Logotipo

Descripción generada automáticamente

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

CURSO INTERSEMESTRAL: MACHINE Y DEEP LEARNING

ACTIVIDAD 1\_04. K-MEANS EN R(PRÁCTICA).

DOCENTE:

FRANCISCO JAVIER LUNA ROSAS

ALUMNO:

EMILIO LUNA PÉREZ, ID:244182

JUNIO-AGOSTO DE 2021

Evidencias de la Practica

Defina el algoritmo K-Means, explique el funcionamiento del algoritmo de K-Means e implemente el algoritmo K-Means en R, el dataset que se utilizará, es el dataset visto en clase de notas escolares.

K-MEANS

Este método agrupa las observaciones en k clusters distintos, donde k lo determina el analista antes de ejecutar el algoritmo. K-means encuentra los k mejores clusters, es el mejor k el que minimiza la distancia intra cluster y maximiza la distancia inter cluster. El algoritmo empleado para ello es:

(1). Especificar el numero k de clusters que se quieren crear.

(2). Seleccionar de forma aleatoria k observaciones del set de datos como centroides iniciales.

(3). Asignar cada una de las observaciones al centroide más cercano.

(4). Para cada uno de los K clusters recalcular su centroide.

(5). Repetir los pasos 3 y 4 hasta que las asignaciones no cambien o se alcance el numero maximo de iteraciones establecido.

ALGORITMO DEL K-MEANS EN R

Texto

Descripción generada automáticamente

Centro de gravedad del cluster 1

Gráfico, Gráfico de barras

Descripción generada automáticamente

Centro de gravedad del cluster 2

Gráfico, Gráfico de barras

Descripción generada automáticamente

Centro de gravedad del cluster 3

Gráfico, Gráfico de barras

Descripción generada automáticamente

Centro de gravedad de los clusters

Gráfico, Gráfico de barras

Descripción generada automáticamente

En grupos verifico a que cluster pertenece cada fila de la tabla de datos

Imagen que contiene Logotipo

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Clustering (Aprendizaje No Supervisado De Análisis De Conglomerados)

El clúster verde: Luis y Sonia: es representado por Luis y Sonia con buenas notas de educación física: son buenos deportistas. Podemos hacer notar que son malos en ciencias y matemáticas porque tienen un ángulo mayor a 180° con respecto a educación física. A su vez son malos en español e historia, pues se oponen negativamente en un Angulo mayor de 45°.

El clúster rojo: Ana, Inés, José y Pedro: parece ser el opuesto del clúster de educación física. Sus miembros son buenos en las ciencias y las matemáticas, pero son malos en deportes.

El clúster azul: Andrés, Carlos, Lucía y María: tiene a los estudiantes destacados en el área de español e historia. No son buenos en deportes, pero no son tan malos.

Conclusiones

En cuanto a aprendizaje no supervisado, este me pareció un método muy visual en cuanto al manejo de datos, el algoritmo de Kmeans con la librería FactoMiner vuelve el trabajo fácil, pero me pude dar cuenta que es algo complejo la implementación de este algoritmo pues necesitamos bases matemáticas en cuanto a distancias y otros relacionados con matemáticas.

Referencias

Download R-4.1.0 for Windows. (2021). The R-project for statistical computing. Consultado en julio 28, 2021, de R-project.org Sitio web: https://cran.r-project.org/bin/windows/base/

Download the RStudio IDE. (2021). Consultado en julio 28, 2021, de Rstudio.com Sitio web: https://www.rstudio.com/products/rstudio/download/

RPubs (2020). Método Kmeans Consultado en agosto 1,2021, de Rpubs.com Sitio web: https://www.rpubs.com/Giorgia96/kmeans

‌

‌