M.S.SANJAY

15BCE0517

L7+L8 MACHINE LEARNING

EXERCISE -2

K MODES CLUSTERING

Sir I have written my own code.please see

I have manually set number of clusters,dimensions and data points. It can also be taken as user input (ex. dim=input(“Enter number of dimensions”) but for convenience ,I have set it manually.

Pts is s 2-d array that represwents data set

dim specifies number fo attributes

n\_cluster specifies number of clusters

centroids\_old and centroids\_new are 2-d array that is used to store centroid points.2 arrays are used for stopping condition checking i.e. to check if centroids have changed or not

pts\_cluster is a array whose lenght is equal to number of data points and each entry specifies to what cluster the data point belongs to(cluster number starts from 0)

I have only fixed what value each attribute can take ('a','b','c' and 'd').rest all can be taken from user input also.

CODE:

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

"""

Spyder Editor

This temporary script file is located here:

/home/likewise-open/VITUNIVERSITY/15bce0517/.spyder2/.temp.py

"""

dim=4

def calcdist(pts,oldcentroid):

ans=[]

print("ponts")

print(pts)

print("old centrpid")

print(oldcentroid)

for i in range(0,len(oldcentroid)):

dist=0;

for j in range(0,dim):

if (pts[j]!=oldcentroid[i][j]):

dist=dist+1

ans.append(dist)

print("distance")

print(ans)

return ans

def calccentroid(oldcentroid,ptscluster,pts):

ans=[]

for i in range(0,len(oldcentroid)):

ans1=[]

modea=[]

modeb=[]

modec=[]

moded=[]

last=[]

for temp in range(0,dim):

ans1.append('a')

for temp in range(0,dim):

modea.append(0)

modeb.append(0)

modec.append(0)

moded.append(0)

for j in range(0,len(ptscluster)):

if (ptscluster[j]==i):

for k in range(0,dim):

if (pts[j][k]=='a'):

modea[k]+=1

elif (pts[j][k]=='b'):

modeb[k]+=1

elif (pts[j][k]=='c'):

modec[k]+=1

elif (pts[j][k]=='d'):

moded[k]+=1

last=pts[j]

for p in range(0,dim):

set=0

if ( (modea[p]>modeb[p]) and (modea[p]>modec[p]) and (modea[p]>moded[p]) ):

ans1[p]='a'

set=1

elif ( (modeb[p]>modea[p]) and (modeb[p]>modec[p]) and (modeb[p]>moded[p]) ):

ans1[p]='b'

set=1

elif ( (modec[p]>modea[p]) and (modec[p]>modeb[p]) and (modec[p]>moded[p]) ):

ans1[p]='c'

set=1

elif ( (moded[p]>modea[p]) and (moded[p]>modeb[p]) and (moded[p]>modec[p]) ):

ans1[p]='d'

set=1

if(set==0):

ans1[p]=last[p]

ans.append(ans1)

return ans

pts=[['a','b','c','d'],

['c','d','a','a'],

['b','a','b','c'],

['a','c','d','b'],

['b','d','a','d'],

#['c','c','d','b'],

#['b','b','b','c']

]

n\_cluster=2

centroids\_old=[];

pts\_cluster=[];

for i in range(0,len(pts)):

pts\_cluster.append([0])

for i in range(0,n\_cluster):

centroids\_old.append(pts[i])

#print(centroids\_old)

centroids\_new=[]

flag=1

while (flag):

for i in range(0,len(pts)):

print("point number:")

print(i)

dist=calcdist(pts[i],centroids\_old)

index=dist.index(min(dist))

print("cluster assigned")

print(index)

pts\_cluster[i]=index

centroids\_new=calccentroid(centroids\_old,pts\_cluster,pts)

print("new centroid")

print(centroids\_new)

if (centroids\_old==centroids\_new):

flag=0

else:

centroids\_old=centroids\_new

for i in range(0,n\_cluster):

print("CLUSTER NUMBER:")

print(i)

print("CENTROID:")

print(centroids\_new[i])

print("POINTS ARE:")

for j in range (0,len(pts\_cluster)):

if (pts\_cluster[j]==i):

print(pts[j])

***OUTPUT:***

>>> runfile('/home/likewise-open/VITUNIVERSITY/15bce0517/Desktop/ML\_LAB\_2\_15BCE0517.py', wdir=r'/home/likewise-open/VITUNIVERSITY/15bce0517/Desktop')

point number:

0

ponts

['a', 'b', 'c', 'd']

old centrpid

[['a', 'b', 'c', 'd'], ['c', 'd', 'a', 'a']]

distance

[0, 4]

cluster assigned

0

point number:

1

ponts

['c', 'd', 'a', 'a']

old centrpid

[['a', 'b', 'c', 'd'], ['c', 'd', 'a', 'a']]

distance

[4, 0]

cluster assigned

1

point number:

2

ponts

['b', 'a', 'b', 'c']

old centrpid

[['a', 'b', 'c', 'd'], ['c', 'd', 'a', 'a']]

distance

[4, 4]

cluster assigned

0

point number:

3

ponts

['a', 'c', 'd', 'b']

old centrpid

[['a', 'b', 'c', 'd'], ['c', 'd', 'a', 'a']]

distance

[3, 4]

cluster assigned

0

point number:

4

ponts

['b', 'd', 'a', 'd']

old centrpid

[['a', 'b', 'c', 'd'], ['c', 'd', 'a', 'a']]

distance

[3, 2]

cluster assigned

1

new centroid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

point number:

0

ponts

['a', 'b', 'c', 'd']

old centrpid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

distance

[3, 3]

cluster assigned

0

point number:

1

ponts

['c', 'd', 'a', 'a']

old centrpid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

distance

[4, 2]

cluster assigned

1

point number:

2

ponts

['b', 'a', 'b', 'c']

old centrpid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

distance

[4, 3]

cluster assigned

1

point number:

3

ponts

['a', 'c', 'd', 'b']

old centrpid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

distance

[0, 4]

cluster assigned

0

point number:

4

ponts

['b', 'd', 'a', 'd']

old centrpid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

distance

[4, 0]

cluster assigned

1

new centroid

[['a', 'c', 'd', 'b'], ['b', 'd', 'a', 'd']]

CLUSTER NUMBER:

0

CENTROID:

['a', 'c', 'd', 'b']

POINTS ARE:

['a', 'b', 'c', 'd']

['a', 'c', 'd', 'b']

CLUSTER NUMBER:

1

CENTROID:

['b', 'd', 'a', 'd']

POINTS ARE:

['c', 'd', 'a', 'a']

['b', 'a', 'b', 'c']

['b', 'd', 'a', 'd']

>>>