

Tema 11. Clustering via Algoritmo k -medias. Ejercicios

Abdelmalik Moujahid, Iñaki Inza, Pedro Larrañaga
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco–Euskal Herriko Unibertsitatea

1. La tabla adjunta contiene 8 casos bidimensionales a partir de los cuales se trata de aplicar tanto el algoritmo propuesto por Forgy como el propuesto por McQueen, en ambos casos con un valor de $k = 3$.

<i>Caso</i>	X_1	X_2
1	1	1
2	2	4
3	3	2
4	3	5
5	4	4
6	4	7
7	6	4
8	6	6

2. Es evidente que el proceso de búsqueda en el que se basa el algoritmo de Forgy es, desde un punto de vista de optimización, local.

Se trata de que diseñes un algoritmo genético para evitar las características de optimizador local del algoritmo de Forgy. Especifica la representación usada, los operadores de cruce y mutación así como la función objetivo.

3. El algoritmo de McQueen se basa en un orden total de los casos almacenados en el fichero, el cual ha sido determinado previamente. Es obvio por tanto que el algoritmo de McQueen es dependiente del orden de almacenamiento de los casos.

Se trata de especificar un algoritmo genético para la búsqueda del mejor orden de almacenamiento de los casos. Estudiar la representación adecuada, así como los operadores de cruce y mutación y la función de evaluación.