Rapport du Projet

Apprentissage non supervise

1. Problematique

Ce projet est pour apprendre la methode de programmation dynamique de Fisher.

Jeu de donnees: sequenceimu, aiquillage

Methode utilise pour la comparaison: CAH-Ward, K-means

2. Explication de l’algorithme

3. Experiment

3.1. Implementation de l’algorithme de programmation dynamique de Fisher

- Fonction diam(X, a, b): retourner le diametre de X[a:b]

- Fonction diam\_matrice(X): cette operation prend bcp de temp ==> pour diminuer le temps de methode du coude => le fonction clustfisher va utiliser parametre diam\_matrice au lieu de donnees X. Retourner la matrice des diametres des classes

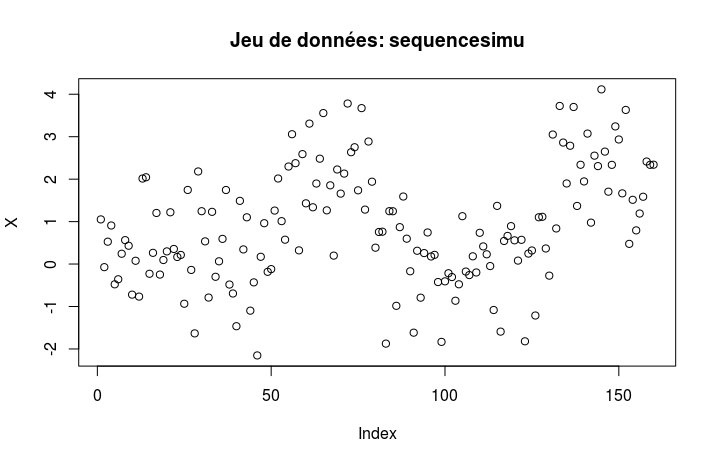
- Fonction clustfisher(diam\_matrice, K): retourner des labels, des instants de changement, l’inertie totale, et l’inertie totale intra-classes

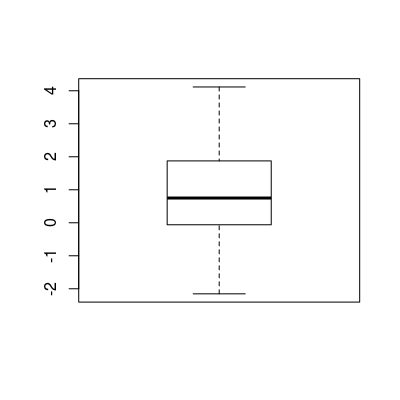
- Fonction clustering(X): il y a 2 jeux de donnees a analyser => pour eviter redoubler du code, cette fonction utilise le jeu de donnees X pour executer methode du coude (base sur Fisher algo) => choisir K manuellement avec l’aide de graphe inerties intra-classes totales, puis il va executer l’algorithme Fisher, CAH-Ward, et K-means avec nombre de classes K choisi. Retourner des objets clustering Fisher, CAH-Ward, K-means, et le numero K

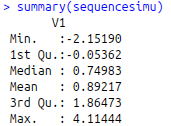
3.2. Resultat

3.2.1. Application a des donnees simulees: sequencesimu

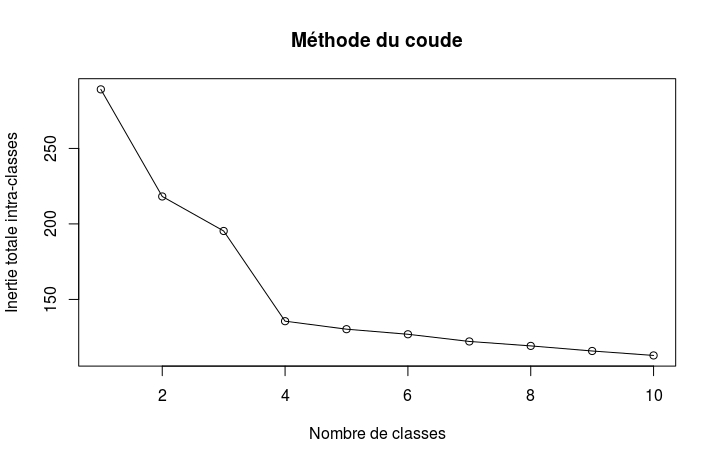
a. Analyse





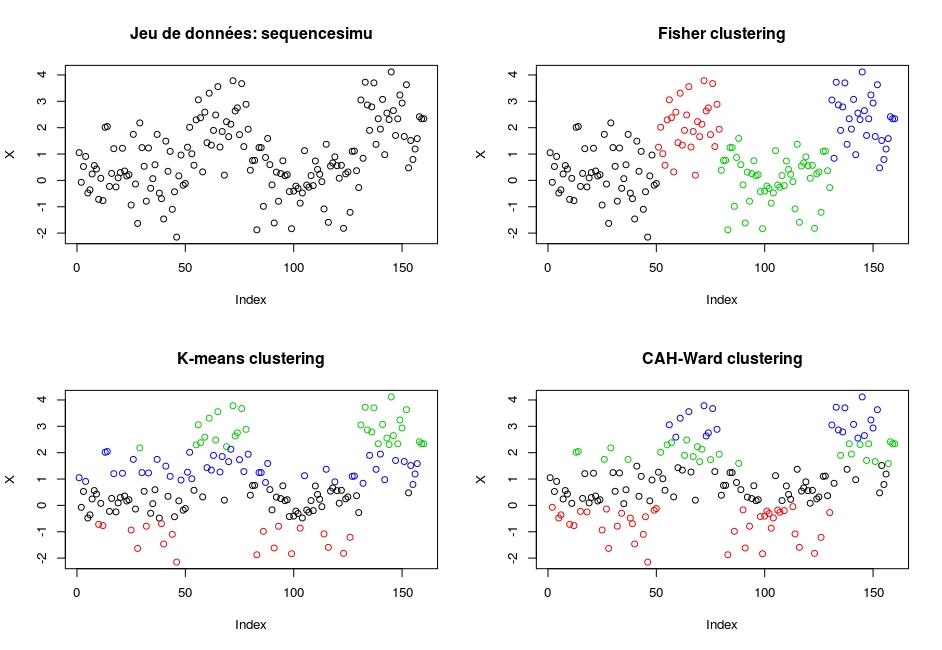


b. Clustering avec des algorithmes: Fisher, CAH-Ward, K-means

Executer clustering => graph suivant

Avec l’aide de ce graphe => choisir manuellement K = 4

Obtenir resultat suivant:

=> Resultat obtenu par Fisher est totalement different

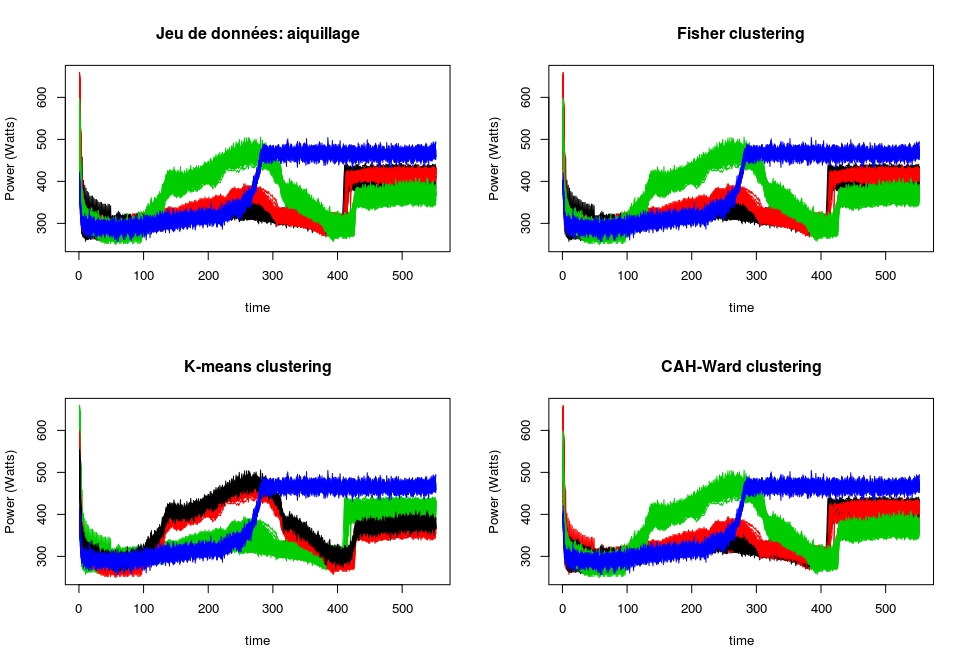
pourquoi?

3.2.2. Application a des donnees reelles: aiquillage

a. Analyse

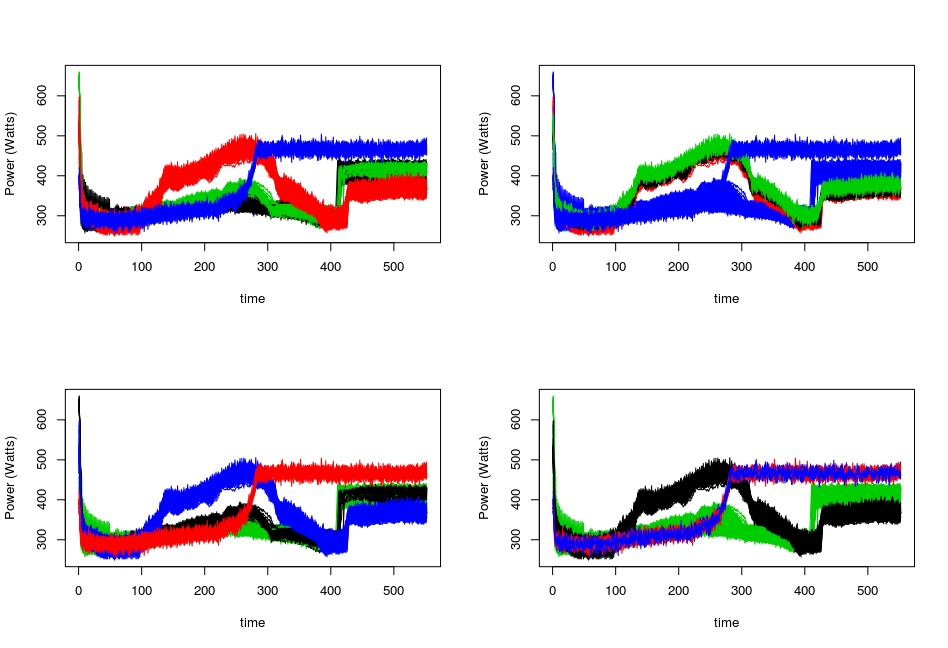
Le jeu de donnees Aiguillage comprend 140 series temporelles d´ecrites par 553 variables : les 552 premi`eres variables correpondent `a l’´energie consomm´ee au cours des mouvements m´ecaniques d’un syst`eme d’aiguillage ferroviaire et la 553e variable correspond `a la classe (1 : sans d´efaut, 2 : d´efaut mineur, 3 : d´efaut critique, 4 : panne).

b. Segmentation en 4 classes



Resultat obtenu par K-means est different (a cause de l’iniatilisation des centres des classes)

Refaire la segmentation avec K-means plusieur fois:



4. Auto-critique: ce qui a ete fait, les defauts, ce qui n’a pas ete fait, ainsi que les ameliorations possibles