

Travaux Pratiques Matlab

Analyse en Composantes Principales

- Traiter les exercices 1, 2 et 3 dans des fichiers de script séparés
- Rédiger un rapport pour les exercices 2 et 3 : ces rapports doivent être rédigés dans un même fichier converti au format pdf et envoyé par mail à l'adresse serge.mazaauric@cpe.fr avant le vendredi 8 mars minuit. Le nom du fichier doit avoir la structure suivante :

NOM1_NOM2_TP_ACP.pdf (majuscules obligatoires)

Pour chaque exercice, le rapport doit être structuré de la façon suivante :

Introduction

- Décrivez le problème (comment le comprenez-vous ? quel est le contexte ?), en rappeler les éléments essentiels.
- Une personne n'ayant pas l'énoncé sous les yeux doit être capable de comprendre la problématique.
- Ne pas copier l'énoncé tel quel, ça ne sert à rien.

Méthode

- Décrivez ce que vous faites (sans nécessairement donner d'explication dans un premier temps) : soyez factuel !
- Vous pouvez éventuellement donner toutes les étapes de votre raisonnement.
- Afficher les parties importantes du code (il n'est pas nécessaire de systématiquement tout afficher). Ce code doit être clairement commenté presque « ligne par ligne » sans nécessairement faire de phrase. Utiliser la police « courier new » pour afficher les codes Matlab.
- **A la suite de ce code, en donner une description globale, synthétique (cette fois-ci en faisant des phrases).**
- Ne pas afficher plusieurs fois le même code sous prétexte que la valeur d'une variable a changé.

Résultats

- Afficher les résultats, les décrire puis en donner une interprétation. Si ces résultats comportent des figures, celles-ci doivent être accompagnées d'une légende (se contenter de « figure 1 », « figure 2 » est insuffisant). Si ces figures sont des graphiques, les unités des axes doivent être renseignées.
- Essayer d'aller « plus loin » dans l'analyse des résultats (par exemple, faire varier un paramètre du problème puis afficher, décrire et interpréter les nouveaux résultats).
- Ne pas hésiter à rappeler des éléments du cours, comme par exemple un théorème ou une formule : ce type de renseignement permet parfois de mieux se faire comprendre (et il est toujours bien de montrer que vous avez bien fait le lien entre l'exercice pratique et la théorie). Attention, écrire « cf. cours page 15 » n'est pas acceptable.

Conclusion :

- Rappelez la problématique ainsi les principaux résultats que vous avez obtenus.
- Essayer de faire une synthèse globale du problème que vous venez d'étudier en mettant en avant ce que vous en avez compris.

Merci également de respecter les consignes suivantes :

- Ecrire en première page les noms des deux personnes qui constituent le binôme et numéroté les pages du rapport.
- Si vous êtes bloqués ou si vous n'arrivez pas à interpréter un résultat, le signaler dans votre rapport. Idem, si vous avez demandé de l'aide auprès d'un autre binôme (donnez les noms de ce binôme). On rappelle que le plagiat est bien évidemment interdit ! La note de zéro est attribuée à tout binôme convaincu de plagiat.
- Relisez-vous : corrigez les fautes d'orthographe, assurez-vous que vos phrases aient un sens (on n'écrit pas comme on parle !).
- Ecrire au présent (n'utiliser le futur que si le contexte le nécessite) : écrire « la fonction est positive » plutôt que « la fonction sera positive ».
- Eviter de trop personnaliser : écrire « la fonction f » ou « la variable x », plutôt que « notre fonction f » ou « ma variable x ».
- Les équations doivent être tapées à l'aide de l'éditeur d'équations fourni avec le logiciel de traitement de texte.
- Ne pas intégrer de copies d'écran dans votre rapport, elles sont souvent illisibles. Idem pour les photographies prises avec un smartphone. Sauvegarder les courbes Matlab, de préférence au format .eps (ou alors en .jpg).
- Ecrire avec une police de caractères de taille raisonnable (10, 11 ou 12), idem pour l'interligne. Un rapport écrit en taille 14 avec un interligne de 2, « histoire de faire du volume », ne trompe personne !

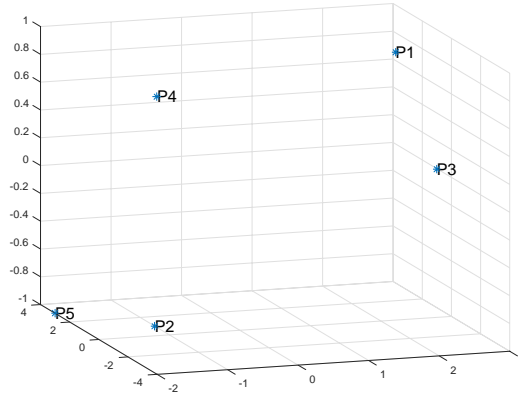
Exercice 1 : un nuage de points presque coplanaires

On considère un nuage de points de \mathbb{R}^3 dont les coordonnées sont les suivantes :

$$P_1(2, -1, 1), \quad P_2(-1, 1, -1), \quad P_3(3, 1, 0), \quad P_4(-2, -4, 1), \quad P_5(-2, 3, -1),$$

1) Afficher le nuage de points dans un repère à trois dimensions (commande `plot3`).

On constate, en faisant tourner le repère, que les points sont presque coplanaires.

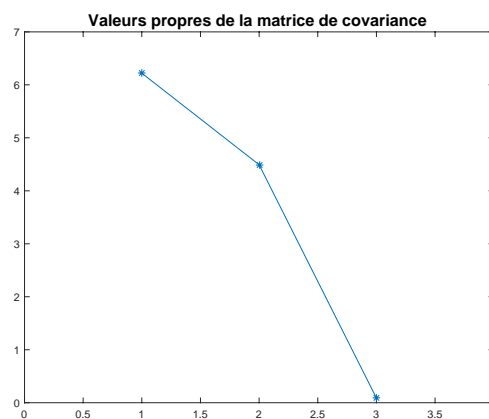


2) On souhaite alors déterminer l'équation du plan passant au plus proche des 5 points du nuage. Pour cela on propose d'abord le problème à l'aide de l'ACP (en reprenant les notations du cours) :

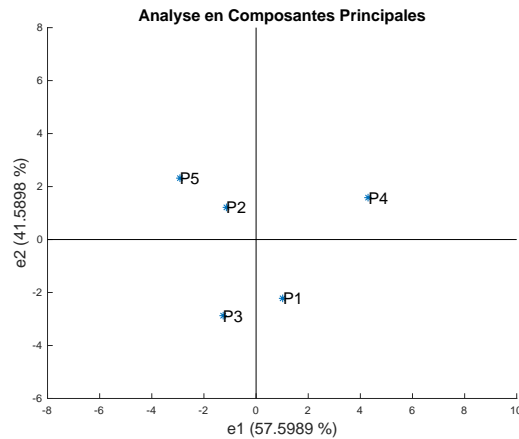
- Identifier les variables et les individus puis en déduire les valeurs de n et m
- Construire la matrice des données (on la notera Y) et affecter les valeurs à n et m (commande `size`)
- Construire la matrice des données centrée (on la notera X , utiliser la commande `mean`, pas de boucle `for`).

Que constatez-vous ?

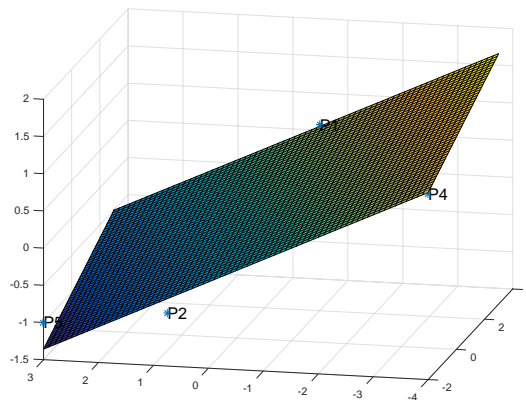
- Construire la matrice de covariance (on la notera M)
- Diagonaliser la matrice de covariance.
- Placer les valeurs propres dans un tableau, que l'on notera λ , dans l'ordre décroissant puis placer dans une matrice, que l'on notera P , les vecteurs propres en respectant l'ordre des valeurs propres (commandes `flipud` et `fliplr`).
- Afficher la courbe des valeurs propres (comme ci-dessous)



- Calculer le taux d'inertie expliquée pour chaque axe factoriel.
- Afficher la matrice P . Que représentent les vecteurs colonnes de cette matrice ?
- Calculer et afficher la matrice des composantes principales (que l'on notera X_{star}).
- Afficher le nuage de points projeté sur les deux premiers axes factoriels (comme ci-dessous)



- L'équation du plan défini par les deux premiers axes factoriels \mathbf{e}_1 et \mathbf{e}_2 a pour équation $ax + by + cz + d = 0$ où a, b et c sont les composantes d'un vecteur normal au plan factoriel $(\mathbf{e}_1, \mathbf{e}_2)$.
 - Quelle est la valeur de d ?
 - Calculer à l'aide de Matlab les valeurs de a, b et c
- Indication :** on peut fixer $a = 1$ et montrer que le problème revient à résoudre un système de la forme $\mathbf{A}\mathbf{U} = \mathbf{V}$ où $\mathbf{U} = (b, c)^T$, \mathbf{A} est une matrice et \mathbf{V} un vecteur que l'on précisera
- Afficher, dans le repère 3D de la question 1, le plan d'équation $ax + by + cz + d = 0$ (commandes `quiver` et `surf`)



- 3) Exécuter à nouveau votre programme en remplaçant le nuage de points précédents par les points suivants et vérifier si tout fonctionne correctement :

$$P_1(2,1,3), \quad P_2(-2,5,-3), \quad P_3(3,-1,0), \quad P_4(5,-1,3), \quad P_5(-2,3,-1), \quad P_6(1,2,-4), \quad P_7(-4,-2,1)$$

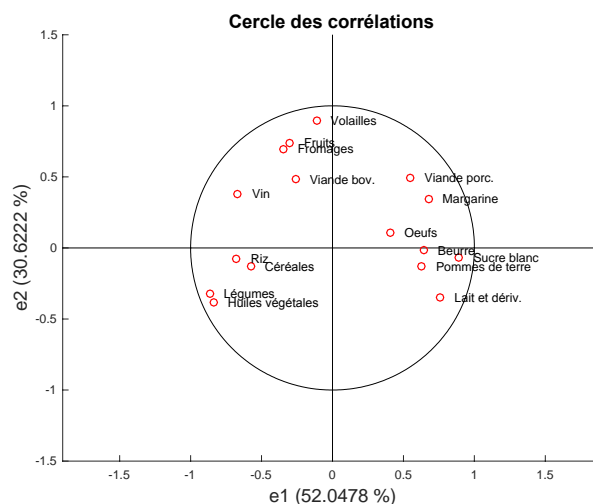
Exercice 2 : habitudes alimentaires

Utiliser l'analyse en composante principale des données présenter dans le tableau ci-dessous afin de voir s'il existe un moyen de distinguer les pays selon leurs habitudes alimentaires.

	RFA	France	Italie	Pays Bas	Belg. Lux.	Russie	Irlande	Danemark	Grèce
Céréales	68	76	127	60	72	68	84	68	82
Riz	2	4	3	3	2	3	2	2	4
Pommes de terre	81	74	41	81	101	102	111	68	70
Sucre blanc	36	34	31	37	35	40	42	44	24
Légumes	64	115	172	82	60	76	83	57	237
Fruits	89	53	69	84	64	33	30	38	57
Vin	25	93	87	13	66	6	3	14	45
Huiles végétales	6	12	20	4	4	8	8	7	22
Margarine	7	3	1	12	10	6	4	12	1
Viande bov.	24	33	25	21	28	23	23	13	20
Viande porc.	58	38	24	41	41	26	32	51	18
Volailles	84	90	80	136	87	14	14	9	13
Lait et dériv.	85	91	84	135	89	134	187	159	64
Beurre	6	7	2	3	8	5	11	6	1
Fromages	12	18	13	13	10	6	3	10	8
Œufs	17	15	11	11	14	14	14	14	12

Quelques indications :

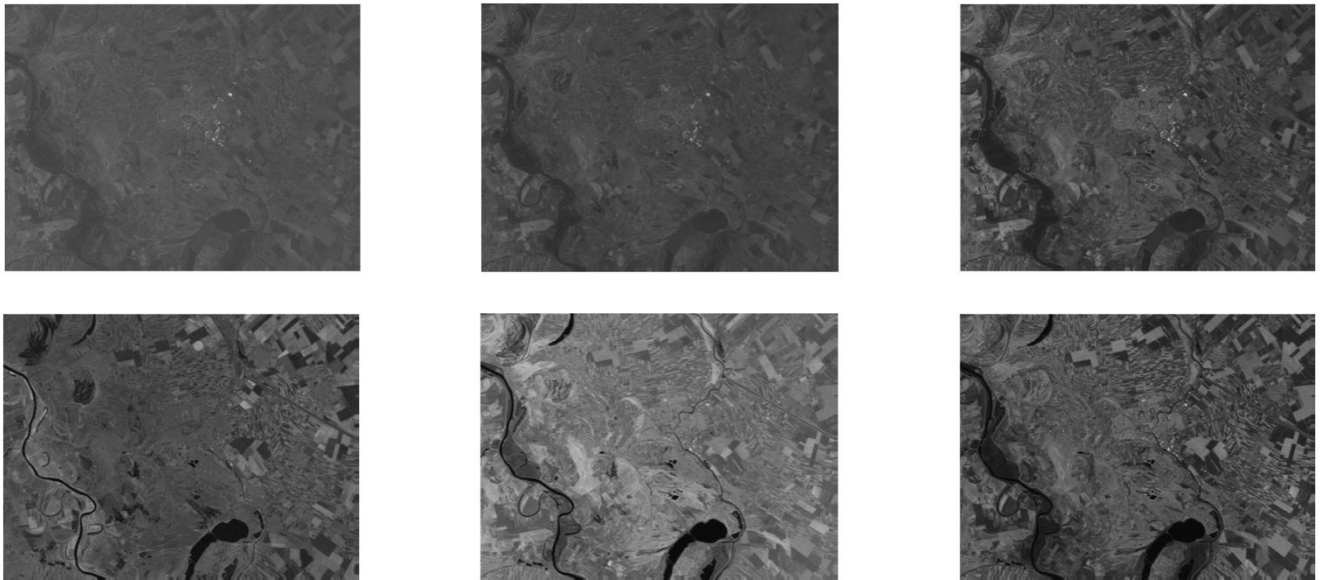
- Récupérer le fichier Ex2_data.m sur CPe-campus
- S'inspirer fortement du travail effectué pour l'exercice 1, notamment pour les notations (qui sont celles du cours), les affichages, etc ...
- Mettre en évidence le taux d'inertie expliquée pour chaque axe factoriel.
- On pourra éventuellement ajouter des graphiques montrant la projection du nuage sur des plans autres que le plan factoriel principal (e_1, e_2) et interpréter ces résultats.
- Afficher le cercle des corrélations (cf. cours) sur le plan factoriel principal et interpréter les résultats (comme ci-dessous).



Exercice 3 : Transformée de Karhunen-Loève (TKL) – d'après Stéphane Grousson

La transformée de Karhunen-Loève est le nom que l'on donne à l'analyse en composantes principales lorsque cette dernière est appliquée au traitement des images.

A partir de 6 images satellitaires d'une même vue mais acquises à travers des canaux fréquentiels différents, nous proposons ici de construire une seule image en conservant le maximum d'informations.



Quelques indications :

- Récupérer sur CPe-campus les fichiers des 6 images `i1.jpg`, ..., `i6.jpg` ainsi que le fichier Matlab `Ex3_tmp.m`
- Exécuter le programme `Ex3_tmp.m`
- Construire la matrice des données (que l'on notera Y) comportant 6 colonnes (1 par image) et n lignes où n est le nombre de pixels d'une image ($n = H \times W$)
- Faire les calculs de l'ACP (matrice des données centrée X , matrice de covariance M , calcul et affichage des valeurs propres dans l'ordre décroissant, matrice de projection P , matrice des composantes principales X_{star})
- Calculer l'inertie expliquée par le premier axe factoriel
- Construire la matrice (que l'on notera I_{rec}) de dimension $H \times W$ à partir de la première composante principale (commande `reshape`)
- Afficher l'image correspondant à la matrice I_{rec} et commenter.