

PRÁCTICA 3 - GEOMETRÍA COMPUTACIONAL - 2022

23 de febrero y 2 de marzo de 2022

DIAGRAMA DE VORONÓI Y CLUSTERING

De uso obligatorio:

- En la plantilla 1 se muestra un ejemplo de aplicación del algoritmo KMeans (librería `sklearn`), utilizando una **métrica euclidiana**.
- En la plantilla 2 se muestra un ejemplo de aplicación del algoritmo DBSCAN (librería `sklearn`), utilizando una *métrica euclidiana* (L^2) o bien una *métrica de Manhattan* (L^1).

El objetivo de esta práctica es clasificar un sistema X de 1000 elementos (con dos estados cada uno) a partir de un determinado número de *vecindades de Voronói*. Para determinar el número óptimo de vecindades o *clusters*, la medida que emplearemos es el *coeficiente de Silhouette* (\bar{s}), que **puede utilizarse directamente desde la librería `sklearn`**.

i) Obtén el coeficiente \bar{s} de X para diferente número de vecindades $k \in \{2, 3, \dots, 15\}$ utilizando 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 Vornói** en esa misma gráfica.

ii) Obtén el coeficiente \bar{s} para el mismo sistema X 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) ¿A qué vecindad pertenecen los elementos con coordenadas $a := (0, 0)$ y $b := (0, -1)$? Comprueba tu respuesta con la función `kmeans.predict`.

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.

El **código fuente debe incluirse** como archivo ‘.py’ independiente. **No** subir archivos comprimidos.