

Programme de remise à niveau sur le thème « Data Mining – Clustering » (classification automatique).

1. Supports à lire

Les supports suivants présentent les bases de la classification automatique (on parle aussi de typologie, ou d'apprentissage non-supervisé). Deux méthodes sont souvent abordées dans ces cours introductifs : la classification ascendante hiérarchique et les techniques de partitionnement par la méthode des centres mobiles. Nous les explorerons longuement dans les exercices de ce thème.

- Chavent M., « La classification automatique de données quantitatives » ;
http://www.math.u-bordeaux.fr/~machaven/wordpress/wp-content/uploads/2013/10/cours_classif_quant.pdf
- Gonzalez P.L., « Méthodes de classification » ;
<http://maths.cnam.fr/IMG/pdf/Classification-2008-2.pdf>
- Gettler-Summa M., Pardoux C., « La classification automatique » ;
https://www.ceremade.dauphine.fr/~touati/SODAS/A_VERIFIER/SEMINAIRES/EDOGEST-seminaires/Classification.pdf
- R. Rakotomalala
<http://tutoriels-data-mining.blogspot.com/2016/07/classification-ascendante-hierarchique.html>
<http://tutoriels-data-mining.blogspot.com/2016/10/clustering-methode-des-centres-mobiles.html>
<http://tutoriels-data-mining.blogspot.com/2016/09/clustering-caracterisation-des-classes.html>
- Reboul L., « Chapitre 3 : Classification » ;
<http://iml.univ-mrs.fr/~reboul/ADD4-MAB.pdf>

2. Exercices à réaliser

Il est illusoire de vouloir tout comprendre avant de vous lancer dans les exercices sur machine. Après avoir lu rapidement les supports, essayez de commencer. Vous aurez alors le loisir de revenir plus spécifiquement sur les différents thèmes quand vous chercherez à répondre précisément aux questions.

Autre conseil important, concentrez-vous sur les aspects opérationnels. Il sera toujours temps par la suite de comprendre les principes fondateurs. A chaque problème à traiter doit correspondre un ou plusieurs fichiers à retourner (Excel ou R). Commentez autant que possible votre travail.

Problèmes à traiter :

1. SISE – DM 2 – Classification ascendante hiérarchique (Autos 2005 – Non 4x4)
2. SISE – DM 2 – K-Means (Image Segmentation)
3. SISE – DM 2 – Etude de cas (Zoo dataset)