

# Principal Component Analysis

## Machine Learning

**Tarek Kouki** tarek.kouki@esen.tn

**Aymen Jemi** aymen.jemi@esen.tn

April 25, 2019

*We Invest In  
Intelligence*

# Sommaire

Introduction

Pourquoi utilisons-nous l'ACP ?

Quelles sont les limites d'ACP ?

Exemple

# Qu'est-ce que l'ACP

## Introduction

- Selon le **Wikipedia** L'analyse des composants principaux (ACP) est une procédure statistique qui utilise une transformation orthogonale pour convertir un ensemble d'observations de variables éventuellement **corrélées** en un ensemble de valeurs de linéairement variables **non corrélées** appelées **composants principaux** .
- Elle permet au praticien de **réduire** le nombre de variables et de rendre l'information **moins redondante** .

# Pourquoi ACP ?

## Réduction de donnée

- ACP est le plus souvent utilisé pour condenser les informations contenues dans un grand nombre de variables d'origine en un ensemble plus petit de nouvelles dimensions composites, avec une perte minimale d'informations.

# Pourquoi ACP ?

## Interprétation

- ACP peut être utilisé pour découvrir les caractéristiques importantes d'un grand ensemble de données.  
Il révèle souvent des relations insoupçonnées auparavant, permettant ainsi des interprétations qui ne résulteraient pas normalement.
- La ACP est généralement utilisée comme étape intermédiaire dans l'analyse des données lorsque le nombre de variables d'entrée est trop important pour permettre une analyse utile.

## Limites

- ACP est basée uniquement sur le vecteur moyen et la matrice de covariance.
- Certaines distributions (normale multivariée) sont caractérisées par cela, mais d'autres ne le sont pas.
- Si les variables sont corrélées, ACP peut atteindre une réduction de dimension. Sinon, l'ACP les ordonne simplement en fonction de leur écart.

## Example

- <https://github.com/jemiaymen/machine-learning/blob/master/pca.ipynb>