

Science des données II : module 7



Analyse en composantes principales

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Laboratoire d'Écologie numérique

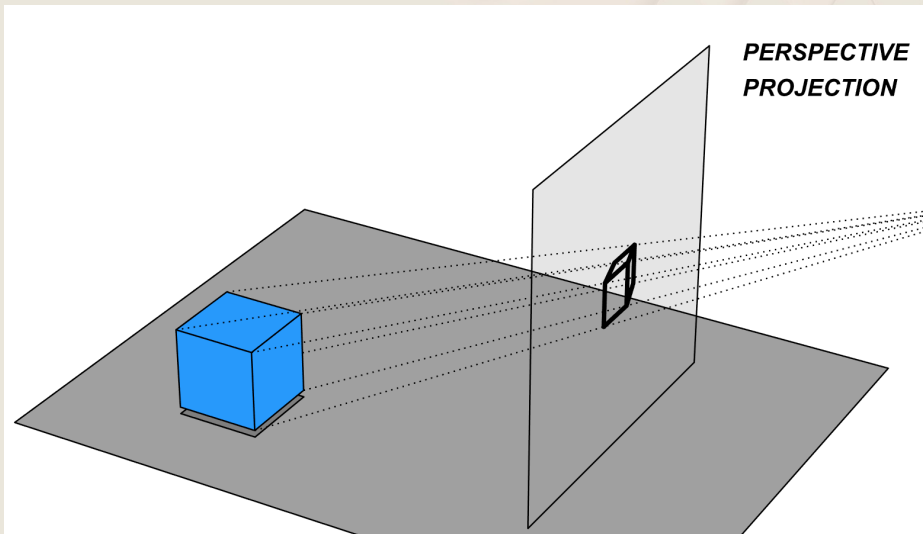


<https://wp.sciviews.org>
sdd@sciviews.org

- Revoir le mécanisme, ni les calculs mathématiques derrière l'ACP
- Nous allons expliciter ensemble l'**idée générale** qui est à l'origine de cette technique
- Objectif : **réduire le nombre de dimensions**, tout en conservant un maximum de l'information utile initiale => **Comment déterminer l'information utile ?**

Photo d'un objet tridimensionnel

Une photo (vue en 2D) d'une scène tridimensionnelle est une réduction de dimension. L'**angle** de prise de vue de l'objet est important.



Qu'est-ce que c'est ?



UMONS

Qu'est-ce que c'est ?



UMONS

Qu'est-ce que c'est ?



Qu'est-ce que c'est ?



L'angle sous lequel la projection de 3D en 2D se fait est important.
Certains angles portent plus d'information que d'autres.

Qu'est-ce que c'est ?



UMONS

Qu'est-ce que c'est ?



Qu'est-ce que c'est ?



UMONS

Qu'est-ce que c'est ?



En général, la projection qui correspond à l'**étalement le plus grand** de l'objet est aussi celle qui porte le plus d'information.

- En général, la projection qui correspond à l'**étalement le plus grand** de l'objet est aussi celle qui porte le plus d'information.
- En statistique, l'étalement se mesure via la variance (ou la covariance si plus d'une variable).
- L'ACP détermine la projection en un nombre de dimensions réduites visant à **étalement le plus possible** le nuage de points au sens de la variance - covariance (ou de la corrélation si les données sont centrées et réduites).
- Note : les présentations sont disponibles sur https://github.com/BioDataScience-Course/sdd_lessons

Vous avez la parole !

