

dendrogram=sc.dendrogram(sc.linkage(X,method='ward'))

# Run until finding best number of clusters then run below code

# applying hierarchical clustering to our dataset

from sklearn.cluster import AgglomerativeClustering

Hierarchical=AgglomerativeClustering(n\_clusters=3,affinity='euclidean',linkage='ward')

PredictH=Hierarchical.fit\_predict(X)

# visualization of our hierarchical clustering

plt.scatter(X[PredictH==0,0],X[PredictH==0,1],s=500,c='red',label='cluster 1')

plt.scatter(X[PredictH==1,0],X[PredictH==1,1],s=500,c='blue',label='cluster 2')

plt.scatter(X[PredictH==2,0],X[PredictH==2,1],s=500,c='green',label='cluster 3')

plt.title('Hierarchical clustering')

plt.xlabel('Category')

plt.ylabel('Age')

plt.legend()

plt.show()



