



CENTRO DE CIENCIAS BÁSICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
APRENDIZAJE INTELIGENTE
6° "A"

PRÁCTICA 5: CLUSTERING JERÁRQUICO

Profesor: Francisco Javier Luna Rosas

Alumnos:

Espinoza Sánchez Joel Alejandro

Gómez Garza Dariana

González Arenas Fernando Francisco

Fecha de Entrega: Aguascalientes, Ags., 7 de abril de 2021

Práctica 5: Clustering Jerárquico

Objetivo:

Mediante el desarrollo de esta práctica, implementar el clustering jerárquico para el análisis de conglomerados.

Introducción:

a

Pregunta de Investigación:

¿Cómo se puede implementar un clustering jerárquico en Python?

Predicción:

Creemos que la implementación consistirá en buscar funciones equivalentes a las existentes en R para trabajar ahora en Python.

Materiales:

Una computadora con Python y el entorno Anaconda.

Método (Variables):

Dependiente: La predicción que se realizará.

Independiente: Los datos a tratar con el conjunto de datos de estudiantes.

Controlada: El algoritmo de clustering jerárquico a implementar.

Seguridad:

Realmente no se trabajó en campo, por lo que no se corren riesgos con la práctica.

Procedimiento:

1.- Se importaron las librerías de trabajo de ciencia de datos de Python:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.cluster.hierarchy as sch
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
```

2.- Se introdujeron los datos del archivo csv:

```
Estudiantes = pd.read_csv('C:/Users/Dell/Desktop/wenas/EjemploEstudiantes.csv',
                          engine = 'python', index_col = 0, sep=';', decimal=',')
print(Estudiantes.head())
```

3.- Se hizo el análisis de las calificaciones:

```
calificaciones = Estudiantes.iloc[:, [0,1,2,3,4]].values
print(calificaciones)
```

4.- Se realizó el análisis de clustering jerárquico por medio de la función de Python:

```
CJ = linkage(calificaciones, 'ward')
dendrogram = sch.dendrogram(CJ)
clusters = fcluster(CJ, t = 2, criterion = 'distance')
```

5.- Se añadieron los datos a la tabla csv original:

```
Estudiantes['CJ'] = clusters
print(Estudiantes)
Estudiantes.to_csv('C:/Users/Dell/Desktop/wenas/EjemploEstudiantesConCJ.csv')
```

Obtención y Procesamiento de Datos:

Al introducir el archivo al código, se probó su funcionamiento:

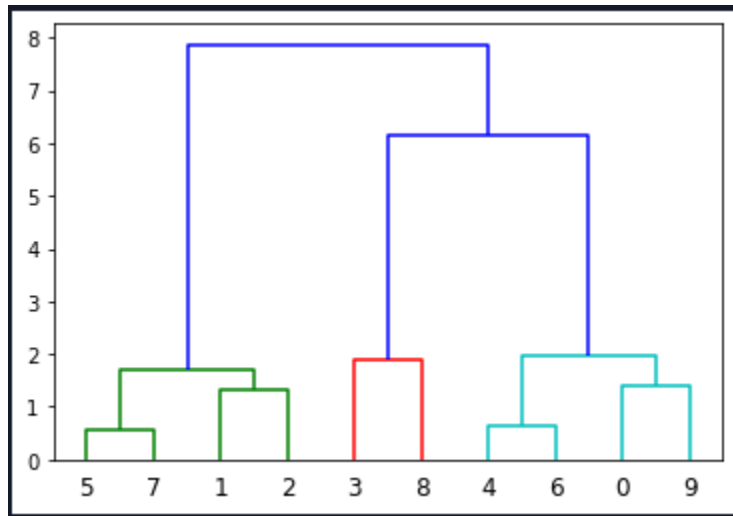
```
<class 'pandas.core.frame.DataFrame'>  
Index: 10 entries, Lucia to Maria  
Data columns (total 5 columns):  
Matematicas      10 non-null float64  
Ciencias         10 non-null float64  
Espanol          10 non-null float64  
Historia         10 non-null float64  
EdFisica         10 non-null float64  
dtypes: float64(5)  
memory usage: 480.0+ bytes
```

	Matematicas	Ciencias	Espanol	Historia	EdFisica
Lucia	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Ines	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andres	6.0	6.0	7.8	8.9	7.3

Se procesaron las calificaciones:

```
[[7.  6.5 9.2 8.6 8. ]  
 [7.5 9.4 7.3 7.  7. ]  
 [7.6 9.2 8.  8.  7.5]  
 [5.  6.5 6.5 7.  9. ]  
 [6.  6.  7.8 8.9 7.3]  
 [7.8 9.6 7.7 8.  6.5]  
 [6.3 6.4 8.2 9.  7.2]  
 [7.9 9.7 7.5 8.  6. ]  
 [6.  6.  6.5 5.5 8.7]  
 [6.8 7.2 8.7 9.  7. ]]
```

Después puede observarse la clasificación hecha por el clustering jerárquico:



Y por último puede observarse la actualización de datos de la tabla:

	Matematicas	Ciencias	Espanol	Historia	EdFisica	CJ
Lucia	7.0	6.5	9.2	8.6	8.0	3
Pedro	7.5	9.4	7.3	7.0	7.0	1
Ines	7.6	9.2	8.0	8.0	7.5	1
Luis	5.0	6.5	6.5	7.0	9.0	2
Andres	6.0	6.0	7.8	8.9	7.3	3
Ana	7.8	9.6	7.7	8.0	6.5	1
Carlos	6.3	6.4	8.2	9.0	7.2	3
Jose	7.9	9.7	7.5	8.0	6.0	1
Sonia	6.0	6.0	6.5	5.5	8.7	2
Maria	6.8	7.2	8.7	9.0	7.0	3

Conclusiones:

Joel Alejandro Espinoza Sánchez: El clustering jerárquico nos ayuda en el análisis de conglomerados que es muy útil. En este trabajo, nos habían encargado adaptar un algoritmo de clustering jerárquico de R a Python que es muy útil para observar cómo se pueden tratar los mismos problemas con diferentes herramientas, ya que parece ser que en R dicho algoritmo es más sencillo, sin embargo es perfectamente replicable en Python aunque sea ligeramente más complicado.

Dariana Gómez Garza: Esta práctica nos ayudó a conocer un poco más el código que el maestro nos enseñó en R pero aplicado en Python.

En la realización de la práctica nos dimos cuenta de muchas funciones que tiene R para ponerse "ahorrar" ciertos pasos que en otros lenguajes.

Al aplicar el CJ en este ejercicio también nos ayudó a reforzar un poco más el tema y los temas pasados, ya que para mi parecer fue un complemento de lo ya visto de CJ, así como el funcionamiento de análisis de conglomerados en un ejercicio más específico y gracias a eso podemos tener una visión más amplia de la aplicación de un clustering jerárquico y que al mismo tiempo nos da sentido de porque es importante su aplicación.

Fernando Francisco González Arenas: En esta práctica vimos un nuevo algoritmo de clusterización para agrupar los datos de una determinada muestra de información. estos datos se agrupan utilizando el algoritmo de clustering jerárquico, el cual agrupa los datos con ayuda de un árbol y un dendrograma, el cual ayuda a definir cuántos clusters se requerirán para agrupar los datos con el corte de la gráfica a cierta altura. estos algoritmos son muy importantes en el aprendizaje no supervisado, para crear aplicaciones con inteligencia artificial que aprendan a partir de los datos proporcionados, sin necesidad de decirle la salida que se desea obtener con los mismos datos.

Referencias:

McCulloch, Warren; Walter Pitts. (1943). *A Logical Calculus of Ideas Immanent in Nervous Activity*. Inglaterra: Bulletin of Mathematical Biophysics

Farley, B.G.; W.A. Clark. (1954). *Simulation of Self-Organizing Systems by Digital Computer*. Chicago: IRE Transactions on Information Theory.

Anexos:

Anexo 1: Código del programa en lenguaje Python:

```
### Presentación
'''
    Universidad Autónoma de Aguascalientes

    Centro de Ciencias Básicas
    Departamento de Ciencias de la Computación
    Aprendizaje Inteligente
    6º "A"

    Práctica 5

    Profesor: Francisco Javier Luna Rosas
    Alumnos:
        Espinoza Sánchez Joel Alejandro
        Gómez Garza Dariana
        González Arenas Fernando Francisco

    Fecha de Entrega: 7 de abril del 2021

    Descripción: CJ
'''

### Librerías
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.cluster.hierarchy as sch
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster

### Lee datos
Estudiantes = pd.read_csv('C:/Users/Dell/Desktop/wenas/EjemploEstudiantes.csv',
                           engine = 'python', index_col = 0, sep=';',
                           decimal=',')

print(Estudiantes.head())

### Procesa calificaciones
calificaciones = Estudiantes.iloc[:,[0,1,2,3,4]].values

print(calificaciones)

### Realiza CJ
CJ = linkage(calificaciones, 'ward')

### Procesa el dendrograma
dendrogram = sch.dendrogram(CJ)
```

```
clusters = fcluster(CJ, t = 2, criterion = 'distance')

#%% Actualiza datos
Estudiantes['CJ'] = clusters

print(Estudiantes)

Estudiantes.to_csv('C:/Users/Dell/Desktop/wenas/EjemploEstudiantesConCJ.csv')
```