

# Práctica 2 de minería de datos

Diego Vera

## I. INTRODUCCIÓN

EN esta práctica se usará un dataset climatológico registrado en el sistema CLICOM, estos datos son del estado de Michoacán y brindan información sobre las precipitaciones y temperatura durante el año de esta región. Se usó R para hacer el análisis de los datos.

## II. DESARROLLO

### A. Obtención de los datos

Para la lectura del archivo se usó `read.table()` junto a la ruta donde se ubicó el archivo 'Estaciones.txt', en este caso fue `C:/Users/diego/OneDrive/Documentos/UNAM/Cursos/Data Mining/R/Estaciones.txt`, en caso de querer usar código, se tiene que cambiar la ruta de lectura del archivo.

**Nota:** Si se está trabajando en RStudio se puede usar la instrucción `dirname(rstudioapi::getSourceEditorContext())$path)` para obtener la ruta actual del script en el que se está trabajando y así obtener un atajo al archivo de datos, si es que se encuentran en el mismo directorio o cercano.

### B. Obtención de matriz de similitudes

Para obtener la matriz de similitudes se usó la distancia euclidiana, basado en el teorema de Pitágoras. Para esto se usó la instrucción `dist()`.

### C. Obtención de la cantidad adecuada de grupos

Para la obtención de la cantidad adecuada de grupos se usó el método del codo, con este método se obtuvo que la cantidad de grupos adecuada es de 5, como es demostrada en la Fig 1.

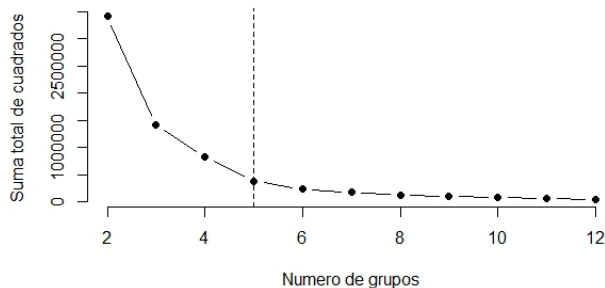


Fig. 1. Método del codo usado en los datos de las estaciones.

### D. Agrupación de los elementos

Los elementos se agruparon mediante el método jerárquico y el particional mediante 5 grupos, para el jerárquico se usó la instrucción `hclust()` junto a la matriz de distancias y para el particional se usó la instrucción `kmeans()` junto a la matriz de datos, el gráfico jerárquico y particional son mostrados en la Fig 2 y la Fig 3, respectivamente.

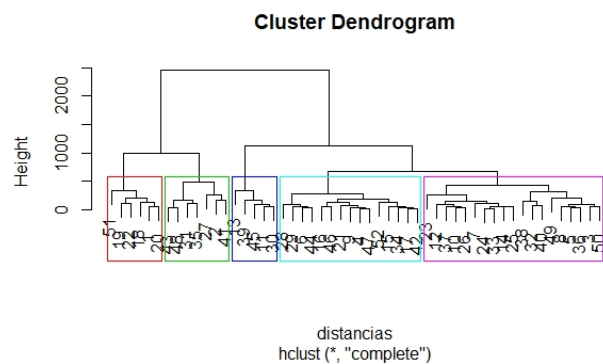


Fig. 2. Agrupación jerárquica en los datos de las estaciones.

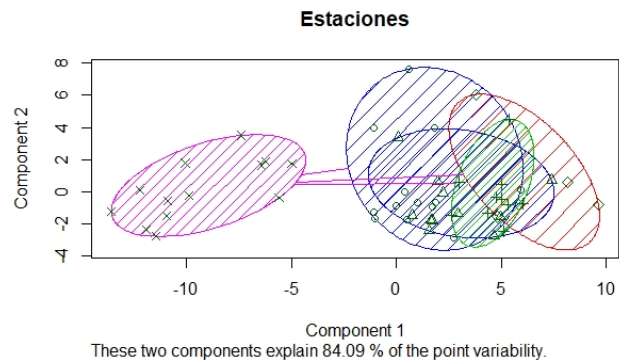


Fig. 3. Agrupación particional en los datos de las estaciones.

### E. Plot de gráficas para el análisis de la información

Se desplegó gráficas para su análisis como las precipitaciones observadas en todas las estaciones climatológicas a través de todo el año, también de la temperatura media observada en todas estas estaciones en el año, de igual manera para la temperatura máxima y mínima, mostradas en las Fig 4, 5, 6 y 7, respectivamente. En las figuras el eje x pertenece a los meses enumerado del 1 al 12 expresando el mes de enero hasta diciembre.

Precipitaciones por mes en las estaciones

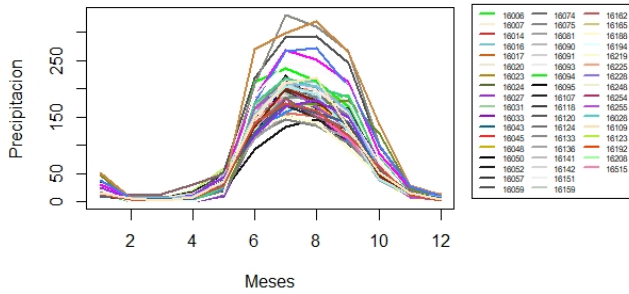


Fig. 4. Precipitaciones observadas en las estaciones climatológicas a través del año.

Temp min por mes en las estaciones

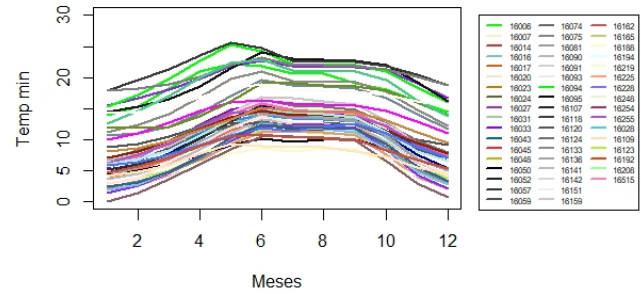


Fig. 7. Temperatura mínima observada en las estaciones climatológicas a través del año.

Temp observada por mes en las estaciones

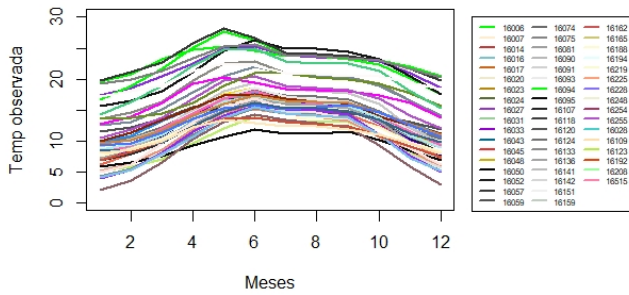


Fig. 5. Temperatura media observada en las estaciones climatológicas a través del año.

Precipitaciones por mes en los clusters

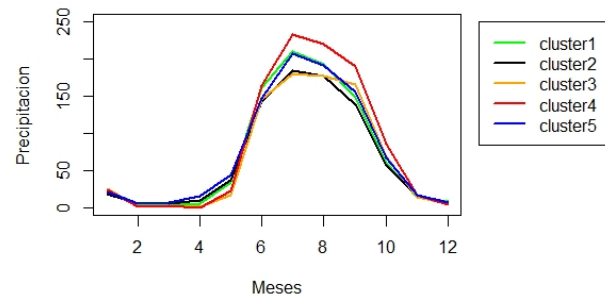


Fig. 8. Precipitaciones obtenidas por cada cluster en todo el año.

Temp max por mes en las estaciones

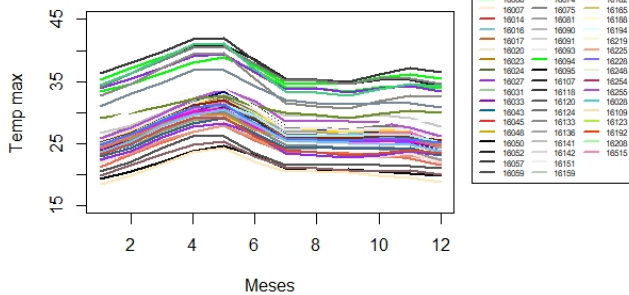


Fig. 6. Temperatura máxima observada en las estaciones climatológicas a través del año.

Temperatura máxima en los clusters

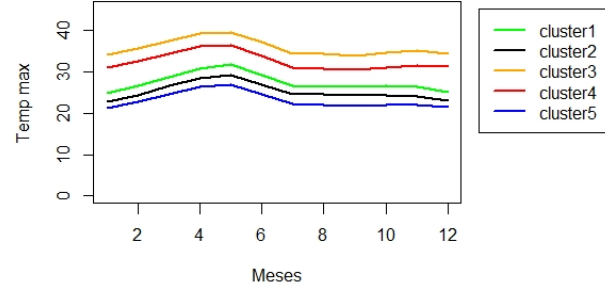


Fig. 9. Temperaturas máximas obtenidas por cada cluster en todo el año.

También se desplegó gráficas de los clusters obtenidos mostrados en la Fig 8, 9 y 10, siendo la gráfica de la precipitación de cada cluster versus los meses en lo que se dio tal cantidad de precipitación, la temperatura máxima obtenida por cada cluster en cada mes y la temperatura mínima obtenida por cada cluster en cada mes, respectivamente.

### III. RESULTADOS

De la Fig 4 se observó que las mayores precipitaciones son desde junio hasta octubre, teniendo un pico en julio y agosto, siendo la mayor precipitación en el mes de julio en

la estación San Carlos I (16107) ubicada en el municipio de Tequicheo De Nicolás Romero seguida en el mes de agosto en la estación Uruapan (CFE) (16165) ubicada en el municipio de Uruapan [1].

En la Fig 6 se observó que la temperatura máxima se registró entre los meses abril y mayo con una temperatura de 41 grados en la estación Los Pinzanes (CFE) (16075) ubicada en el municipio Huetamo.

De la Fig 7 se observó que la temperatura mínima se registró entre los meses de diciembre y enero con una

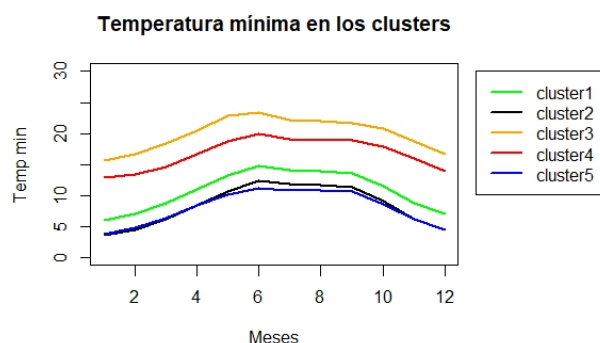


Fig. 10. Temperaturas mínimas obtenidas por cada cluster en todo el año.

temperatura de 0.72 y 0.16 grados respectivamente, en la estación Temascales (16124) ubicada en el municipio Contepec.

En Fig 8 se puede observar que el cluster 4 es quien posee las mayores precipitaciones en los meses de julio y agosto, mientras que los cluster 2 y 3 son los que obtienen las menores precipitaciones en estos meses. De la Fig 9 se observó que el cluster 3 es quien presenta las temperaturas más altas, mientras que el cluster 5 es quien presenta las temperaturas más bajas, este último expresado en Fig 10.

De los clusters formados, mostrados en la Fig 3, se observó que el primer cluster conformado por 19 estaciones está ubicado en la zona norte de Michoacán, distribuidas entre la subprovincia de Chapala, Zamora de Hidalgo y la Subprovincia Sierras y Bajíos Michoacanos, estas presentan un clima seco y semiseco con abundantes lluvias en verano y altas temperaturas en primavera. En este grupo hay precipitaciones mínimas de 2.4mm hasta máximas de 319.9mm y la media durante todo el año de las estaciones pertenecientes a este cluster es un aproximado de 72.9mm. Su temperatura media en este cluster fue de 13.8, la temperatura media máxima fue de 27.3 y la media mínima de 10.8 grados centígrados.

El segundo cluster conformado por 15 estaciones está ubicado en la zona noreste de Michoacán, distribuidas entre la Subprovincia Sierras y Bajíos Michoacanos, la Subprovincia Neovolcánica Tarasca y Heroica Zitácuaro, estas presentan un clima templado subhúmedo con lluvias en verano, a pesar de ser de las menos abundantes de todos los clusters junto al cluster 3, y temperaturas más altas en primavera, llegando a casi 30 grados centígrados. En este grupo hay precipitaciones mínimas de 2.9mm hasta máximas de 220.1mm y la media durante todo el año de las estaciones pertenecientes a este cluster es un aproximado de 67.1mm. Su temperatura media en este cluster fue de 11.6, la temperatura media máxima fue de 25.2 y la media mínima de 8.4 grados centígrados.

El tercer cluster conformado por 7 estaciones está ubicado

en la zona central de Michoacán y sureste, distribuidas entre las zonas de Apatzingán de la Constitución y la Subprovincia Depresión del balsas, estas presentan un clima seco y semiseco con abundantes lluvias en verano, pero a pesar de ser las menos abundantes a comparación de los otros clusters, estas se mantienen hasta setiembre en comparativa con los otros clusters que disminuyen sus precipitaciones en el mes de agosto. Este grupo es el que posee las mayores temperaturas en todo el año a comparación de los demás clusters, siendo su pico más alto en primavera donde fue cercano a los 40 grados centígrados. En este grupo hay precipitaciones mínimas de 0mm hasta máximas de 235.4mm y la media durante todo el año de las estaciones pertenecientes a este cluster es un aproximado de 67.6mm. Su temperatura media en este cluster fue de 22.1, la temperatura media máxima fue de 35.7 y la media mínima de 19.9 grados centígrados.

El cuarto cluster conformado por 6 estaciones está ubicado en la zona central, este y suroeste de Michoacán, distribuidas entre las zonas de la Subprovincia Cordillera Costera del Sur, Apatzingán de la Constitución y la Subprovincia de la Escarpa Limítrofe del Sur, estas presentan un clima cálido subhúmedo con abundantes lluvias en verano, siendo este grupo el que posee las más altas precipitaciones que todos los demás, y con altas temperaturas en primavera, este es el segundo cluster con más altas temperaturas. En este grupo hay precipitaciones mínimas de 0.1mm hasta máximas de 330.3mm y la media durante todo el año de las estaciones pertenecientes a este cluster es un aproximado de 81.1mm. Su temperatura media en este cluster fue de 19, la temperatura media máxima fue de 32.4 y la media mínima de 16.7 grados centígrados. Cluster con altas precipitaciones en verano, siendo las más altas de todos los demás clusters, y temperaturas más altas en todo el año, teniendo en primavera su pico más alto superando los 30 grados centígrados.

El quinto cluster conformado por 5 estaciones está ubicado en la zona norte y noreste de Michoacán, distribuidas entre las zonas de la Subprovincia Llanuras y Sierras de Querétaro e Hidalgo, Heroica Zitácuaro, Subprovincia Depresión del balsas, la Subprovincia Neovolcánica Tarasca y Subprovincia Sierras y Bajíos Michoacanos, en este cluster cada estación está muy distribuido geográficamente, estas presentan un clima templado subhúmedo con altas precipitaciones en verano y las más bajas temperaturas en primavera e invierno, siendo el cluster con temperaturas más bajas en todo el año con el pico más bajo en invierno llegando a casi 0 grados en algunas estaciones. En este grupo hay precipitaciones mínimas de 2.7mm hasta máximas de 293.3mm y la media durante todo el año de las estaciones pertenecientes a este cluster es un aproximado de 74.5mm. Su temperatura media en este cluster fue de 10.7, la temperatura media máxima fue de 23 y la media mínima de 8 grados centígrados.

#### IV. CÓDIGO

Este reporte y el código fuente (.R) se conseguirá en la carpeta asignada, en caso de que el código fuente esté dañado,

entrar al siguiente URL donde se encuentra el código en GitHub: [https://github.com/DiegoStevenVera/practica2\\_DM](https://github.com/DiegoStevenVera/practica2_DM).

## V. CONCLUSIONES

Las técnicas de agrupamiento nos ayudan a realizar un análisis más rápido cuando se posee una gran cantidad de elementos en un dataset. En este caso la gran cantidad de estaciones que se poseía se pudo agrupar en 5 clusters, de estos se pudo obtener las características más resaltantes de estas estaciones que les pertenecían, como:

- Cluster 1: Grupo de estaciones que pertenecían a un clima más balanceado, poseían temperaturas medias y precipitaciones medias a comparación de los otros clusters.
- Cluster 2: Cluster con menores precipitaciones y temperaturas bajas en todo el año.
- Cluster 3: Cluster con las precipitaciones más bajas pero con las más altas temperaturas en todo el año.
- Cluster 4: Cluster con las más altas precipitaciones que los demás y temperaturas altas a comparación de los demás.
- Cluster 5: Cluster con precipitaciones medias, ni tan altas ni bajas como el cluster 1, pero con las temperaturas más bajas de todos los demás grupos.

## REFERENCES

- [1] Gobierno de México, "Información climatológica por estado," <https://smn.conagua.gob.mx/es/informacion-climatologica-por-estado?estado=mich>.