Algèbre Linéaire et Analyse de Données

Examen Machine 2021 - 2022, Sujet 1, Durée : 1h20
Licence 2 MIASHS

Guillaume Metzler
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Résumé

L'examen comporte deux exercices qui sont indépendants et qui nécessitent de reprendre les notions vus en cours depuis le début de l'année.
Le premier exercice porte 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

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

et

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

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. Que constatez-vous ? Que peut-on en déduire ?
   (c) Calculer la norme des différents vecteurs.
   (d) Déterminer la projection du vecteur $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ sur le vecteur $\mathbf{u}_1$.
2. **Etude de la matrice A**
   (a) La matrice A est-elle inversible ?
   (b) Expliquer (sans calculs) pourquoi la matrice A est diagonalisable.
   (c) Déterminer les valeurs propres de la matrice A et expliquer pourquoi la matrice est diagonalisable.
   (d) Calculer les produits $A\mathbf{u}_1, A\mathbf{u}_2$ et $A\mathbf{u}_3$. Que remarquez-vous ?
   (e) Que peut-on dire de la famille de vecteurs $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ vis-à-vis de la matrice A ?

Exercice 2

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 :

- *titreCours* : le titre du cours.
- *idCours* : l'identifiant du cours.
- *inscription* : nombre de jours écoulés depuis votre inscription au cours.
- *progression* : votre progression sur le cours (en pourcentage).
- *moyenneDeClasse* : moyenne de la classe aux évaluations (en pourcentage).
- *duree* : durée estimée du cours (en heures).
- *difficulte* : difficulté estimée du cours (1 : facile... 3 : difficile).
- *nbChapitres* : nombre de chapitres.
- *nbEvaluations* : nombre d'évaluations dans le cours (comprend les quiz et les activités).
- *ratioQuizEvaluation* : proportion de quiz par rapport au nombre total d'évaluations (nombre d'évaluations : nombre de quiz + nombre d'activités).

Dans la suite, nous noterons X 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. Quelle est la norme de Frobenius de la matrice Z ?

Analyse du nuage des individus

Dans la suite de l'étude vous allez étudier **le nuage des individus**. 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 individus*, on la notera C. Quelle est sa dimension ?
2. Diagonaliser la matrice C.
3. Quelle est lien entre la somme des valeurs propres de C et la norme de Frobenius de Z ?
4. Définir des objets $f1$ et $f2$ qui représentent les composantes principales des individus sur le premier plan factoriel (on désignera par $\mathbf{u}_1$ et $\mathbf{u}_2$ les vecteurs qui forment la base de ce premier plan factoriel).
5. Représenter ces individus sur le premier plan factoriel.
6. Quelle est la quantité d'information préservée lors de la projection des données sur le premier plan factoriel ?
7. Rappeler le lien qui existe entre *les composantes principales des individus* et *les axes principaux* lors de l'étude du *nuage des variables*
8. On admettra que les axes principaux peuvent s'interpréter comme suit :
   — : l'axe $\mathbf{u}_1$ s'interprète comme étant la *longueur du cours*
   — : l'axe $\mathbf{u}_2$ s'interprète comme étant la *facilité du cours*

   A partir de ces informations et du graphe des individus que vous aurez généré. Interpréter ce nuage de points (ressemblances et différences entre les individus, caractéristiques des individus, est-ce qu'il y a des groupes qui se dessinent, ...)