

Proiectarea Algoritmilor 2011-2012

Laborator 12 - Aplicații de laborator

Algoritmi aleatori

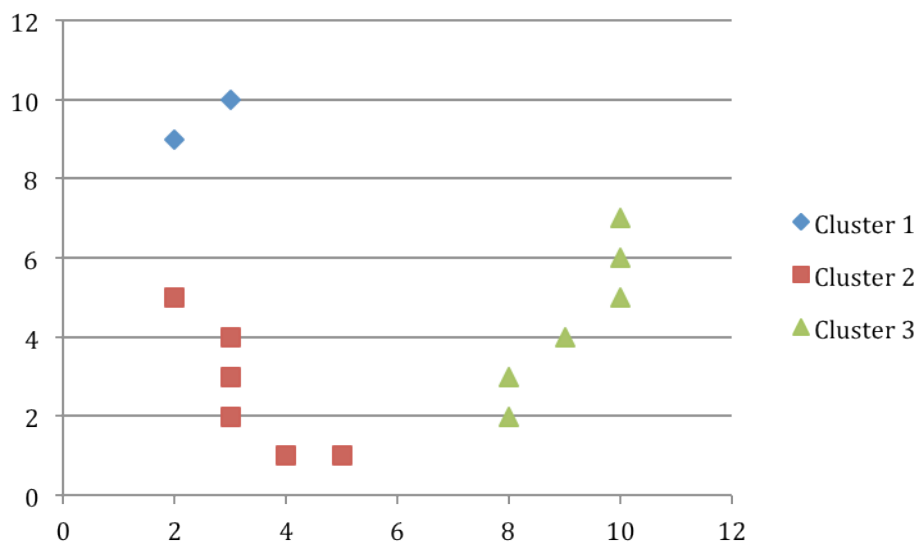
k-Means ++

Fie n puncte într-un spațiu bi-dimensional. Se dorește o grupare a acestora în k **cluster**e - un grup de puncte situate într-o vecinătate spațială care să maximizeze coeziunea intra-cluster și să asigure o cuplare slabă inter-cluster.

Spre exemplu, pentru setul de intrare:

X	2	3	2	3	3	3	4	5	8	8	9	10	10	10
Y	9	10	5	4	3	2	1	1	2	3	4	5	6	7

Se poate observa o grupare naturală în 3 **cluster**e:



Coeziunea internă este determinată drept $1 /$ media distanțelor către clustroid (punctul din cluster cel mai apropiat de toate celelalte noduri), iar cuplarea inter-cluster este $1 /$ distanța dintre **cluster**e (distanța dintre clustroidii aferenți).

Pașii algoritmului k-Means sunt următorii (pentru N și k date):

1. Se selectează k puncte random din spațiu care vor fi centroizii inițiali ai **cluster**elor.
2. Se efectuează iterativ următorii pași cât timp atribuirile fiecărui nod la un **cluster** rămân neschimbate:

- a. **Asignarea** fiecărui nod unui cluster (distanța minimă euclidiană către centrozii din pasul curent este minimă);
- b. **Recalcularea centroidului** drept media aritmetică a coordonatelor punctelor asignate.

Pentru un set de date și un k stabilit se vor determina clusterelor aferente.

Se va implementa și o optimizare a selecției inițiale de puncte pornind de la principiul că acestea ar trebui să fie geografic cât mai dispersate (se dorește maximizarea distanței între centrozii inițiali).