

Laboratorio 1: Métodos clásicos

Fecha de entrega: ver Mediación Virtual

Entrega: individual

Instrucciones: Resuelva los ejercicios propuestos en el Laboratorio. Guarde el script respectivo, con su nombre y súbalo en el sitio del curso. Además, guarde los resultados y súbalos también en el sitio.

En este laboratorio trabajaremos varios enfoques clásicos de optimización que se usan en métodos de análisis de datos empleando .

1. **k-medias**

Escoja alguna librería de R que tenga el método de k-means con el método original de Forgy.

Considere las siguientes tablas de datos:

- Notas escolares (10 estudiantes \times 5 variables).
- Peces de Amiard.
- Iris de Fisher.

- (a) Para cada tabla haga un multistart de 1000 corridas con 3 y 4 clases.
- (b) Reporte el valor $W(P)$, inercia intraclases, de la mejor solución obtenida.
- (c) Reporte la tasa de atracción τ del mejor óptimo obtenido.
- (d) Reporte el valor promedio $\overline{W(P)}$ de cada experimento.
- (e) Reporte el tiempo de cada multistart.
- (f) Reporte el número de iteraciones que en promedio dura cada multistart.
- (g) Busque una manera de graficar el comportamiento de los óptimos locales para ilustrarlo y se vea la variedad que puede haber.

2. **Escalamiento multidimensional (MDS)** Aplicaremos el MDS en alguna librería de R a las siguientes tablas de disimilitudes:

- Las colas.
- Los teléfonos móviles de Alemania.
- Las ciudades de Costa Rica.

La tabla `mobile.txt` tiene 3 columnas: las primeras dos indican dos puntos i, j y la tercera columna es la disimilitud calculada por un equipo de ingenieros que determinan la calidad de la conexión entre i y j .

La tabla `colas.txt` contiene los juicios de 10 personas sobre la disimilitud (en una escala de 0 a 100) entre pares de refrescos tipo colas. Las etiquetas de los refrescos están en el archivo `colas.eti`

Para la tabla de las ciudades de Costa Rica puede construirla con Google Maps calculando distancias en kilómetros en carretera para las siguientes 12 ciudades:

Alajuela	Cartago	Golfito	Guápiles
Heredia	Liberia	Limón	Nicoya
San Carlos	San José	Pérez Zeledón	Puntarenas

Considere que se puede usar el ferry para cruzar a la Península de Nicoya.

Se quiere minimizar el 1-stress de Kruskal:

$$\sigma^2(\mathbf{X}) = \sum_{i,j} w_{ij} [\delta_{ij} - d(\mathbf{x}_i, \mathbf{x}_j)]^2$$

con δ las disimilitudes dadas y d la distancia Euclídea en 2 dimensiones.

Para las inicializaciones puede tomar un rectángulo definido por las disimilitudes máximas de cada tabla y hacer tirajes al azar de las coordenadas horizontal y vertical de cada punto. Como referencia, haga una corrida iniciando con la solución clásica de Torgerson.

- En cada caso, aplique el MDS métrico de Kruskal en un multistart de 100 corridas con inicializaciones al azar y soluciones en \mathbb{R}^2 .
- Reporte el valor $\sigma(\mathbf{X})$, el 1-stress de Kruskal, de la mejor solución obtenida.
- Reporte la tasa de atracción τ del mejor óptimo obtenido.
- Reporte el valor promedio $\overline{\sigma(\mathbf{X})}$ de cada experimento.
- Reporte el tiempo de cada multistart.
- Reporte el número de iteraciones que en promedio dura cada multistart.
- Busque una manera de graficar el comportamiento de los óptimos locales para ilustrarlo y se vea la variedad que puede haber.