Fondements du Machine Learning

Florentin Goyens

Licence 3 Informatique et Mathématiques pour la Décision et les Données

6 septembre 2022

Contact et liens utiles

Liens utiles

- Mon adresse email: florentin.goyens@dauphine.psl.eu
- Page du cours :
  https://flgoyens.github.io/teach/FML2022
- Notes de cours.
- Equipe sur Microsoft Teams

Organisation du cours

Programme

- Cours : 10 séances de 1h30. Attention, changement de programme !
  - Mardi 13h45-15h15;
  - pas de cours les mardi 20 septembre, 27 septembre et 1er novembre; dernier cours le mardi 29 novembre.
- TD : avec Lucas GNECCO HEREDIA
  - Mardi 15h30-17h00;
- Flexible

Évaluation

- Note=0.3*Controle Continu + 0.7*Examen;
- Contrôle continu : en presentiel ou devoir à la maison.
- Examen : En janvier.

Informations pratiques

Notes de cours (polycopié)

- Disponibles sur la page du cours;
- mises à jour de semaine en semaine;

Ne pas hésiter à poser des questions pendant le cours

- C'est bon pour vous;
- et pour moi aussi.

Objectifs du cours

- Présenter des modèles simples d'**analyse de données**...
- ...très fréquents en apprentissage machine *(machine learning)*.
- **Outils :** algèbre linéaire, probabilités/statistiques.

Terminologie

Différents domaines

- Analyse de données *(data analysis)*;
- Apprentissage [machine/automatique/artificiel] *(machine learning)*;
- Intelligence artificielle *(AI)*;
- Masses de données *(Big Data)*;
- etc.

Terminologie

Différents domaines

- Analyse de données *(data analysis)*;
- Apprentissage [machine/automatique/artificiel] *(machine learning)*;
- Intelligence artificielle *(AI)*;
- Masses de données *(Big Data)*;
- etc.

Sciences des données *(data science)*

Machine learning pour ce cours

Ensemble de **techniques** qui visent à extraire de l'information d'un jeu de **données et prendre des décisions de manière automatique**.

Principe 1: Grande quantité de données

- Pour avoir de l'information à extraire;
- Pour être statistiquement représentatif.

Principe 2: Utilisation d'algorithmes

- Traitement systématique et efficace;
- Théorie mathématique + implémentation.

Pourquoi s'intéresser aux données ?

Essentiel pour les entreprises

- Modèle économique des GAFA(M);
- Service gratuit mais valeur dans l'exploitation des données.

Important pour la recherche

- Quantité massive de données générées en biologie, médecine,...
- Difficultés mathématiques et informatiques.

Approches guidées par les données

Ou data-driven, drivées par les données, etc.

- Pallie le manque de modèles formels;
- Pourrait remplacer la modélisation de certains systèmes physiques à terme.

Exemples: systèmes de recommendation

But : Suggérer du contenu en se basant sur les préférences.

- Séries/Films/Vidéos (Netflix, Youtube);
- Produits commerciaux (Ebay, Amazon);
- Hôtels/Restaurants/Voyages (Google, Facebook).

Outil : Une matrice des avis (Clients/Produits).

Les grandes questions

Comment suggérer du contenu pertinent, et donc...

1. Quels sont les éléments principaux de nos préférences ?
2. Comment gérer un grand nombre d'avis ?
3. Les avis reflètent-ils vraiment la réalité ?

Quid du cours ?

Fondamentaux de l'apprentissage (*machine learning*)

- Des modèles linéaires de l'information dans les données;
- Les données sont vues comme des réalisations de variables aléatoires (typiquement gaussiennes).

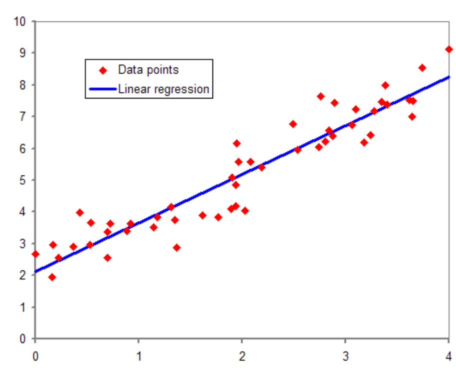
Modèles linéaires

- Souvent efficace en pratique;
- Très souvent le premier cas considéré en recherche;
- Utilise des savoirs fondamentaux en algèbre linéaire, statistiques (et optimisation).

Premier modèle d'apprentissage

- Distribution (de probabilité) des données connue;
- Apprentissage supervisé.

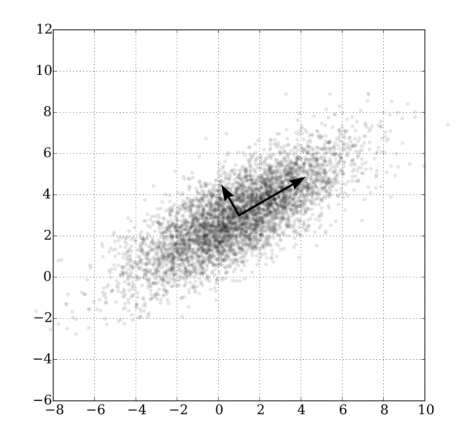
$\Rightarrow$ **Régression linéaire.**



Second modèle d'apprentissage

- Pas de distribution connue;
- Extraction d'information des données;
- Fréquent en apprentissage non supervisé.

$\Rightarrow$ **Analyse en composante principales.**



En bref

Machine learning/Apprentissage

- Décrire le comportement de données;
- Prédire les propriétés de données futures.

Notre approche

- Le cas linéaire : souvent efficace et populaire en pratique.
- Le modèle linéaire : permet de présenter les enjeux et les outils fondamentaux.

Table des matières

- Introduction et motivation (Chapitre 1);
- Algèbre linéaire et modèles linéaires :
  - Décomposition en valeurs propres/valeurs singulières (Chap. 2);
  - Moindres carrés linéaires (Chap. 3).
- Outils statistiques (Chapitre 4) :
  - Statistique multidimensionnelle;
  - Compromis biais-variance.
- **Régression linéaire** (Chapitre 5):
  - Simple, multiple;
  - Régularisation.
- Réduction de dimension (Chapitre 6):
  - **Analyse en composantes principales**;
  - Applications.

Sommaire

Sommaire

Les notations pour ce cours

Conventions

- Scalaires : $a, b, c, \ldots, \alpha, \beta, \gamma$;
- Vecteurs : $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \ldots, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$;
- Matrices : $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$.

Notations vectorielles

- $\mathbb{R}^n$: ensemble des vecteurs à $n \geq 1$ composantes réelles;
- Par convention, $\boldsymbol{x} \in \mathbb{R}^n$ est un vecteur colonne, et on note $\boldsymbol{x}^{\mathrm{T}}$ le vecteur ligne correspondant.
- Pour tout $\boldsymbol{x} \in \mathbb{R}^n$ et tout $i \in \{1, \dots, n\}$, on notera $x_i \in \mathbb{R}$ sa i-ème coordonnée (dans la base canonique de $\mathbb{R}^n$) $\Rightarrow$ $\boldsymbol{x} = [x_i]_{1 \leq i \leq n}$.

Notations vectorielles

- $\mathbb{R}^n$: ensemble des vecteurs à $n \geq 1$ composantes réelles;
- Par convention, $\boldsymbol{x} \in \mathbb{R}^n$ est un vecteur colonne, et on note $\boldsymbol{x}^\mathrm{T}$ le vecteur ligne correspondant.
- Pour tout $\boldsymbol{x} \in \mathbb{R}^n$ et tout $i \in \{1, \ldots, n\}$, on notera $x_i \in \mathbb{R}$ sa i-ème coordonnée (dans la base canonique de $\mathbb{R}^n$) $\Rightarrow$ $\boldsymbol{x} = [x_i]_{1 \leq i \leq n}$.

Structure d'espace vectoriel normé

- *Addition dans* $\mathbb{R}^n$ *:* $\boldsymbol{x} + \boldsymbol{y} := [x_i + y_i]_{1 \leq i \leq n}$;
- *Multiplication par un réel :*

$$\lambda \boldsymbol{x} := [\lambda x_i]_{1 \leq i \leq n} \;;$$

Notations vectorielles (suite)

Produit scalaire sur $\mathbb{R}^n$

Le produit scalaire est défini pour tous vecteurs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ par

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} := \sum_{i=1}^{n} x_i y_i.$$

C'est une forme bilinéaire symétrique définie positive (NB : $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$).

Norme Euclidienne d'un vecteur :

Pour tout $x \in \mathbb{R}^n$

$$\|\boldsymbol{x}\| := \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Notations vectorielles (suite)

Sous-espace engendré

Soient $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ p vecteurs de $\mathbb{R}^n$. Le *sous-espace engendré par les vecteurs* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ est le sous-espace vectoriel

$$\mathrm{range}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) := \left\{ \boldsymbol{x} = \sum_{i=1}^{p} \alpha_i \boldsymbol{x}_i \;\middle|\; \alpha_i \in \mathbb{R}\ \forall i \right\}.$$

Ce sous-espace est de dimension au plus $\min\{n, p\}$.

Definition d'une base de $\mathbb{R}^n$

Une base de $\mathbb{R}^n$ est une suite de vecteurs linéairement indépendants et générateurs de $\mathbb{R}^n$

Application linéaire

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ ensemble des matrices à m lignes, n colonnes et $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^m$.

Représentation de $\boldsymbol{A}$ en colonnes

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

ou en lignes

$$A = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix}.$$

Un produit matrice vecteur $\boldsymbol{Ax}$ est une combinaison linéaire des colonnes de $\boldsymbol{A}$

Noyau, image, rang

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

- Le **noyau** (*kernel/null space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\ker(\boldsymbol{A}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \quad \boldsymbol{A}\boldsymbol{x} = \mathbf{0}_{\mathbb{R}^m}\}\,.$$

- L'**image** (*range space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\mathrm{Im}(\boldsymbol{A}) := \{\boldsymbol{y} \in \mathbb{R}^m \mid \exists \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}\,.$$

- Le **rang** (*rank*) de $\boldsymbol{A}$, noté $\mathrm{rang}(\boldsymbol{A})$, est la dimension du sous-espace vectoriel $\mathrm{Im}(\boldsymbol{A})$. On a $\mathrm{rang}(\boldsymbol{A}) \leq \min\{m, n\}$.

Noyau, image, rang

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

- Le **noyau** (*kernel/null space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\ker(\boldsymbol{A}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}_{\mathbb{R}^m}\}.$$

- L'**image** (*range space*) de $\boldsymbol{A}$ est le sous-espace vectoriel

$$\mathrm{Im}(\boldsymbol{A}) := \{\boldsymbol{y} \in \mathbb{R}^m \mid \exists \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\}.$$

- Le **rang** (*rank*) de $\boldsymbol{A}$, noté $\mathrm{rang}(\boldsymbol{A})$, est la dimension du sous-espace vectoriel $\mathrm{Im}(\boldsymbol{A})$. On a $\mathrm{rang}(\boldsymbol{A}) \leq \min\{m, n\}$.

Théorème du rang

Pour toute matrice $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, on a

$$\dim(\ker(\boldsymbol{A})) + \mathrm{rang}(\boldsymbol{A}) = n.$$

Normes matricielles

Norme opérateur

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{Ax}\|$$

Norme de Frobénius

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} = \text{trace}(A^\top A)$$

Transposée, symétrie

Définitions

Soit $\boldsymbol{A} = [\boldsymbol{A}_{ij}] \in \mathbb{R}^{m \times n}$ une matrice à m lignes et n colonnes.

- La *transposée de* $\boldsymbol{A}$, notée $\boldsymbol{A}^{\mathrm{T}}$, est la matrice à n lignes et m colonnes telle que

$$\forall i = 1, \ldots, m, \ \forall j = 1, \ldots, n, \qquad \left[\boldsymbol{A}^{\mathrm{T}}\right]_{ij} = \boldsymbol{A}_{ij}.$$

Transposée, symétrie

Définitions

Soit $\boldsymbol{A} = [\boldsymbol{A}_{ij}] \in \mathbb{R}^{m\times n}$ une matrice à m lignes et n colonnes.

- La *transposée de* $\boldsymbol{A}$, notée $\boldsymbol{A}^{\mathrm{T}}$, est la matrice à n lignes et m colonnes telle que

$$\forall i = 1, \ldots, m,\ \forall j = 1, \ldots, n, \qquad \left[\boldsymbol{A}^{\mathrm{T}}\right]_{ij} = \boldsymbol{A}_{ij}.$$

Cas des matrices carrées

- $\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{n\times n}$;
- $\boldsymbol{A}$ est dite *symétrique* si $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.

Ex) Matrices diagonales, de covariance, d'adjacence, etc.

Matrices inversibles, définies, positives

Une matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ est dite *inversible* s'il existe $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ telle que $\boldsymbol{BA} = \boldsymbol{AB} = \mathbf{I}_n$, où $\mathbf{I}_n$ est la matrice identité de $\mathbb{R}^{n \times n}$.
La matrice $\boldsymbol{B}$ est alors unique : elle est appelée l'*inverse de* $\boldsymbol{A}$ et se note $\boldsymbol{A}^{-1}$.

Une matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ est dite *semi-définie positive* si

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \geq 0.$$

Elle est dite *définie positive* lorsque $\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} > 0$ pour tout vecteur $\boldsymbol{x}$ non nul.

Matrices orthogonales

Une matrice carée $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ tel que $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}_n$ est dite orthogonale.
Une matrice $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ avec $m \geq n$, est dite orthogonale si ces colones sont des vecteurs orthonormés, $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_n$

Une transformation orthogonale conserve la norme

Soit $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ et deux matrices orthogonales $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ et $\boldsymbol{V} \in \mathbb{R}^{n\times n}$. Alors

$$\|\boldsymbol{A}\| = \|\boldsymbol{U}\boldsymbol{A}\| = \|\boldsymbol{A}\boldsymbol{V}\|$$

et

$$\|\boldsymbol{A}\|_F = \|\boldsymbol{U}\boldsymbol{A}\|_F = \|\boldsymbol{A}\boldsymbol{V}\|_F$$

Sommaire

Valeurs propres et vecteurs propres

Definition

Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. On dit que $\lambda \in \mathbb{R}$ est une *valeur propre de* $\boldsymbol{A}$ si

$$\exists \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{v} \neq \vec{0}, \qquad \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

Le vecteur $\boldsymbol{v}$ est alors un *vecteur propre associé à la valeur propre* λ.
L'ensemble des valeurs propres de $\boldsymbol{A}$ s'appelle le *spectre*.

Valeurs propres et vecteurs propres

Definition

Soit $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. On dit que $\lambda \in \mathbb{R}$ est une *valeur propre de* $\boldsymbol{A}$ si

$$\exists \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{v} \neq \vec{0}, \qquad \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

Le vecteur $\boldsymbol{v}$ est alors un *vecteur propre associé à la valeur propre* λ. L'ensemble des valeurs propres de $\boldsymbol{A}$ s'appelle le *spectre*.

- Toute matrice de $\mathbb{R}^{n \times n}$ possède n valeurs propres complexes, mais pas nécessairement n valeurs propres réelles.
- Les valeurs propres réelles d'une matrice semi-définie positive (resp. définie positive) sont positives (resp. strictement positives).
- Le noyau de $\boldsymbol{A}$ est engendré par les vecteurs propres associés à la valeur propre 0.

Décomposition en valeurs propres

Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ symétrique admet une décomposition dite **spectrale** de la forme :

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1},$$

avec

- $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ matrice orthogonale ($\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$), dont les colonnes $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ forment une *base orthonormée* de vecteurs propres.
- $\boldsymbol{\Lambda}$ matrice diagonale, qui contient les n valeurs propres de $\boldsymbol{A}$ $\lambda_1, \ldots, \lambda_n$ sur la diagonale.

Décomposition en valeurs propres

Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ symétrique admet une décomposition dite **spectrale** de la forme :

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1},$$

avec

- $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ matrice orthogonale ($\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$), dont les colonnes $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ forment une *base orthonormée* de vecteurs propres.
- $\boldsymbol{\Lambda}$ matrice diagonale, qui contient les n valeurs propres de $\boldsymbol{A}$ $\lambda_1, \ldots, \lambda_n$ sur la diagonale.

- Il n'y a pas unicité de la décomposition spectrale;
- Aux permutations près, l'ensemble des valeurs propres est unique.

Sur la décomposition spectrale

Chaque valeur propre λ_i caractérise l'effet de $\boldsymbol{A}$ sur le vecteur $\boldsymbol{p}_i$:

- $|\lambda_i| >> 1 \Rightarrow \|\boldsymbol{A}\boldsymbol{p}_i\| >> \|\boldsymbol{p}_i\|$;
- $|\lambda_i| << 1 \Rightarrow \|\boldsymbol{A}\boldsymbol{p}_i\| << \|\boldsymbol{p}_i\|$.

Interprétation géométrique

L'action de la matrice $\boldsymbol{A}$ sur un vecteur $\boldsymbol{x}$

- Allonge les composantes de $\boldsymbol{x}$ dans la base des vecteurs propres associées aux plus grandes valeurs propres en valeur absolue;
- Réduit les composantes de $\boldsymbol{x}$ associées aux valeurs propres les plus faibles en valeur absolue;
- Cas extrême : si $\ker(\boldsymbol{A}) \neq \{\mathbf{0}\}$, toute composante de $\boldsymbol{x}$ selon un vecteur du noyau est réduite à zéro par $\boldsymbol{A}$.

Diagonalisation comme changement de base

Interpretation comme un changement de base

La decomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}$ exprime l'application $\boldsymbol{A}$ dans la base $p = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n)$, où $\boldsymbol{\Lambda}$ est diagonale.

En effet, $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{b}$ signifie que $(\boldsymbol{b})_e = (\boldsymbol{x})_p$. De même on a $\boldsymbol{x} = \boldsymbol{P}^{-1}\boldsymbol{b}$ qui signifie que $\boldsymbol{P}^{-1}\boldsymbol{b}$ represente les coefficients du vecteur $\boldsymbol{b}$ dans la base $\boldsymbol{p}$.

$$P = {}_e(\boldsymbol{I})_p$$

et

$$P^{-1} = {}_p(\boldsymbol{I})_e$$

Donc

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} = {}_e(\boldsymbol{I})_p\, {}_p(\boldsymbol{A})_p\, {}_p(\boldsymbol{I})_e$$

Sommaire

Le cas des matrices rectangulaires

Notion de valeur propre

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. On peut parler :

- des valeurs propres de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \in \mathbb{R}^{n \times n}$;
- des valeurs propres de $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{m \times m}$.

Peut-on s'en servir pour obtenir une décomposition de $\boldsymbol{A}$?

Le cas des matrices rectangulaires

Notion de valeur propre

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. On peut parler :

- des valeurs propres de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \in \mathbb{R}^{n \times n}$;
- des valeurs propres de $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \in \mathbb{R}^{m \times m}$.

Peut-on s'en servir pour obtenir une décomposition de $\boldsymbol{A}$?

Observations concernant $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$

- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ est symétrique;
- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ est semi-définie positive;
- $\ker(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \ker(\boldsymbol{A})$;
- $\mathrm{rang}(\boldsymbol{A}^{T}\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A})$;
- $\mathrm{Im}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{Im}(\boldsymbol{A}^{\mathrm{T}})$;.

(Résultats similaires pour $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.)

Décomposition en valeurs singulières

Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ admet une **décomposition en valeurs singulières** (*SVD*) de la forme

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}.$$

- $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ est orthogonale;
- $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ est orthogonale;
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ est diagonale par blocs, avec des coefficients nuls sauf ceux de la diagonale $\{[\boldsymbol{\Sigma}]_{ii}\}_i$ qui sont positifs (ou nuls). Ces éléments, notés $\{\sigma_i\}$, s'appellent les **valeurs singulières de $\boldsymbol{A}$**.

Décomposition en valeurs singulières

Théorème

Toute matrice $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ admet une **décomposition en valeurs singulières** (*SVD*) de la forme

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}.$$

- $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ est orthogonale;
- $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ est orthogonale;
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ est diagonale par blocs, avec des coefficients nuls sauf ceux de la diagonale $\{[\boldsymbol{\Sigma}]_{ii}\}_i$ qui sont positifs (ou nuls). Ces éléments, notés $\{\sigma_i\}$, s'appellent les **valeurs singulières de $\boldsymbol{A}$**.

- $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} \Leftrightarrow \boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{\Sigma}$;
- Une *SVD* n'est pas définie de façon unique mais ses valeurs singulières le sont.

Pour la suite

Cours

- La décomposition en valeurs singulières (SVD) : Preuve constructive;
- outil fondamental;
- .

TD

- Quelques exercices d'algèbre linéaire;
- Applications de la SVD.

Pour finir...

Exercice (I-1 polycopié)

Soit $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Démontrer les propriétés suivantes sans utiliser la décomposition en valeurs singulières :

a) $\ker(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \ker(\boldsymbol{A})$;

b) $\mathrm{rang}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rang}(\boldsymbol{A})$;

c) $\mathrm{rang}(\boldsymbol{A}^{T}\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A})$, et en déduire que $\mathrm{Im}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{Im}(\boldsymbol{A}^{\mathrm{T}})$.

Comment les démontrer en utilisant maintenant la SVD ?