

PRÁCTICA 2 - GEOMETRÍA COMPUTACIONAL - 2023

15 de febrero, 22 de febrero y 1 de marzo de 2023

DIAGRAMA DE VORONÓI Y CLUSTERING

De uso obligatorio:

- En “Personas_en_la_facultad_matematicas.txt” se presentan dos variables de estado, $X_1 =$ “nivel de estrés” y $X_2 =$ “afición al rock”, para un conjunto de $A = \{a_i, \{X_j\}_{j=1}^2\}_{i=1}^{1500}$ de 1500 personas de la Facultad de Matemáticas. Estos datos deben utilizarse como valores de entrada del problema.
- En “Grados_en_la_facultad_matematicas.txt” se presenta el listado de pertenencia de cada persona a uno de los 4 grados o doble-grados principales de la Facultad de Matemáticas, etiquetados como $\{0, 1, 2, 3\}$ para preservar el anonimato de cada caso. Estos datos sólo deben utilizarse para medir el error de predicción, en caso de solicitarse.

Determina el número ideal de *Grados* de nuestra Facultad de Matemáticas (sistema A) a partir del número óptimo de *clusters* o vecindades de Voronói. Para ello, utiliza el *coeficiente de Silhouette* (\bar{s}), que **puede emplearse directamente desde la librería sklearn**:

i) Obtén el coeficiente \bar{s} de A para diferente número de vecindades $k \in \{2, 3, \dots, 15\}$ usando el algoritmo KMeans. Muestra en una gráfica el valor de \bar{s} **en función de k** y **decide** con ello cuál es el número óptimo de vecindades. En una segunda gráfica, muestra la clasificación (clusters) resultante con diferentes colores y representa el **diagrama de Voronói** en esa misma gráfica.

ii) Obtén el coeficiente \bar{s} para el mismo sistema A usando ahora el algoritmo DBSCAN con la métrica ‘euclidean’ y luego con ‘manhattan’. En este caso, el parámetro que debemos explorar es el **umbral de distancia** $\epsilon \in (0.1, 0.4)$, fijando el número de elementos mínimo en $n_0 = 10$. Comparad gráficamente con el resultado del apartado anterior.

iii) ¿De qué Grado diríamos que son las personas con coordenadas $a := (0, 0)$ y $b := (0, -1)$? Comprueba tu respuesta con la función `kmeans.predict`.

De uso voluntario, únicamente quien lo necesite:

- La plantilla 1 aplica un ejemplo de KMeans (sklearn), con *métrica euclidiana* (L^2).
- La plantilla 2 aplica un ejemplo de DBSCAN (sklearn), con *métrica de Manhattan* (L^1).

Observaciones:

La memoria debe entregarse antes de que transcurran **21 días** desde el inicio de la práctica, salvo que se indique lo contrario.

La memoria, siempre en **pdf**, debe incluir **al menos** la siguiente información: (1) Introducción (motivación/objetivo de la práctica), (2) Material usado (método y datos), (3) Resultados, (4) Conclusión y (5) Anexo con el script/código utilizado.

La extensión máxima de la memoria **no superará las 2 páginas**, sin contar el código anexado (ilimitado). El total de la superficie de las figuras/tablas (si las hubiese) no podrán excederse del 50 % de la memoria (cuerpo principal).

Además, el **código fuente debe incluirse** en un archivo ‘.py’ independiente. **No** subir archivos **comprimidos**.