SY19 Automne 2011

TP 1: Positionnement multidimensionnel

Exercice 1. Exercice théorique

On considère le tableau de données suivant :

$$X = \begin{pmatrix} 8.5 & 1.5 \\ 3.5 & 5.0 \\ 2.0 & 6.5 \\ 9.5 & 1.5 \\ 8.5 & 2.5 \\ 3.0 & 6.5 \\ 9.0 & 2.5 \\ 2.0 & 5.5 \end{pmatrix}.$$

Première partie: ACP

On associe les mêmes pondérations à tous les individus et on munit \mathbb{R}^p de la métrique euclidienne.

- 1. En utilisant la fonction eigen, calculer les axes factoriels de l'ACP du nuage de points ainsi défini. Quels sont les pourcentages d'inertie expliquée par chacun de ces axes?
- 2. Calculer les composantes principales; en déduire la représentation des huit individus dans le premier plan factoriel. Quelle est la transformation appliquée par l'ACP à un jeu de données de \mathbb{R}^2 représenté dans le plan factoriel?
- 3. Calculer l'expression $\sum_{\alpha=1}^k \mathbf{c}_{\alpha} \mathbf{u}'_{\alpha}$ pour les valeurs k=1 et k=2. Que remarque-t-on?

Deuxième partie: MDS

- 1. Calculer le tableau D^2 des distances euclidiennes associé à ces données.
- 2. Calculer la matrice W des produits scalaires, d'une part directement à partir de X, d'autre part, à partir de D^2 .
- 3. Vérifier que W (ou $\frac{1}{n}W$) est semi définie positive.
- 4. Déterminer la matrice V des vecteurs propres normées (au sens de D_p) de $\frac{1}{n}W$ et L la matrice diagonale des valeurs propres.
- 5. En déduire la représentation fournie par l'AFTD.
- 6. Tracer le nuage associé au tableau initial X et le nuage associé à la représentation fournie par l'AFTD. Comparer ces deux représentations.
- 7. En utilisant les résultats précédents, écrire en R la fonction aftd ayant comme argument d'entrée la distance D et comme argument de sortie la représentation dans \mathbb{R}^2 associée par l'AFTD et la qualité de cette représentation.

Exercice 2. Les données de mutations

Les principaux outils R permettant d'appliquer les méthodes de positionnement multidimensionnel sont les suivants :

- la classe dist permet de manipuler les tableaux de proximités (distance, dissimilarités, similarités).
 Les fonctions as.dist et as.matrix permettent de passer de la représentions matricielle classique à la représentation en tableau de distance.
- concernant les fonctions :
 - la fonction dist permet d'effectuer le calcul des principales distances (euclidienne, ...) à partir d'un tableau de données;
 - la fonction cmdscale réalise l'AFTD,
 - la fonction sammon (module MASS), la projection de Sammon,
 - et la fonction isoMDS (module MASS), la projection de Kruskal;
 - enfin, le module Shepard (module MASS) permet de tracer un diagramme de Shepard.

On se propose d'utiliser ces outils sur les données de mutations2.txt. Charger les données et les stocker dans un tableau de données (data frame) :

```
mutations <- read.table("mutations2.txt", header=F, row.names=1)</pre>
```

- 1. Effectuer l'AFTD en utilisant la fonction AFTD écrite dans l'exercice 1 et en utilisant la fonction cmdscale. Comparer les résultats alors obtenus.
- 2. Visualiser les données précédentes dans un espace à deux dimensions en utilisant les projections de Sammon et de Kruskal. Comparer aux résultats de l'AFTD.
- 3. Tracer les diagrammes de Shepard, et les comparer.

Exercice 3. Les données de distances entre aéroports

Les données contenues dans le fichier airport2.txt représentent les distances de vol entre des aéroports situés en différents points du globe. L'objectif de cet exercice est d'appliquer les différentes méthodes de positionnement multidimensionnel à ces données et de comparer les résultats obtenus. Charger les données et les stocker dans un tableau de données (data frame) :

```
airports <- read.table("airports2.txt", header=F, row.names=1)</pre>
```

- 1. Effectuer l'AFTD sur les données au moyen de la fonction cmdscale. Analyser les résultats obtenus.
- 2. Visualiser les données dans un espace à deux dimensions en utilisant les projections de Sammon et de Kruskal. Comparer à la représentation obtenue par AFTD.
- 3. Restreindre les données aux seuls aéroports européens; effectuer l'AFTD et représenter les données dans un espace à deux dimensions. Que constate-t-on? Comparer ces résultats obtenus à ceux obtenus en utilisant les projections de Sammon et de Kruskal.

```
MUTATION DISTANCES AMONG 20 SPECIES (FITCH AND MARGOLIASH)
     The source of this data is a paper by Fitch and Margoliash
| in Science(1967). For a more recent reference see Scientific
| American (1972?).
| Every species has a protein molecule, Cytochrome c, which varies |
| from species to species but has a similar function for all. It
| consists of a long chain of amino acids. There are only a few
| acids, but different molecules are obtained by varying the
\mid acids in each \mid position in the chain. The \mid number of positions
| with different acids measures distance between two species.
| these distances are given in the data below.
| For example, the amino acids in Cytochrome c for two species look |
| like this:
| Moth
                 XXYVPLY .....SEXI
| Screwworm fly XXYVPLY .....LSEI
| where the whole chain is 110 in length, and the letters represent |
| particular amino acids. Each difference contributes to mutation
| distance according to the minimum number of nucleotides that would|
| need to be changed to convert one into the other.
| Fitch & Margoliash used these data to construct a phylogenetic
| tree.
| Ref: Science, v. 155, 279-284.
+-----
```

Man 0 Monkey 01 0 13 12 0 Dog 17 16 10 0 Horse 16 15 08 01 0 Donkey 13 12 04 05 04 0 Pig Rabbit 12 11 06 11 10 06 0 12 13 07 11 12 07 07 0 Kangaroo 17 16 12 16 15 13 10 14 0 Pekin Duck Pigeon 16 15 12 16 15 13 08 14 03 0 18 17 14 16 15 13 11 15 03 04 0 Chicken 18 17 14 17 16 14 11 13 03 04 02 0 King Penguin 19 18 13 16 15 13 11 14 07 08 08 08 0 Snapping Turtle 20 21 30 32 31 30 25 30 24 24 28 28 30 0 Rattlesnake 31 32 29 27 26 25 26 27 27 27 26 27 27 38 0 Tuna 33 32 24 24 25 26 23 26 26 26 26 28 30 40 34 0 Screwworm Fly 36 35 28 33 32 31 29 31 30 30 31 30 33 41 41 16 0 Moth Bakers Mould 63 62 64 64 64 64 62 66 59 59 61 62 65 61 72 58 59 0 Bread Yeast 56 57 61 60 59 59 59 58 62 62 62 61 64 61 66 63 60 57 0 66 65 66 68 67 67 67 68 66 66 65 67 69 69 65 61 61 41 0 Skin Fungus