

Evaluation des dossiers sur les mathématiques du Big Data

Les « 12 travaux d'AstéRix »

Daif Hakim

27/01/2021

ACP

1. Critères d'évaluation

Dans ce dossier, nous allons parcourir et faire une évaluation des travaux de mes camarades en mathématiques du Big Data:

Nos critères seront :

1. Visuel et l'organisation du dossier: Es ce que la mise en page est propre, agréable à lire ou non ? Le travail est-il organisé (chapitres, sections, sous-sections, Parties)

2. Qualité du latex et des formules mathématiques : Es ce que les formules sont claires ou pas ? Es ce que le dossier contient le code Latex ou le rapport sous format Rmd ? **3. Compréhension de l'idée générale:** Les Auteurs arrivent-ils à transmettre l'idée générale ?

4. Explication et compréhension des formules mathématique et des concepts fondamentaux: Es ce que les formules et les notions sont bien expliquées ?

5. Difficulté et originalité du thème : Le sujet abordé est-il original/difficile ou pas ? ## 2. Lien vers le document commenté

Le document évalué dans le cadre de ce rendu a été fait par Benjamin GUIGON sur l'analyse des composantes principales.

lien vers son Github: <https://github.com/benjaminiguigon/PSBX/blob/main/Article%20ACP%20.pdf>

le dossier en question est **Article ACP .pdf**

3. Synthèse de la présentation

L'ACP est une méthode factorielle de réduction de dimension pour l'exploration statistique de données quantitatives complexes dont l'objectif est de revenir à un espace de dimension réduite en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales.

banjamen, commence tout d'abord par une petite introduction sur l'ACP puis il entame les différentes étapes pour construire le modèle associé à ce dernier en explicitant à chaque étape son aspect mathématique,

l'auteur de ce document a consacré toute une partie pour montrer les objectifs de l'ACP, malgré qu'il a bien expliqué, et d'une façon très simple, l'objectif général de l'analyse par les composantes principales dans la partie Introduction, ce qui facilite vraiment la compréhension de rapport surtout pour un débutant.

Banjamin revient ensuite dans une autre partie sur l'aspect théorique des modèles de l'ACP

L'auteur a aussi explicité les différentes métriques mathématiques essentiels qui sont au cœur du processus de la réduction.

au final il expose les différentes représentations graphiques du modèle construit, qui vont nous permettre d'interpréter les données multidimensionnelles.

Explication des formules

La première étape de l'analyse des composantes principale consiste à représenter les données au sein du matrice de variance-covariance qui va permettre de réaliser ce résumé pertinent.

Matrice des covariances :

$$S = \sum_{i=1}^n \omega_i (x_i - \bar{x}) (x_i - \bar{x})' = \bar{X}' D \bar{X} \quad (1)$$

ne fois la matrice covariance est définie, on détermine les valeurs propres de chaque vecteurs (sachant la sommes de ces valeurs représente l'inertie total en d'autres termes : centre de gravité) puis on cherche les vecteurs propres des individus qui vont nous permettre de faire un graph des variables et des individus sur un espace plus restreint (en déformant le moins possible la configuration globale des individus selon l'ensemble des variables initiales) afin de d'interpréter les données multidimensionnelles

Les vecteurs $f^k = M \mathbf{v}^k$ sont les facteurs principaux. Ils permettent de définir les combinaisons linéaires des Xjoptimales au sens ci-dessus.

- Les vecteurs $c^k = \bar{X} f^k$ sont les composantes principales.
- Les variables C^k associées sont centrées, non corrélées et de variance λ_k :

ce sont les variables principales:

$$\begin{aligned} \text{cov}(C^k, C^l) &= (\bar{X} f^k)' D \bar{X} f^l = \mathbf{f}^{k'} \mathbf{S} \mathbf{f}^l \\ &= v^{k'} \mathbf{M} \mathbf{S} \mathbf{M} v^l = \lambda_l v^{k'} \mathbf{M} v^l = \lambda_l \delta_k^l \end{aligned}$$

5. Evaluation

1. Visuel et l'organisation du dossier: Le visuel est moyen, travail mal organisé (section, titres, exemples), on a du mal à comprendre l'enchainement des idées (on reconnait la difficulté d'exposer les notions mathématiques d'un tel thème), les parties et les sections sont mal placées, on trouve par exemple un partie nommée "partie 1" alors qu'il y'a même pas une deuxième partie, "PROPOSITION 2" sans qu'on trouve "PROPOSITION 1"

Trop de formules sans explications concrètes ce qui ne nous laisse pas suivre le raisonnement de l'auteur.

2. Qualité du latex et des formules mathématiques :

les formules sont claires, on a l'impression que l'auteur maitrise les codes latex concernant les formules malgré qu'on ne trouve pas dans ce dossier le code latex ce qui ne nous permet pas d'évaluer sa maitrise de ce code.

3. Compréhension de l'idée générale : avant de critiquer un tel sujet il convient tout d'abord de reconnaître sa difficulté, peut-être on peut facilement comprendre les objectifs de l'ACP et l'idée général de

ce thème à travers la première partie (introduction et partie: modèle), mais si on suit l'enchainement de ses idées notamment les notions mathématiques qu'il a exposé on aura du mal à comprendre les détails et on se perd très facilement, surtout pour une personne qui ne maîtrise pas bien les notions algébrique.

Il aurait pu ajouter un cas pratique ou une application pour faciliter la compréhension.

4. Explication et compréhension des formules mathématique et des concepts fondamentaux:

Des formules très claire sur le plan visuel, mais très difficile à comprendre car y'en trop, puis la plupart sont sans explication concrète (manque d'exemples simples).

5. Difficulté et originalité du thème : le sujet abordé est original malgré que c'est un sujet très classique, en termes de difficultés, c'est l'un des sujets les plus difficile à comprendre parce que ca nécessite un niveau très avancé en algèbres et en mathématique, personnellement j'ai eu l'occasion d'avoir des modules ou on a eu l'ACP comme cours et ca m'a pris beaucoup de temps à comprendre une petite parties des notions mathématiques ce qui montre la difficulté de ce sujet.

6. Conclusion

Il s'agit d'un document original et très utile lorsque l'on est en présence d'une somme importante de données quantitatives à traiter et interpréter, c'est un outil extrêmement puissant de compression et de synthèse de l'information.

on apprécie bien le choix de Benjamin et l'effort qu'il fournit pour ce travail, on sent qu'il maîtrise bien l'algèbre et quelques notions mathématique de l'ACP.

Cependant il aurait pu détailler encore plus son sujet avec des exemples simples et concrets et des explication précises pour chaque formule, cela facilitera au mieux la compréhension du sujet.