

Master of Science in Data Science

Métodos de aprendizaje de máquinas en Data Science

Tarea 1

Clusters espaciales

El territorio de Chile está subdividido en unidades administrativas, como regiones, provincias y comunas. La definición de estas unidades actuales obedece a diversos motivos históricos, políticos y circunstanciales, por lo que no existe homogeneidad entre ellas y existen comunas de diversos tamaños, poblaciones y condiciones de vida. La Subsecretaría de Desarrollo Regional y Administrativo de Chile ha decidido que para optimizar el diseño y ejecución de políticas públicas necesitan crear unidades administrativas que tengan un mayor grado de homogeneidad interna, también conocido como cohesión, y que estas regiones estén cohesionadas también en el espacio.

Para estos efectos, la SUBDERE lo ha contratado a usted como consultor para que diseñe un algoritmo de clusterización que permita crear regiones en el territorio que cumplan con estas características. Para cumplir esto, se le otorgará una base de datos de todas las manzanas de la comuna de Las Condes (~1.600), donde cada manzana contiene 39 atributos que reflejan diferentes dimensiones demográficas y de condiciones de vida, según lo reportado en el censo 2017 y lo calculado por la plataforma de [Bienestar Humano Territorial](#).

El objetivo principal de este encargo es crear un programa computacional que permita crear clústers de manzanas que sean similares entre sí para reemplazar a los 13 cuadrantes actuales.

Problema 1) Aplique un algoritmo de clustering y seleccione 13 clusters sin considerar las variables de posición de las manzanas. Muestre el resultado del proceso de clustering (grafique los puntos del espacio con sus respectivas etiquetas de clusters), y vea cuáles son sus ventajas y desventajas para la creación de las manzanas.

Problema 2) Aplique un algoritmo de clustering y seleccione 13 clusters considerando solo las variables de posición de las manzanas. Muestre el resultado del proceso de clustering (grafique los puntos del espacio con sus respectivas etiquetas de clusters), y vea cuáles son sus ventajas y desventajas para la creación de las manzanas.

Problema 3) Aplique un algoritmo de clustering y seleccione 13 clusters considerando las variables descriptivas y las variables de posición. Sin embargo, para este problema, se le pide que pueda integrar la distancia entre las manzanas como algo restrictivo. Específicamente, que la distancia entre dos puntos que pertenecen a un mismo cluster no sea mayor a 2,000 metros (en la medida de lo posible). Muestre el resultado del proceso de clustering (grafique los puntos del espacio con sus respectivas etiquetas de clusters), y vea cuáles son sus ventajas y desventajas para la creación de las manzanas.

Elija uno de los tres clustering obtenidos (justificando su decisión) y describa los clústeres generados.

Los puntajes asignados a cada tarea corresponden a:

- Limpieza de datos y selección (1 punto, código y justificación)
- Problema 1 (1 punto, código y justificación)
- Problema 2 (1 punto, código y justificación)
- Problema 3 (2 puntos, código y justificación)
- Análisis de los resultados (1 puntos, código y justificación).

Los datos se pueden descargar desde [aquí](#).

La fecha de entrega es el 12 de Septiembre a las 15:30 horas.

La tarea se puede realizar hasta en grupos de 3 personas.

¡Mucha suerte!