|  |
| --- |
| 16 octobre 2016  Caoi VALENTE  Rochd MALIKI |

|  |
| --- |
| **Compte rendu TP2**  MAJ INFO – UV2  413 : Algorithmique avancé |

1, rue de l’Institut

75000 PARIS – France

Tél. +33 (0)1 00 00 00 00

Siret : 000 000 000 000 00

APE : 0000 A

[**www.mines-telecom.fr**](http://www.mines-telecom.fr)

Mines ParisTech

Mines Albi

Mines Alès

Mines Douai

Mines Nantes

Mines Saint Étienne

Télécom ParisTech

Télécom Bretagne

Télécom SudParis

Télécom Ecole de Management

Télécom Lille1

Eurecom

Sommaire

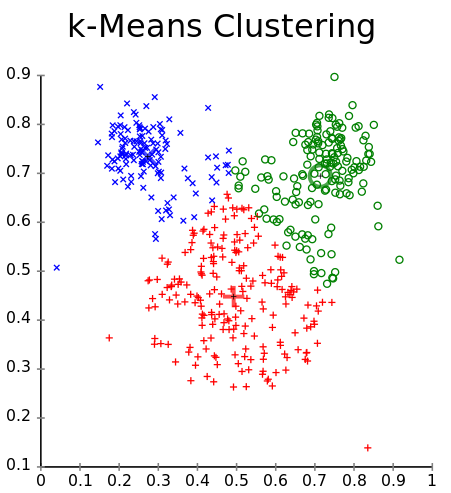
Table des matières

[1. Présentation de l’algorithme k-means 2](#_Toc464001177)

[2. structure du code 2](#_Toc464001178)

# 

# Présentation de l’algorithme k-means



L’algorithme k-means permet de construire une collection d’objets similaires au sein d’un même groupe. Autrement dit, l’algorithme permet de classifier les objets tels que les objets de grande similarité sont dans un même groupe, et les objets de groupes différents sont de faible similarité.

# structure du code

Notre algorithme doit être capable de :

* Générer aléatoirement les données
* Choisir des centres aléatoires à partir des données générées
* Chargement et sauvegarde des données
* Calcul de la distance entre deux objets
* Affectation des objets aux différents groupes
* Mettre à jour les centres à chaque fois afin de bien regrouper les objets

# Fonctions et complexités

## générer des données aléatoires :

Fonction:

On fournit le nombre d’observations ainsi que le nombre d’attributs qui caractérise ces observations, ces attributs sont générés aléatoirement dans un intervalle qu’on peut fournir en entrées, cet intervalle est mis pas défaut [-1000,1000] :

On représente chaque données par un tableau de taille égale au nombre d’attributs +1 sur lequel l’indice 0 représente le numéro de la donnée.

Pseudo-code :

generateDatas(n,k, min\_=-1000, max\_=1000){

tableau\_de\_données = [ ] **1**

for j in range(n) { **n\*(2+k\*1)**

donnée = [ 0 for i in range(k+1) ] 1

donnée [ 0 ] = j+1 1

for h in range(1,k+1) { k

donnée[h]=uniform( min\_,max\_) 1

}

tableau\_de\_données.append(donnée) **1**

}

return tableau\_de\_données

}

Complexité :

La complexité de la fonction est de 1+n\*(2+k) donc c’est une complexité de O(n\*k)

## Choisir centres :

Cette fonction permettra de choisir un nombre k de centres parmi les données générées

|  |  |
| --- | --- |
|  | Technopôle Brest-Iroise CS 83818 29238 Brest Cedex 3 France  +33 (0)2 29 00 11 11  **www.telecom-bretagne.eu** |
| INSTITUT-MINES-TELECOM_Logo_Buro_petit | |