

Separando la paja del trigo

En *aprendizaje por computadora* uno de los problemas más fundamentales es el de *clusterizar*. Clusterizar significa agrupar un conjunto de objetos en diferentes subconjuntos por alguna noción de similaridad intraconjunto (o disimilaridad interconjunto). Por ejemplo, si tenemos una nube de puntos en un espacio métrico, nos gustaría realizar divisiones (*clusters*) que agrupen a los conjuntos de puntos que se encuentran juntos.

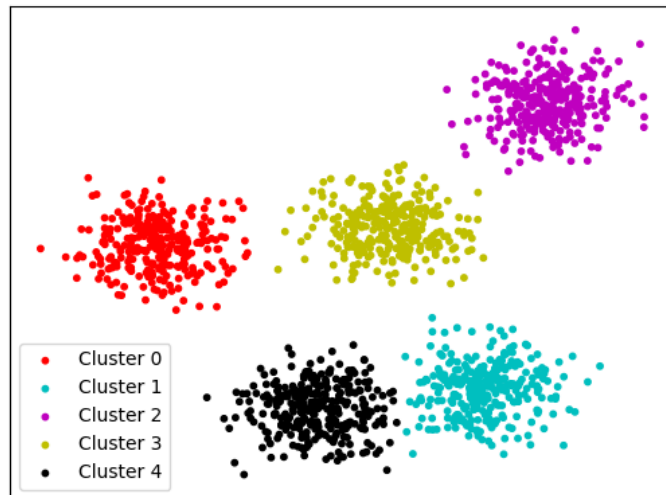


Figura 1: Puntos clusterizados

En la figura 1 vemos un ejemplo de este problema. En esta figura los clusters son evidentes a simple vista, aunque existen algunos puntos que no son tan fáciles de clasificar.

En otros casos los clusters pueden parecer claros intuitivamente pero difíciles de separar de una manera algorítmica. El problema de clustering resulta de gran interés ya que son muchas las aplicaciones en donde se puede utilizar. Puntos como los de la figura 1 pueden representar situaciones muy variadas como dígitos manuscritos que queremos identificar o pacientes que se sometieron a un determinado análisis médico.

En nuestro caso, buscaremos abordar este problema de una perspectiva de grafos, utilizando la noción de Árbol Generador Mínimo para generar los clusters deseados.

Parámetros y formato de entrada/salida

La entrada consistirá de una primera línea con un entero n , correspondiente a la cantidad de puntos a clusterizar. Luego le sucederán n líneas con dos números reales, x_i , y_i . x_i y y_i indicarán las coordenadas del punto i .

La salida consistirá de n líneas. Cada línea i tendrá un número que debe indicar a qué cluster pertenece el punto i .