Algèbre Linéaire et Analyse de Données

Examen Machine 2023 - 2024
Sujet A, Durée : 1h30
Licence 2 MIASHS

Guillaume Metzler

Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Résumé

L'examen comporte trois exercices qui sont indépendants et qui nécessitent de reprendre les notions vus en cours depuis le début de l'année.
Les deux premiers exercices portent sur la partie *Prise en Main de* R. Le troisième exercice porte sur l'*Analyse en Composantes Principales.*

Vous répondrez directement dans le fichier R qui accompagne ce sujet. Ce fichier R contiendra aussi bien le code que les commentaires qui permettra de répondre aux questions du sujet.

Quand vous aurez terminé, vous déposerez directement votre examen sur l'espace de dépôt prévu à cet effet, directement sur Moodle.

Exercice 1

Dans cet exercice, on va considérer la famille de vecteurs $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ et la matrice A définies par

$$\mathbf{u}_1 = \begin{pmatrix} 2 \\ -2 \\ -5 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -2 \\ 2 \\ -1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -5 \\ 3 \\ 4 \end{pmatrix}$$

et la matrice $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} 4 & -1 & -3 \\ -1 & 3 & -6 \\ -3 & -6 & 2 \end{pmatrix}.$$

1. **Famille de vecteurs**
   (a) Est-ce que la famille de vecteurs $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ est une famille libre ?
   (b) Calculer le produit scalaire entre les différents vecteurs. Est-ce que la famille de vecteurs est une famille orthogonale ?
   (c) Calculer la norme des différents vecteurs.
   (d) Déterminer la projection du vecteur $\begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix}$ sur le vecteur $\mathbf{u}_1$.
2. **Etude de la matrice A**
   (a) La matrice $\mathbf{A}$ est-elle inversible ?
   (b) Expliquer (sans calculs) pourquoi la matrice $\mathbf{A}$ est diagonalisable.
   (c) Déterminer les valeurs propres de la matrice $\mathbf{A}$ et expliquer pourquoi la matrice est diagonalisable à l'aide de ce nouveau calcul.

Exercice 2

On considère le système suivant

$$\begin{cases} 2x - 3y + 6z - 3u & = & 2 \\ z + u & = & -5 \\ 4u & = & -8 \\ 7y - 3z + 2u & = & -1 \end{cases}$$

1. Rappeler la définition de système de Cramer.
2. Le système est-il un système de Cramer ?
3. Résoudre le système S à l'aide des formules de Cramer puis par une méthode de votre choix.

Exercice 3

Ce dernier exercice a pour objectif de vous faire étudier un jeu de données à l'aide de l'Analyse en Composantes Principales.
Les données utilisées se trouvent directement dans le fichier R qui accompagne ce sujet. On présente rapidement les différentes variables, seules les huit dernières seront utilisées pour l'analyse :

- *Intensité odeur*
- *Typicité odeur*

- *Caractère pulpeux*
- *Intensité goût*
- *Caractère acide*
- *Caractère amer* : chrono au 110m haies (en s).
- *Caractère sucré*
- *Glucose*
- *Fructose*
- *Saccharose*
- *pH*
- *Titre*
- *Acide citrique*
- *Vitamine C*

Dans la suite, nous noterons **D** la matrice des données (ou encore la matrice de design). On rappelle que les individus sont représentés en ligne et que les variables sont représentées en colonne.

Préparation des données

1. Déterminer le barycentre du jeu de données.
2. Calculer l'écart-type de chaque variable.
3. Créer une matrice **Z** qui contiendra votre jeu de données *centré, réduit* et *normé.*
4. Déterminer la norme de Frobenius de la matrice **Z**. Qu'est-ce que représente cette quantité ?

Analyse du nuage des individus

Dans la suite de l'étude vous allez étudier le nuage des variables. Il sera judicieux de se reporter aux commandes graphiques qui ont été vues en cours pour faire la représentation et notamment pour l'interprétation.

1. Définir la matrice qui permet d'analyser le nuage des variables, on la notera **R**. Quelle est sa dimension ?
2. Diagonaliser la matrice **R**.
3. Quelle est lien entre la somme des valeurs propres de **R** et la norme de Frobenius de **Z** ?
4. Définir des objets $g1$ et $g2$ qui représentent les composantes principales des individus sur le premier plan factoriel (on désignera par $\mathbf{v}_1$ et $\mathbf{v}_2$ les vecteurs qui forment la base de ce premier plan factoriel).
5. Rappeler le lien qui existe entre *les composantes principales des variables* et *les axes principaux* lors de l'étude du *nuage des individus*
6. Représenter ces individus sur le premier plan factoriel.
7. Quelle est la quantité d'information préservée lors de la projection des données sur le premier plan factoriel ? Quid sur le deuxième plan factoriel ?
8. Au regard du graphique obtenu, quel sens pourriez-vous donner aux axes $\mathbf{v}_1$ et $\mathbf{v}_2$? Justifiez votre réponse en indiquant les variables initiales qui ont permis votre interprétation.
9. Mise à part graphiquement, comment auriez-vous pu également faire cette interprétation ?