

## Tarea 5

### Aglomeración

#### I Descripción

Visite el sitio de scikit-learn:

`plot_kmeans_digits`

Allí usted encontrará un ejemplo de aglomeración sobre los datos de imágenes de  $8 \times 8$  px de dígitos, que además evalúa al algoritmo de  $k$ -medias con varios criterios.

1. Descargue, ejecute, y analice el ejemplo.
2. Investigue qué significan los criterios de evaluación.
3. Describa con sus palabras, brevemente qué significa cada criterio de evaluación.
4. Investigue qué hace cada modo de inicialización.
5. Describa con sus palabras, brevemente qué hace cada modo de inicialización.
6. Elimine la normalización de los datos, para que estos puedan mantener su naturaleza como imágenes.
7. Agregue código al ejemplo para visualizar los centroides como imágenes.
8. Muestre los centroides utilizando  $k = 3$ ,  $k = 10$  y  $k = 20$ .
9. En realidad el ejemplo así como está, hace “trampa” con ilustrar los centroides de los conglomerados, pues no está mostrando la *proyección* de los centroides en 64 dimensiones hacia el espacio de los dos primeros componentes principales, sino que está realizando la aglomeración directamente con los datos ya proyectados en 2D.

Agregue a la figura anterior los centroides encontrados en 64D, reproyectados a 2D.

#### II Entregables

Un archivo comprimido (`.zip` o `.tar.gz` o `.tar.bz2`) con:

1. Archivos de Python modificado.
2. Archivo PDF con descripciones de criterios de evaluación y modos de inicialización.
3. Archivo PDF con las figuras de los centroides.

Esta tarea es en parejas.