

Cette séance de 1h30 de TP est consacrée à:

- L'ACP en fonction du nuage de point,
- La validation et à la stabilité de l'Analyse en Composantes Principales et son approche dite 'dual' .

Vous devez disposer des codes du précédent TP1 ACP que vous avez développé.

Vous devez déposer votre TP par trinôme sur campus : le(s) script(s) développé(s) + le fichier de compte rendu (max 6 pages) . **Rendus de ce TP en fin de la semaine : 10 octobre**

Partie 1 - ACP et étude de nuage de point

Il s'agit d'étudier la qualité de la réduction en fonction des propriétés intrinsèques à l'échantillon. Ceci fait suite au TP1 sur la mise en œuvre d'une ACP centrée et d'une ACP normée et de sa qualité de réduction pour les variables et les individus.

1. Etude de la forme du nuage initiale et de sa répercussion sur la réduction de dimension

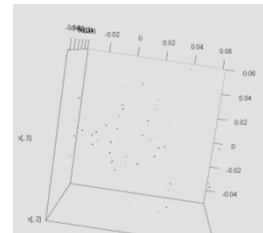
1.1 Nuage isotrope

On veut générer des données pour 3 variables d'un nuage de points proches d'une sphère. Pour cela vous pouvez choisir plusieurs méthodes mais une possibilité est de générer un échantillon de taille n des 3 variables X, Y, Z comme 3 vecteurs gaussiens indépendants de loi $N(0,1)$ puis de récupérer le vecteur V à 3 composantes tel que $V/\|V\|$. Car la densité $f(v)$ d'un vecteur gaussien est isotrope et dépend que de v .

Vous pouvez visualiser votre nuage par la :

`open3d()`

`plot3d(x[,1], x[,2], x[,3], col = rainbow(n))`



-Vous devez écrire le script qui génère ces données, puis mettre en œuvre l'ACP

- En fonction de la taille de l'échantillon, suivre l'évolution de la matrice de covariance ou de corrélation, la cascade des valeurs propres, ainsi que la qualité de la projection des individus. Interpréter.

1.2 Nuage non isotrope

On cherche maintenant à voir sur des formes non isotropes si des corrélations plus fortes entre certaines variables changent l'ACP et comment. On génère des données pour 3 variables qui constituent un nuage de points pour lesquelles deux variables sont corrélées fortement. Pour cela vous pouvez choisir plusieurs méthodes mais une possibilité est de générer un échantillon de taille n pour 3 variables X, Y,Z puis appliquer une relation linéaire entre les deux premières X et Y , puis entre Y et Z. Vous pouvez modifier votre corrélation entre deux variables en rajoutant un bruit.

Tester l'ACP sur les données générées et en fonction de la taille de l'échantillon suivre l'évolution de la matrice de covariance ou de corrélation, la cascade des valeurs propres, ainsi que la qualité de la projection des individus. Interpréter.

On vous donne le cas de 3 variables générées selon le script suivant :

```
x <- sort(rnorm(1000))
```

```
y <- rnorm(1000)
```

```
z <- rnorm(1000) + atan2(x, y)
```

- Que vous donne une telle ACP ? Devez-vous normer les données ?

1.3 Points extrémaux

Dans le cas d'un nuage dont la forme est non isotrope (choisir votre générateur de nuage) vous choisissez soit de rajouter certains points extrémaux pour une ou deux variables (vous allez étirer ainsi votre nuage selon cette direction) soit au contraire ôter des points extrémaux (ou quelques points) pour compacter votre nuage selon quelques variables.

- Proposer quelques essais d'ACP centrée ou ACP normée sur ces données, qu'observez-vous ? Que pouvez-vous conclure sur le fait de centrer ou normer les données.

2. Partie 2. Etude de la forme du nuage initiale sur la réduction de dimension dans les deux espaces

Il s'agit :

- De mettre en œuvre le volet 'dual', qui sera à la base de la décomposition par ACP
- De visualiser la qualité de la réduction en fonction des propriétés intrinsèques à l'échantillon selon ces deux approches R^p ou R^n
- De comprendre l'enjeu de la validation d'une méthode de traitement de données (plus généralement d'un modèle de prévision) (robustesse et sensibilité de la réduction)

Dans cette partie les données sont celles des deux nuages de la partie 1 :

Cas 1 : Nuage isotrope

Pour générer des données pour 3 variables d'un nuage de points proches d'une sphère vous pouvez choisir plusieurs méthodes mais une possibilité est de générer un échantillon de taille n des 3 variables X, Y, Z comme 3 vecteurs gaussiens indépendants de loi $N(0,1)$ puis de récupérer le vecteur V à 3 composantes tel que $V/\|V\|$. Car la densité $f(v)$ d'un vecteur gaussien est isotrope et dépend que de v .

Cas 2 Nuage non isotrope

Pour générer un nuage de points dans un R^3 pour lesquelles deux variables sont corrélées fortement, plusieurs méthodes existent. Une possibilité est de générer un échantillon de taille n pour 3 variables X, Y, Z puis appliquer une relation linéaire entre les deux premières X et Y , puis entre Y et Z . Vous pouvez modifier votre corrélation entre deux variables en rajoutant un bruit.

On vous donne le cas de 3 variables générées selon le script suivant :

```
x <- sort(rnorm(1000))
```

```
y <- rnorm(1000)
```

```
z <- rnorm(1000) + atan2(x, y)
```

Vous avez développé l'ACP centrée et centrée réduite sur l'espace R^p en TP1 : on vous demande successivement de :

- 1) Mettre en œuvre de l'ACP sur espace R^n sur les deux cas
- 2) De comparer avec les résultats dans R^p sur les deux cas
- 3) Retrouver les formules de passage à partir du cours et de vos résultats successifs
- 4) Valider les résultats sur vos exemples
- 5) Tester sur les deux cas : l'effet de la dimension sur les deux décompositions
- 6) Proposer la projection de points supplémentaires (espace sur R^p) et de points colonnes (variables) dans ce nouvel espace R^n . Pour cela vous devez proposer de générer soit des colonnes supplémentaires soit des lignes supplémentaires comme individus lignes :

Cas 1 : un ou deux vecteurs gaussiens V indépendants de loi $N(0,1)$ tel que $V/\|V\|$ (densité $f(v)$ d'un vecteur gaussien est isotrope et dépend que de v) permettront ici de créer de nouvelles variables et/ou étendre l'échantillon pour placer ensuite ces nouveaux points dans l'espace réduit.

Cas 2 : générer des points supplémentaires pour les 3 variables X, Y, Z et/ou générer une ou deux variables liées aux 3 précédentes. Attention ne pas refaire l'ACP sur ces données car elles sont dites non actives pour la réduction.

Modèle de rendu

Compte-rendu du TP n°2 - Analyse en composantes principales partie 2

Noms du groupe

Partie 1 - l'ACP et étude de nuage de point dans R^p

3. Etude de la forme du nuage initiale et de sa répercussion sur la réduction de dimension

1.1 Nuage isotrope

Méthode de génération de la matrice de données : code ou pseudo code

Principaux résultats de l'ACP :

En fonction de la taille de l'échantillon, suivre l'évolution de la matrice de covariance ou de corrélation, la cascade des valeurs propres, ainsi que la qualité de la projection des individus.

Interprétation

1.2 Nuage non isotrope

1) Méthode de génération de la matrice de données : code ou pseudo code

Principaux résultats de l'ACP :

En fonction de la taille de l'échantillon, suivre l'évolution de la matrice de covariance ou de corrélation, la cascade des valeurs propres, ainsi que la qualité de la projection des individus.

Interprétation

2) Cas du nuage proposé : $x \leftarrow \text{sort}(\text{rnorm}(1000))$ $y \leftarrow \text{rnorm}(1000)$ $z \leftarrow \text{rnorm}(1000) + \text{atan2}(x, y)$

*Résultats de l'ACP
normées les données ?*

3.3 Points extrémaux

*Description de la démarche adoptée
essais d'ACP centrée ou ACP normée sur ces données,
résultats
Conclusion sur le fait de centrée ou normée les données.*

4. Partie 2. Etude de la forme du nuage initiale sur la réduction de dimension dans les deux espaces \mathbb{R}^p ou \mathbb{R}^n

Le compte rendu devra restituer les différentes étapes, pas de modèle de rendu mais il doit contenir :

- L'ACP sur espace \mathbb{R}^n sur les deux types de nuage isotrope et non isotrope
 - La comparaison de l'ACP dans \mathbb{R}^p sur les deux cas
 - Les formules de passage à partir du cours et de vos résultats successifs
 - Tester sur les deux cas : l'effet de la dimension sur les deux décompositions
 - Proposer la projection de points supplémentaires (espace sur \mathbb{R}^p) et de points colonnes (variables) dans ce nouvel espace \mathbb{R}^n . Pour cela vous devez proposer de générer soit des colonnes supplémentaires soit des lignes supplémentaires comme individus lignes :
 - Cas 1 : un ou deux vecteurs gaussiens V indépendants de loi $N(0,1)$ tel que $V/\|V\|$ (densité $f(v)$ d'un vecteur gaussien est isotrope et dépend que de v) permettront ici de créer de nouvelles variables et/ou étendre l'échantillon pour placer ensuite ces nouveaux points dans l'espace réduit.
 - Cas 2 : générer des points supplémentaires pour les 3 variables X, Y, Z et/ou générer une ou deux variables liées aux 3 précédentes. Attention ne pas refaire l'ACP sur ces données car elles sont dites non actives pour la réduction.
-