

Taller 3

Estructura Unión-Búsqueda

Clusterización por el vecino más cercano

Los métodos de clusterización en analítica de datos buscan agrupar conjuntos de puntos “similares”. Para efectos del ejercicio, se tiene como entrada un conjunto de puntos en el plano en forma de un arreglo de dimensión $N \times 2$.

El agrupamiento por vecinos cercanos se implementa de la siguiente forma:

1. Cada punto del dataset es inicialmente un cluster.
2. Se busca la pareja de puntos más cercanos que no se encuentren “conectados” y se “unen”.
3. Se repite el paso 2 hasta no encontrar pares de puntos con una distancia menor a $DMAX$ (parámetro del algoritmo especificado por el cliente).

El programa debe arrojar las siguientes salidas:

Un arreglo $N \times 1$ indicando el cluster al que pertenece cada punto.

El número de clusters encontrados.

Ejercicios a desarrollar

1. Implementar el algoritmo de clusterización por vecino más cercano como una función de biblioteca. Para generar los puntos 2D de entrada, se puede utilizar el [gist](#) indicado más abajo.
2. Graficar los puntos del cluster asignando distintos colores a los puntos de cada uno de los clusters.
3. Estimar analíticamente el desempeño del algoritmo en función de N .
4. Evaluar empíricamente el desempeño del algoritmo. Incluir tabla de mediciones en función de N , gráfica y curva de mejor ajuste. Comparar el resultado con el valor obtenido analíticamente.

Generación de clusters

<https://gist.github.com/jmlon/856375ca62722f361efa8bc75062c98b>

Entregables

Remitir el código fuente de la solución implementada y la hoja de cálculo con los resultados de las pruebas experimentales (se aceptan Word, Excel, OpenOffice, LibreOffice, PDF). Nombrar el archivo comprimido Taller3-<Nombre1>-<Nombre2>... (.zip .rar .7z o .tgz). Para estandarizar la forma de invocar el programa, ubicar el método main y las funciones de biblioteca solicitadas en la clase Taller3.

En caso de utilizar estructuras de las bibliotecas del texto (algs4.jar) **no** anexar la biblioteca.

Grupos máximo de 3 personas.

Ejemplo:

