

Exercice 1

ACP de verres (part 1)

6 points

Le but de cet exercice est d'analyser globalement le tableau de données D qui contient des mesures réalisées sur différents échantillons de verre. Le recueil de ces données sur différents types de verre a été motivée par des recherche en criminologie. Sur les lieux d'un crime, le verre laissé peut être utilisé comme preuve...s'il est correctement identifié ! Les données sont disponibles dans le fichier `median_m8_2018.mat`, des données du médian 2018 dans l'onglet Annales du Moodle de M8. Voici une description détaillée des données. Rappelons que *Instances* = individus et *Attribute* = variables.

```

1. Title: Glass Identification Database

2. Relevant Information:n
The study of classification of types of glass was motivated by
criminological investigation. At the scene of the crime, the glass left
can be used as evidence...if it is correctly identified!

3. Number of Instances: 214

4. Number of Attributes: 10 continuously valued attributes (including an Id#)
plus the class attribute (labelled 1 to 7)

5. Attribute Information:
1. Id number: 1 to 214
2. RI: refractive index
3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as
are attributes 4-10)
4. Mg: Magnesium
5. Al: Aluminum
6. Si: Silicon
7. K: Potassium
8. Ca: Calcium
9. Ba: Barium
10. Fe: Iron
11. Type of glass: (class attribute)
-- 1 building_windows_float_processed
-- 2 building_windows_non_float_processed
-- 3 vehicle_windows_float_processed
-- 4 vehicle_windows_non_float_processed (none in this database)
-- 5 containers
-- 6 tableware
-- 7 headlamps
    
```

Chargez les données et vérifiez que la matrice D et le vecteur $Names$ ont la bonne taille. Combien y a t'il d'échantillon de verre de type 1 (`building_windows_float_processed`)

```

load median_m8_2018
length(D)
length(Names)
length(find(D(:,end)==1))
    
```

1. Lors de la mise au point de cet examen, Il semblerait que six erreurs se soient cachées dans le code ci-dessous. Saurez vous les retrouver ?

```

x = D(:,end);
mean(x) % calcul de la moyenne de la derniere variable
X = D(:,1:end-1); % extraction de la matrice des donnees
[n,p] = size(X);
Xr = X-ones(n,1)*mean(X)./(ones(n,1)*std(X)); % normalisation des donnees
C = Xr'*Xr/n % calcul de la matrice des correlation
[V L] = eig(C); % calcul des axe factoriels
V = V(:,end:-1:1); % on les range du plus important au moins important
U = X*V; % calcul des composantes principales
Vn = V(:,1:2)*sqrt(L(1:2,1:2))/sqrt(n); % representation des variables
    
```

Exercice 2**ACP de verres (part 2)****10 points**

1. Magnesium
 - a) tracez la fonction de répartition empirique de la variable 4 Mg et commentez là
 - b) dessinez la boîte à moustache de la variable 4 Mg et commentez là
 2. Calculez les axes factoriels des variables aléatoires quantitatives du tableau D
 3. Quel est le nombre minimal d'axes factoriels qu'il faut considérer pour représenter au moins 50% de l'information contenue dans les variables aléatoires quantitatives du tableau D ?
 4. Les deux premières composantes principales.
 - a) Visualisez sur les deux premières composantes principales, les individus correspondant à chacun des classes, sur un même graphique avec des couleurs différentes.
 - b) proposez une interprétation de ces axes ?
 5. Comment interpréter les axes en terme de variables ?
 - a) Calculez les corrélations entre données et les composantes principales.
 - b) Visualisez les 4 premiers axes factoriels et commentez les.
 6. Refaites une ACP les variables aléatoires quantitatives mais uniquement des échantillons de verre de type 1 (`building_windows_float_processed`) tableau D . Quelles sont les différences les plus significatives avec les résultats de l'ACP sur le tableau complet ?
-

Exercice 3**Assimilation Bayésienne****4 points**

Soit m une moyenne connue et λ un facteur de confiance connu lui aussi. On souhaite connaître la valeur centrale d'un ensemble de points $x_i, i = 1, n$, sachant que cette valeur doit être proche de m avec une confiance λ . On modélise ce problème à travers une fonction coût J :

$$J(c) = \sum_{i=1}^n (x_i - c)^2 + \lambda(c - m)^2$$

1. Donnez l'expression de c qui minimise le coût J .
 2. Application numérique : calculez c pour $m = 1$, $\lambda = 5$, $n = 10$ et $x_1, \dots, x_{10} =$
 $-0.3380, -0.9348, -0.1982, -0.5399, -0.8393, -0.2697, -0.3243, -0.5245, -0.6172, -0.4141$
-

Annexes inutiles Float glass = Verre flotté (d'après wikipedia)

Le verre flotté ou verre « float » est un type de verre plat fabriqué par le principal procédé de fabrication de ce type de verre.

Dans le procédé « float » inventé par Sir Alastair Pilkington en 1952, un mélange de matières premières est chargé en continu dans le four de fusion. À la sortie du four, le verre forme un ruban flottant à la surface de l'étain fondu. La surface de l'étain fondu est extrêmement lisse, donnant au verre une planéité de surface parfaite. Le ruban de verre est ensuite lentement refroidi jusqu'à complet durcissement et recuit. Le ruban ainsi obtenu est d'épaisseur régulière et présente des surfaces parfaitement polies. Le ruban est ensuite découpé en plaques pour livraison.