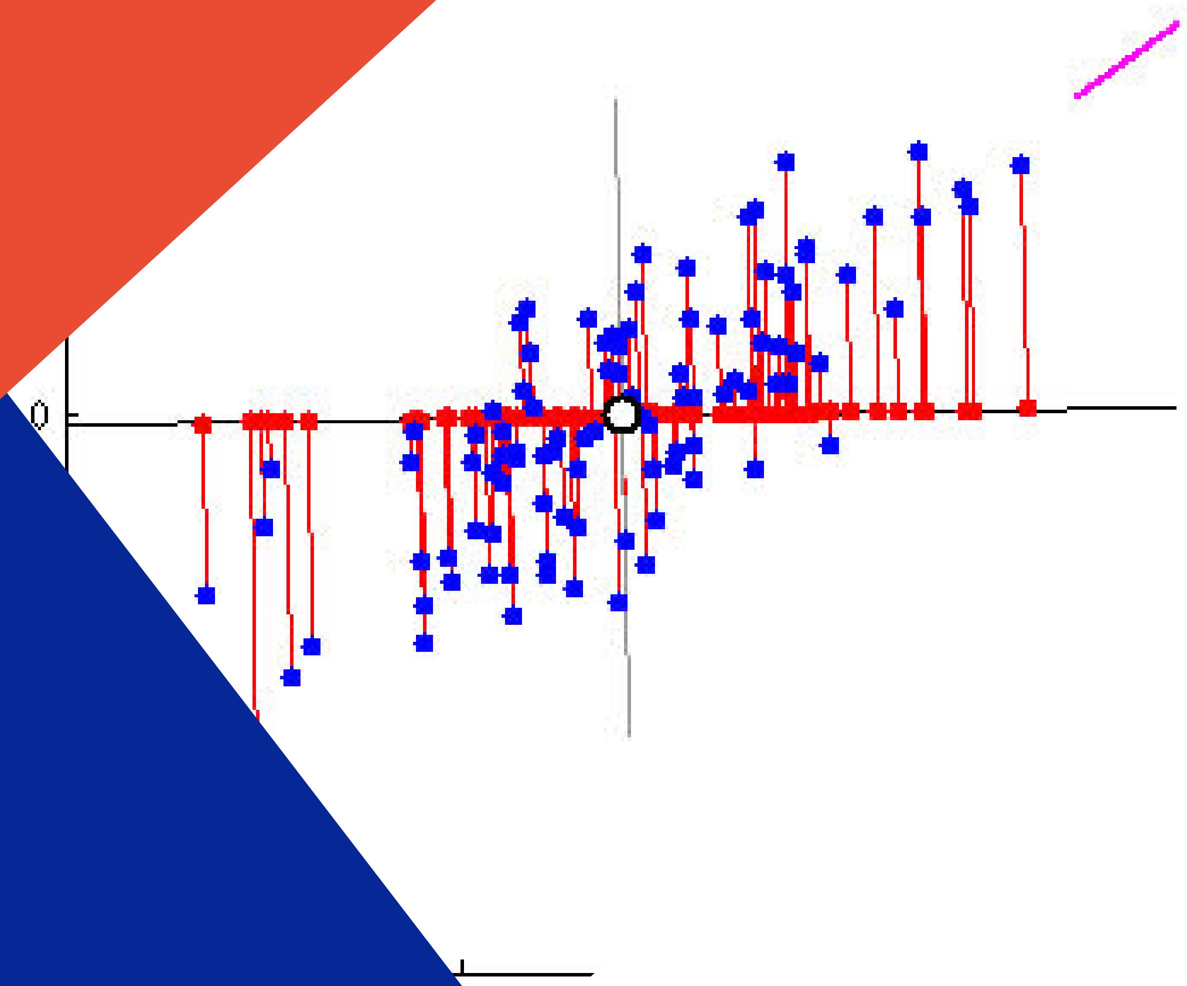


Analyse en composantes principales

Réalisé Par: Aroubite - Hrouch - Ghoundal - Rakib

Encadré par: M. Ghazdali

Sommaire

Présentation concernant la théorie générale de l'ACP

Objectif

L'analyse en composantes principales est une méthode de la famille de l'analyse des données et plus généralement de la statistique multivariée, dont l'objectif est de résumer un ensemble de données.

Objectif

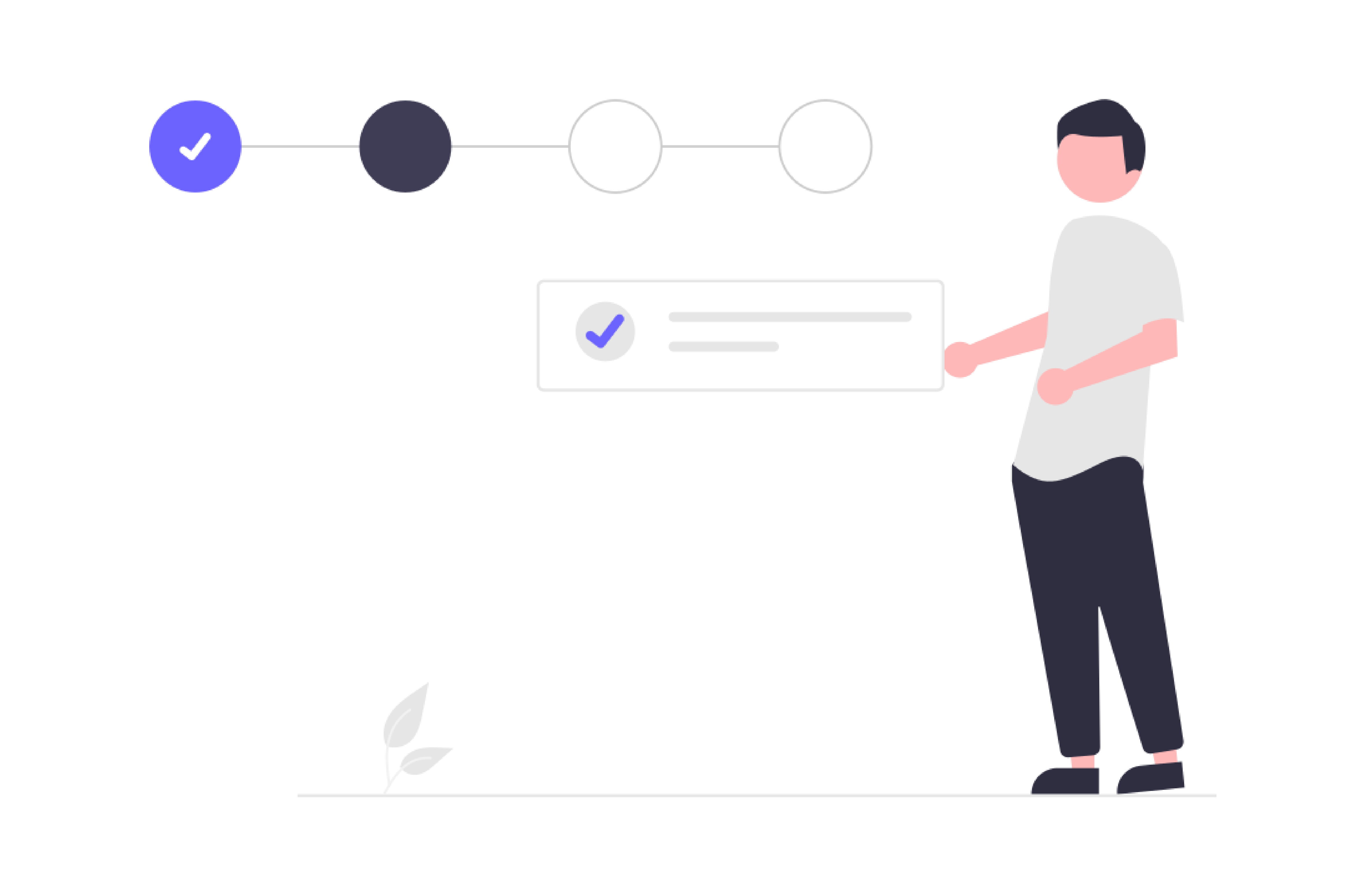
L'idée centrale de l'analyse en composantes principales (ACP) est de réduire la dimensionnalité d'un ensemble de données composé d'un grand nombre de variables interdépendantes tout en conservant autant que possible la variation présente dans l'ensemble de données.

Objectif

Ceci est réalisé en transformant en un nouvel ensemble de variables, les composantes principales (PC), qui ne sont pas corrélées et qui sont ordonnées de sorte que les premières conservent la majeure partie de la variation présente dans toutes les variables d'origine.

Les étapes

Pour déterminer les composantes principales



Les étapes de l'ACP

01

Tableau des données

02

Centrer les données

03

Matrice de covariance

04

Diagonalisation

Trouver les valeurs/vecteurs propres.

05

Définir les composantes principales

Choisir la dimension à retenir et determiner les axes principaux

01

Tableau des données

On considére les données suivantes:

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

01

Tableau des données

La matrice correspendante est:

$$\mathbf{A} = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

02

Calcul de la moyenne

$$\bar{A} = [\,66 \quad 60 \quad 60\,]$$

03

Matrice de covariance

Rappel:

La formule pour calculer la covariance:

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n} (Xi - \bar{x})(Yi - \bar{y})$$

03

Matrice de covariance

C'est quoi la matrice de covariance?

La matrice de covariance est une matrice symétrique p × p (où p est le nombre de dimensions) qui a comme entrées les covariances associées à toutes les paires possibles des variables initiales. Par exemple, pour un ensemble de données tridimensionnel avec 3 variables x, y et z, la matrice de covariance est une matrice 3 × 3 de ceci à partir de :

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

03

Matrice de covariance

Rappels:

Cov(a,a)=Var(a)

Cov(a,b)=Cov(b,a)

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix} \longleftrightarrow \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix}$$

03

Matrice de covariance

Pourquoi la matrice de covariance est util?

C'est en fait le signe de la covariance qui compte :

- si positif alors : les deux variables augmentent ou diminuent ensemble (corrélé)
- si négatif alors : L'un augmente quand l'autre diminue (Inversement corrélé)

03

Matrice de covariance

Calcul de la matrice de covariance:

$$\begin{array}{cc} & \begin{array}{ccc} Math & English & Arts \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} \end{array} \longrightarrow \begin{array}{cc} & \begin{array}{ccc} Math & English & Art \end{array} \\ \begin{array}{c} Math \\ English \\ Art \end{array} & \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix} \end{array}$$

04

Diagonalisation

Le but est de résoudre l'equation:

$$\det(A-\lambda I) = 0$$

$$\det\left(\begin{pmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

$$\det\begin{pmatrix} 504-\lambda & 360 & 180 \\ 360 & 360-\lambda & 0 \\ 180 & 0 & 720-\lambda \end{pmatrix}$$

04 Diagonalisation

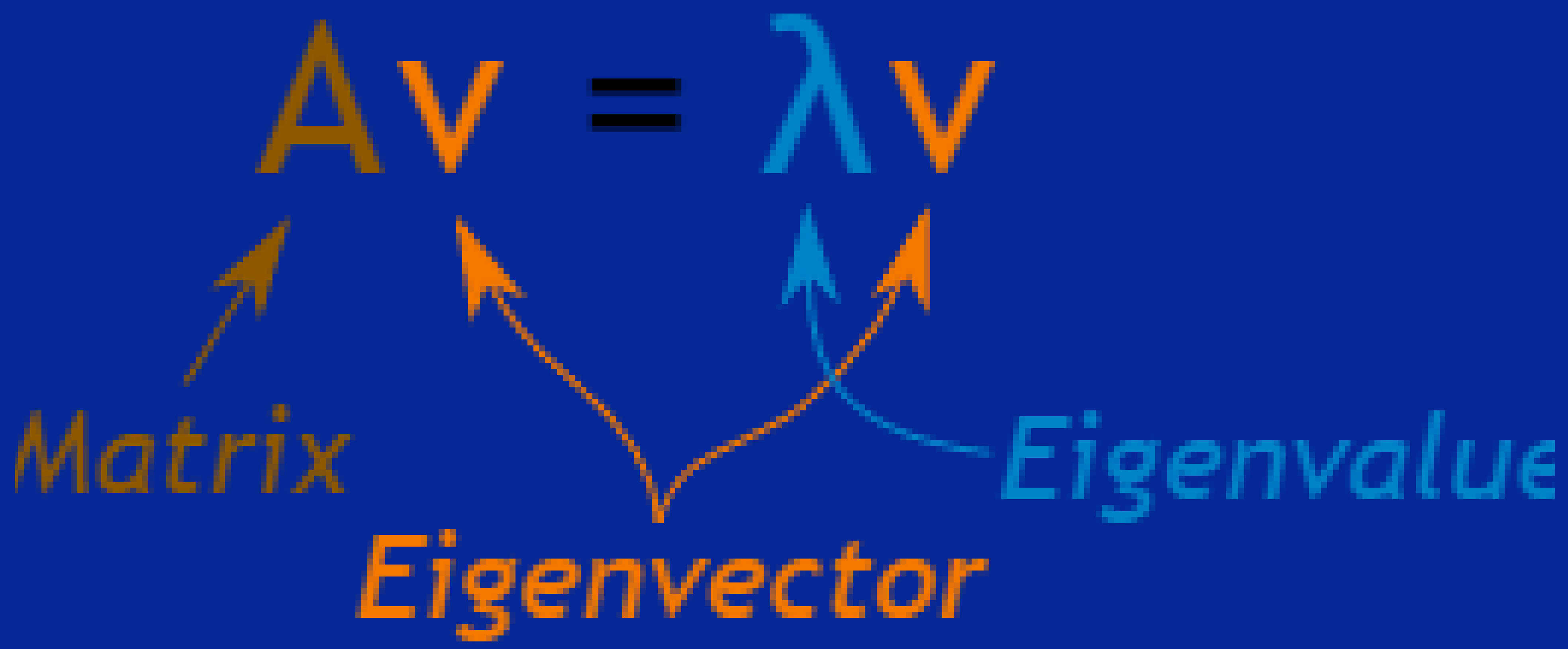
$$-\lambda^3 + 1584\lambda^2 - 641520\lambda + 25660800 = 0$$

$$\lambda \approx 44.81966\ldots, \lambda \approx 629.11039\ldots, \lambda \approx 910.06995\ldots$$

05

Diagonalisation

Maintenant on calcule les vecteurs propres, suivant la formules suivantes:



05

Diagonalisation

$$\begin{pmatrix} -3.75100\ldots \\ 4.28441\ldots \\ 1 \end{pmatrix}, \begin{pmatrix} -0.50494\ldots \\ -0.67548\ldots \\ 1 \end{pmatrix}, \begin{pmatrix} 1.05594\ldots \\ 0.69108\ldots \\ 1 \end{pmatrix}$$

On fait un tri décroissant des valeurs propres:

$$\lambda \approx 44.81966\ldots, \lambda \approx 629.11039\ldots, \lambda \approx 910.06995\ldots$$

$$\begin{pmatrix} 910.06995 \\ 629.11039 \\ 44.81966 \end{pmatrix}$$

Définir les composantes principales

$$\mathbf{W} = \begin{bmatrix} 1.05594 & -0.50494 \\ 0.69108 & -0.67548 \\ 1 & 1 \end{bmatrix}$$