

# Minería de datos: Tarea 5

Profesor Guillermo Herrera

---

## Agrupamiento jerárquico

**Ejercicio 1.** En clase se hizo la deducción del método de agrupamiento jerárquico. Describe los 4 tipos de enlace (linkage) y su respectivo parámetro en sklearn para utilizarlo en Python.

**Ejercicio 2.** Grafica 5 datos en  $\mathbb{R}^2$  e ilustra manualmente, paso a paso el método de agrupamiento jerárquico y cuáles serían los dendrogramas asociados para esos 5 datos utilizando cada uno de los 4 tipos de enlace.

---

## k-means

**Ejercicio 3.** Implementa el algoritmo de k-means para hacer agrupamiento. Debes utilizar la función de distancia euclidiana al cuadrado como medida de similitud.

**Ejercicio 4.** En clase vimos el uso de k-means para segmentar imágenes por colores. Con la implementación del ejercicio 3 vas a segmentar alguna de las imagenes anexas utilizando 3 agrupamientos, cada agrupamiento debe ilustrarse de un color diferente como se hizo en clase.

**Ejercicio 5.** Una manera de convertir una imagen RGB a escala de grises, es poner en todas las entradas de cada pixel el mismo valor y calcular este valor a partir la fórmula

$$N = (0.3 \cdot R) + (0.59 \cdot G) + (0.11 \cdot B).$$

Así por ejemplo para el pixel que tiene 3 entradas con valores  $[R, G, B] = [0.1, 0.3, 0.2]$ , entonces con la fórmula dada cambiará los valores a  $[N, N, N] = [0.229, 0.229, 0.229]$ . El ejercicio es generar una nueva imagen a partir de la segmentación. El objeto principal (la flor) debe estar con sus colores originales y el resto de la imagen debe estar en escala de grises. Ustedes deben decidir el valor de  $k$  a partir del cual ya les genera la imagen deseada. Tomen en cuenta que puede tardar unos minutos en ejecutarse su programa.

---

## Reporte

Finalmente deben escribir un reporte con todos los ejercicios respectivos de la tarea. El reporte debe cumplir con las especificaciones que hemos platicado en clase y el código debe estar comentado. No deben hablar del código en el reporte ni poner capturas de pantallas del mismo.

En la primera sección deben hablar de los ejercicios 1, 2. **(4 puntos)**

En la segunda sección deben hablar de los ejercicios 3, 4 y 5. Deben agregar en su reporte la imagen que eligieron y las segmentaciones. **(6 puntos)**