Modélisation de Données Complexes

Matthieu Boyer

Cours TalENS 1 2025-2026

Table des matières

Ce polycopié ne doit pas être vu comme un remplacement au cours, mais simplement comme un résumé du contenu qui permet d'être sûr de ne rien manquer. On ne parlera ici des espaces vectoriels que dans un cadre très restreint adapté à ce qu'on va vouloir en faire. En particulier, on ne fera pas d'algèbre linéaire, on ne parlera pas de base quelconque ni d'isomorphismes. Je vous renvoie au cours donné avec Clément Allard l'an dernier pour plus de détails et des preuves.

1 Calcul vectoriel

1.1 Espaces euclidiens

Autour de 300 avant notre ère, Euclide, mathématicien grec écrit les Éléments, un traité de géométrie qui construit le monde autour de 5 axiomes (postulats indémontrables par les autres), sous forme de constructions géométriques réalisables :

1. On peut tracer une ligne droite de tout point à tout autre point ;
2. On peut étendre une ligne droite finie continuement en une ligne droite ;

3. On peut décrire un cercle de tout centre avec tout rayon ;
4. On suppose que tous les angles droits sont égaux ;
5. On suppose qu'étant donnée une droite et un point n'appartenant pas à la droite, il y a exactement une droite qui passe par ce point et qui ne croise pas la droite de départ.

Les espaces vectoriels euclidiens sont une manière algébrique de décrire un espace vérifiant les axiomes d'Euclide.

Définition 1.1 Un espace vectoriel (réel) est un ensemble E de points (appelés vecteurs) munis d'une addition interne associative commutative unitale et d'une multiplication externe distributive sur l'addition.

Autrement dit, dans l'espace E :

- Il existe un vecteur origine ou vecteur nul, dénoté 0_E ou simplement 0 ;
- L'addition de deux vecteurs se fait par la relation de Chasles [1] ;
- La multiplication par $\lambda \in \mathbb{R}$ d'un vecteur x est l'agrandissement/le rétrécissement de x par un facteur λ.

Il n'y a pas de "points" dans E, mais on peut les construire en voyant un point P comme le vecteur partant de 0 et allant jusqu'à P.

R Il est important de noter que cette définition ne présume rien sur le nombre de vecteurs, ni sur ce qu'ils représentent : pour tout ensemble X, l'ensemble des fonctions de X dans $\mathbb{R}$ est un espace vectoriel (muni de l'addition point à point et de la multiplication du résultat). On va beaucoup, dans la suite, faire appel à l'intuition géométrique des espaces vectoriels, mais il nous est utile d'avoir une définition très générale.

Définition 1.2 Une base d'un espace vectoriel E est une famille $(e_i)_{i\in I}$ d'éléments de E telle que tout vecteur x de E s'écrit de manière unique comme une somme finie de e_i. On dit que E est de dimension finie si E possède une base finie.

Tous les espaces vectoriels possèdent une base, si l'on accepte l'axiome du choix (ou plutôt le lemme de Zorn, équivalent). Tous les espaces vectoriels ne possèdent pas de base finie, comme c'est par exemple le cas de l'espace des fonctions. Toutes les bases d'un espace vectoriel ont même cardinal, appelé dimension de l'espace.

Définition 1.3 L'espace engendré par une famille (e_i) est l'ensemble $\mathrm{Vect}(e_i)$ des combinaisons linéaires des (e_i).

R Être une base, ce n'est pas juste engendrer E (dans ce cas, la famille est génératrice), c'est engendrer E sans avoir de degré de liberté superflu. Cela revient à dire qu'il n'existe pas d'indice $k \in I$ tel que $e_k \in \mathrm{Vect}(e_i)_{i\in I\setminus\{k\}}$, et dans ce cas là, on dit que la famille est libre. Dans la suite, on ne considèrera que des familles libres (sauf mention contraire).

L'espace engendré par un vecteur est une droite, celui engendré par 2 vecteurs est un plan et celui engendré par $n-1$ vecteurs dans un espace de dimension n est appelé un hyperplan.

Définition 1.4 Une application $\varphi : E \to F$ est dite linéaire si pour tout $x, y \in E$, pour tout $\lambda \in \mathbb{R}$: $\varphi(\lambda x + y) = \lambda\varphi(x) + \varphi(y)$. Une application $\psi : E \times F \to G$ est dite bilinéaire si pour tout $x \in E$ sa restriction à F est linéaire (et de même pour tout $y \in F$).

Définition 1.5 Un produit scalaire sur E est une application bilinéaire symmétrique définie positive de $E \times E$ dans $\mathbb{R}$. Sa norme engendrée est l'application $\|\cdot\| : x \mapsto \sqrt{\langle x, x\rangle}$

Symmétrique signifie que $\langle x, y\rangle = \langle y, x\rangle$, positive signifie que $\langle x, x\rangle \geq 0$ et définie signifie que $\langle x, x\rangle = 0 \Rightarrow x = 0$.

1. Qui est un théorème de géométrie affine et non vectorielle.

Définition 1.6 Deux vecteurs sont orthogonaux si leur produit scalaire est nul. Une base est orthonormale si tous ses vecteurs sont orthogonaux deux à deux et sont de norme 1.

Tout espace vectoriel admet une base orthonormale.

Proposition 1.1 Dans le cas où E admet une base orthonormée (e_i), si $x = \sum x_i e_i$ et $y = \sum y_i e_i$ alors :

$$\langle x, y \rangle = \sum x_i y_i$$

Toutes les définitions ci-dessus sont une extension simple de ce qui a été vu en cours pour le produit scalaire en dimension 2 et 3.

Théorème 1.1 — Pythagore Si la famille des u_i est orthogonale :

$$\left\| \sum u_i \right\|^2 = \sum \|u_i\|^2$$

Lorsque notre espace vectoriel euclidien est de dimension finie, on a le résultat suivant :

Définition 1.7 Si E est un espace euclidien de dimension d, alors E est isomorphe à $\mathbb{R}^d$. Autrement dit, tout calcul fait dans E est équivalent à un calcul fait dans $\mathbb{R}^d$.

Ce résutlat est important car dans la suite, nous ne travaillerons que sur des espaces dont nous pouvons connaître la structure : ou bien ils ont des propriétés particulières (espaces de fonctions, de suites, etc...) ou bien ils sont $\mathbb{R}^d$ pour un certain d. Nous pouvons donc nous limiter à définir ce que nous voulons sur $\mathbb{R}^d$, en offrant possiblement un sens dérivé de l'espace de départ à nos applications sur $\mathbb{R}^d$.

Il existe une base dite canonique de $\mathbb{R}^d$, notée $e_1, \ldots, e_d$ et orthonormale.

1.2 Applications sur les espaces euclidiens

Dans la suite, toutes les fonctions sur lesquelles nous travaillerons seront en dimension finie. On va chercher à généraliser rapidement la définition de dérivée :

Définition 1.8 La dérivée directionnelle de $f : \mathbb{R}^n \to \mathbb{R}$ selon $v \in \mathbb{R}^n$ en $x \in \mathbb{R}^n$ est donnée par :

$$D_v(f)(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} \in \mathbb{R}$$

Définition 1.9 Le gradient de f en x est le vecteur $\nabla(f)(x)$ des dérivées directionnelles de f en x selon chacun des e_i :

$$\nabla(f)(x) = \begin{pmatrix} D_{e_1}(f)(x) \\ \vdots \\ D_{e_d}(f)(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{pmatrix}$$

C'est un vecteur de $\mathbb{R}^d$.

Le gradient est le vecteur qui indique la direction de plus forte progression de la fonction.

Proposition 1.2 L'opérateur $\nabla(\cdot)(x)$ est linéaire.

Cette proposition indique qu'en effet, l'opérateur ∇ désigne bien une sorte de dérivée. D'ailleurs, lorsque $d = 1$, c'est

exactement la dérivée dont vous avez l'habitude. Voici juste un exemple utile :

$$\nabla\left(\|\cdot\|^2\right)(x) = 2x \quad \nabla(\|\cdot\|)(x) = \nabla\left(\|\cdot\|\right)(x) \times \frac{1}{2\|x\|} = \frac{x}{\|x\|}$$

Attention, le gradient n'existe pas toujours, comme le montre l'exemple ci-dessus, qui n'a pas de gradient en 0. L'expression pour le gradient de la norme découle de la règle de la chaîne : $\frac{\partial f\circ g}{\partial x} = \frac{\partial g}{\partial x} \times f'(g(x))$ où $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R}^d \to \mathbb{R}$. Il y a une forme plus générale de la règle, mais elle n'est pas nécessaire pour comprendre la suite.

Définition 1.10 La Hessienne de f en x est en quelque sorte la dérivée seconde de f, c'est la matrice (vecteur 2-dimensionnel ou tableau) :

$$H(f)(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1 \partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_d \partial x_1}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_1 \partial x_d}(x) & \cdots & \frac{\partial f}{\partial x_d \partial x_d}(x) \end{pmatrix} = \left(\frac{\partial f}{\partial x_i \partial x_j}(x)\right)_{i,j}$$

où $\frac{\partial f}{\partial x_i \partial x_j}$ désigne la dérivée selon x_j de la fonction dérivée selon x_i de f.

Théorème 1.2 Comme sur $\mathbb{R}$, un point x est un extremum local de f si $\nabla(f)(x) = 0$. De même, un point x est un minimum local si $H(f)(x)$ est définie positve.

2 Plongements euclidiens

On s'intéresse maintenant au problème de la construction d'ensembles de données sur lesquelles travailler.

2.1 Plongements

Il n'y a pas de définition formelle de plongement qui n'est pas spécifiique à un type de plongement. En général cependant, un plongement d'une structure A dans une structure B est une forme de morphisme injectif (ou monomorphisme).

Définition 2.1 Un plongement ι de X dans un espace euclidien $\mathbb{R}^F$ dit espace de représentation ou espace de caractéristiques (feature space), est une application injective qui vérifie l'heuristique suivante : Étant donnés deux objets a, b de X, les deux vecteurs $\iota(a), \iota(b)$ de l'espace euclidien vérifient, pour toute fonction qu'on souhaite étudier :

$$f(\iota(a), \iota(b)) = \Theta(f(a, b))$$

où Θ désigne la notation de Landau pour l'encadrement à constante près ($Cf(a, b) \leq f(\iota(a), \iota(b)) \leq C'f(a, b)$)

R C'est une forme d'application continue injective pour la topologie engendrée par tiré en arrière des métriques que nous allons étudier sur l'espace de représentation. Si cette phrase ne veut rien dire pour vous, c'est normal pour l'instant, nous reviendrons sur ça lors d'un cours ultérieur.

L'acte de plonger ses données de départ dans un espace euclidien n'est pas à prendre à la légère, c'est lui qui engendre la plupart des propriétés que nous pouvons étudier : D'une part on va chercher à reduire au maximum la dimension F de l'espace de caractéristiques pour limiter les besoins en mémoire et en temps des calculs ; D'autre part, si le plongement n'est pas fidèle aux données de départ, et n'est pas réellement injectif (par exemple si on réduit trop la dimension en ne se basant que sur un échantillon fini d'un ensemble réellement infini de données).

2.2 Zoologie de métriques

Dans cette partie on va regarder rapidement quelques métriques intéressantes sur des ensembles de points dans un espace vectoriel E. Il faut toujours se souvenir que définir une métrique sur un espace vectoriel doit se faire en ayant en tête l'ensemble qui a été plongé dans l'espace vectoriel et ce que la métrique peut nous apprendre sur l'espace.

2.2.1 Similarité angulaire

On va chercher à mesurer l'angle entre deux droites engendrées par u et v.

Proposition 2.1 — Inégalité de Cauchy-Schwarz. Si u, v sont des éléments d'un espace euclidien alors :

$$|\langle u, v\rangle| \leq \|u\| \, \|v\|$$

avec égalité si et seulement si $u = \lambda v$ avec $\lambda \geq 0$.

En particulier cela signifie que

$$\frac{\langle u, v\rangle}{\|u\| \, \|v\|} \in [-1, 1]$$

et donc que $\arccos\left(\frac{\langle u,v\rangle}{\|u\|\|v\|}\right)$ est bien une mesure d'angle entre les deux droites.

2.2.2 Norme L-p

Définition 2.2 Si $p \geq 1$ et $x \in E$.

$$\|x\|_p = \sqrt[p]{\sum x_i^p}$$

Pour tout p, on obtient une distance entre nos points, par $d_p(x, y) = \|x - y\|_p$. Pour $p = \infty$ on pose $\|x\|_\infty = \max |x_i|$.

2.2.3 Représentation fréquentielle

En divisant x par la somme des x_i, on construit une distribution de probabilité à partir de x. Ceci nous permet d'utiliser tout un tas de métriques sur l'ensemble des distributions de probabilité, notamment les distances de Wasserstein, la divergence de Kullback-Leibler ou la métrique de Fisher-Rao.

3 Algorithmes sur les immersions

3.1 PCA

Le but de l'algorithme PCA est de trouver une meilleure base de l'espace pour représenter nos données. Pour ce faire on va construire une base dont chacune des composantes augmente le plus possible l'espace précédent. Ceci nous permet d'ailleurs de construire un espace de plongement de plus faible dimension, au coût de la signification des caractéristiques.

procedure PCA(n, X)
  for $i = 1 \to n$ **do**
    $u_i \leftarrow \text{argmin}_{u \in \text{Vect}(u_j)_{j<i}^{\perp}} \sum_{x \in X} \langle x, u\rangle$
end procedure

PCA est facile à implémenter et préserve la variance globale des données, en restant linéaire et peu coûteux mais fonctionne assez mal dans le cas où les données ne sont pas du tout linéaire.

3.2 t-SNE

Pour t-SNE, l'algorithme est plus complexe et moins interprétable, mais préserve la similitude locale de l'ensemble. On plonge les données dans un ensemble de probabilités conditionnels représentant des similitudes pour une distribution gaussienne. On peut ensuite mesurer la différence entre une distribution apprise (en 2 ou 3 dimensions) et cette distribution, par la divergence de Kullback-Leibler. La divergence de Kullback-Leibler est simplement une mesure de la probabilité que deux distributions de probabilité soit fortement différentes.