Master M2, SITN et Data Science-maths, Université Claude Bernard, Lyon1
Analyse factorielle,
année 2016-2017

PROJET(individuel ou par binôme)
à rendre au plus tard le 17 Février 2017

Les problèmes de ce projet utilisent les méthodes: ACP, AFD, AFCM, ACC.
***Pour tous les problèmes**, il faut d'abord faire une **description de l'expérience** qui a permis d'obtenir les données. Ensuite, il faut réaliser une **étude descriptive (avec interprétations)**.*

Tous les quatre problèmes utilisent des données du site "Center for Machine Learning and Intelligent Systems at the University of California":

http://cml.ics.uci.edu/

de l'Université de Californie, Etats Unis.
Pour chaque problème, vous trouvez le fichier de données à la rubrique "Data Folder" et la description des données à la rubrique "Data Set description".

Problème 1. *(Primar Tumor)*
A l'adresse internet

http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass

vous trouvez des données concernant des résultats obtenus à la suite des mammographies.
On veut étudier si les variables qualitatives sont corrélées, et dans le cas affirmatif, trouvez les liens entre les modalités. Une attention particulière sera accordée à la sévérité du cancer. Le médecin voudrait savoir si la gravité du cancer est décrite par les caractéristiques obtenues à la suite de la biopsie.

Problème 2. *(Stone Flakes)*
A l'adresse internet

http://archive.ics.uci.edu/ml/datasets/StoneFlakes

vous trouvez des données concernant des fouilles archéologiques.
Une partie des données se trouve à l'adresse mentionnée plus haut et l'autre partie à l'adresse (il faut cliquer sur "Data Folder")

http://archive.ics.uci.edu/ml/machine-learning-databases/00299/StoneFlakes.dat

Les données manquantes sont symbolisées par "?". Pour exploiter les données il faut d'abord les mettre dans le même tableau (éventuellement le même fichier de données).
1) Une fois que vous avez réussi à mettre les données sous une forme exploitable, répondez à la question mentionnée dans le texte: *"Does the data reflect the technological progress during several hundred thousand years? "*
2) Il y a-t-il une relation entre les variables: *LBI, RTI, WDI, FLA, PSF, FSF, ZDF1, PROZD*? En utilisant les relations trouvées, caractérisez chaque variable *group, age, dating, mat, region, site* des éclats (flakes) de pierre obtenus à la suite de la taille des pierres.

Problème 3. *(Parkinson)*
A l'adresse internet

http://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring

vous trouvez des données concernant la maladie de Parkinson.
Etudiez le lien entre les variables *"motor UPDRS"* et *"total UPDRS"* d'une côté et les autres variables numériques de l'autre côté.

Problème 4. *(Wine)*
A l'adresse internet

http://archive.ics.uci.edu/ml/datasets/Wine

vous trouvez des données concernant les résultats d'une analyse chimique des vins produits dans la même région en Italie, mais par trois vignerons différents.
1) Il y a-t-il une relation entre les variables qui mesurent les caractéristiques chimiques des vins? Quelles sont ces relations?
2) De point de vue des caractéristiques chimiques, les vins des trois vignerons sont-ils différents? Pourquoi?
3) Caractérisez les vins produits par chacun des trois vignerons.