

# Géométrie des Données et Apprentissage Machine

## Lecture 1 – Introduction et Notions Fondamentales

Youssef MESRI - MINES Paris - PSL

November 23, 2025

# Plan

- 1 Lecture 1 : Introduction à la géométrie des données
- 2 Motivations
- 3 Concentration des distances : preuve par récurrence
- 4 Notions fondamentales
- 5 Conclusion
- 6 Réduction de dimension

# Lecture 1 : plan

- Notion de données en haute dimension
- Malédiction de la dimension et concentration de distances
- Variétés et distances géodésiques
- Réduction de dimension
  - PCA
  - Isomap
  - Diffusion Maps
  - t-SNE
- Labs

# Pourquoi une géométrie des données ?

- Données modernes : **images, sons, textes** → espaces  $\mathbb{R}^d$  avec  $d \gg 1$ .
- Pourtant, elles possèdent souvent une **structure intrinsèque** de faible dimension.
- **Idée clé** : les données sont souvent concentrées sur une **variété** de dimension intrinsèque  $k \ll d$ .
- **Exemple** : images de chiffres manuscrits (784 dimensions) mais proches d'une variété de dimension *beaucoup* plus petite  $\approx 10$ .

MNIST

7 4 3 1 0 8 3 4  
7 2 4 9 4 8 6 5  
6 9 0 3 1 5 8 1  
0 0 8 3 4 9 8 0  
5 3 8 5 0 1 0 5  
4 6 4 3 6 6 1 7  
7 3 9 9 6 9 5 0

# Explosion dimensionnelle : malédiction de la dimension

- Lorsque la dimension  $d$  augmente, le volume de l'espace croît **exponentiellement**.
- Dans un cube unité  $[0, 1]^d$ , la majeure partie du volume se concentre près des **bords**.
- Les distances deviennent moins discriminantes :

## Illustration mathématique

Soient  $n$  points tirés uniformément dans  $[0, 1]^d$ . On définit :

$$d_{\min} = \min_{i \neq j} \|x_i - x_j\|,$$

$$d_{\max} = \max_{i \neq j} \|x_i - x_j\|.$$

On observe que :

$$\lim_{d \rightarrow \infty} \frac{d_{\max} - d_{\min}}{d_{\min}} \rightarrow 0.$$

**Conséquence** : toutes les distances deviennent presque égales !

# Mise en place

Deux points  $x, y \in [0, 1]^d$  i.i.d. uniformes. Définir la distance au carré :

$$S_d := \|x - y\|_2^2 = \sum_{i=1}^d (x_i - y_i)^2 = \sum_{i=1}^d X_i,$$

où  $X_i$  sont i.i.d. copies de  $X := (U - V)^2$  avec  $U, V \sim \mathcal{U}[0, 1]$  indépendants. On montrera :  $\mathbb{E}[S_d] = d\mu$ ,  $\text{Var}(S_d) = d\sigma^2$  et la dispersion relative  $\rightarrow 0$ .

## Cas $d = 1$ : loi de la différence et moments

Posons  $Z := U - V$ . Alors  $Z \sim$  loi triangulaire sur  $[-1, 1]$  de densité  $f_Z(z) = 1 - |z|$ .

Comme  $X = Z^2$  :

$$\mathbb{E}[X] = \int_{-1}^1 z^2(1 - |z|) dz = 2 \int_0^1 z^2(1 - z) dz = 2\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{1}{6},$$

$$\mathbb{E}[X^2] = \int_{-1}^1 z^4(1 - |z|) dz = 2 \int_0^1 z^4(1 - z) dz = 2\left(\frac{1}{5} - \frac{1}{6}\right) = \frac{1}{15}.$$

Donc  $\mu = \mathbb{E}[X] = \frac{1}{6}$  et  $\sigma^2 = \text{Var}(X) = \frac{1}{15} - \frac{1}{36} = \frac{7}{180}$ .

## Vérification alternative (moments de l'uniforme)

Moments utiles pour  $W \sim \mathcal{U}[0, 1]$  :

$\mathbb{E}[W] = \frac{1}{2}$ ,  $\mathbb{E}[W^2] = \frac{1}{3}$ ,  $\mathbb{E}[W^3] = \frac{1}{4}$ ,  $\mathbb{E}[W^4] = \frac{1}{5}$ . Avec indépendance de  $U, V$  :

$$\mathbb{E}[(U - V)^4] = 2\mathbb{E}[U^4] - 4\mathbb{E}[U^3]\mathbb{E}[V] + 6\mathbb{E}[U^2]\mathbb{E}[V^2] - 4\mathbb{E}[U]\mathbb{E}[V^3] = \frac{1}{15}.$$

On retrouve donc  $\mathbb{E}[X^2] = \frac{1}{15}$  et  $\text{Var}(X) = \frac{7}{180}$ .

## Étape de récurrence (somme de i.i.d.)

Supposer vrai pour  $d$  :  $\mathbb{E}[S_d] = d\mu$ ,  $\text{Var}(S_d) = d\sigma^2$ . Pour  $d + 1$  :

$$S_{d+1} = S_d + X_{d+1} \Rightarrow \mathbb{E}[S_{d+1}] = (d+1)\mu, \quad \text{Var}(S_{d+1}) = (d+1)\sigma^2,$$

par indépendance. Par récurrence simple, les formules valent pour tout  $d \geq 1$ .

# Concentration relative et Chebyshev

Coefficient de variation :

$$\frac{\sqrt{\text{Var}(S_d)}}{\mathbb{E}[S_d]} = \frac{\sqrt{d\sigma^2}}{d\mu} = \frac{C}{\sqrt{d}}, \quad C = \frac{\sqrt{\sigma^2}}{\mu} = 6\sqrt{\frac{7}{180}}.$$

**Chebyshev** : pour tout  $\varepsilon > 0$ ,

$$\Pr(|S_d - \mathbb{E}[S_d]| \geq \varepsilon \mathbb{E}[S_d]) \leq \frac{\sigma^2}{\varepsilon^2 \mu^2} \cdot \frac{1}{d} \xrightarrow{d \rightarrow \infty} 0.$$

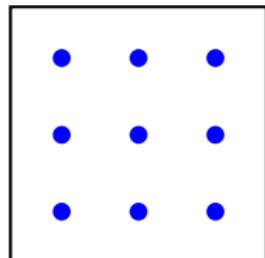
Donc **concentration relative** de  $S_d$  autour de sa moyenne.

# Interprétation géométrique

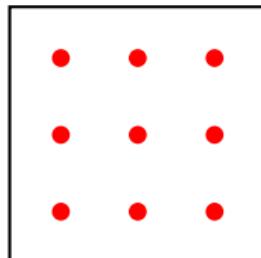
- $\mathbb{E}[S_d] = d/6$  croît linéairement en  $d$  (donc  $\mathbb{E}[\|x - y\|_2] \asymp \sqrt{d}$ ).
- L'écart-type est  $\asymp \sqrt{d}$  aussi, mais **relativement** à la moyenne il vaut  $O(1/\sqrt{d})$ .
- En grande dimension, les distances se ressemblent **relativement** : perte de pouvoir discriminant.

## Exemple numérique : distances en haute dimension

- Simulation : on génère  $10^4$  points dans  $[0, 1]^d$ .
- On calcule la distance moyenne  $\bar{d}$  et son écart-type  $\sigma_d$ .
- Résultat :  $\sigma_d/\bar{d} \rightarrow 0$  quand  $d \rightarrow \infty$ .



2D



3D (projection)

Any illustration : concentration des distances quand  $d$  augmente (ici 2D et projection 3D).

# Conséquences pratiques

- Les méthodes classiques basées sur des distances euclidiennes deviennent inefficaces.
- Clustering et k-plus proches voisins perdent leur sens.
- Nécessité de méthodes tenant compte de la **structure intrinsèque** des données.
- D'où la géométrie des données : travailler dans une *dimension intrinsèque*  $k \ll d$ .

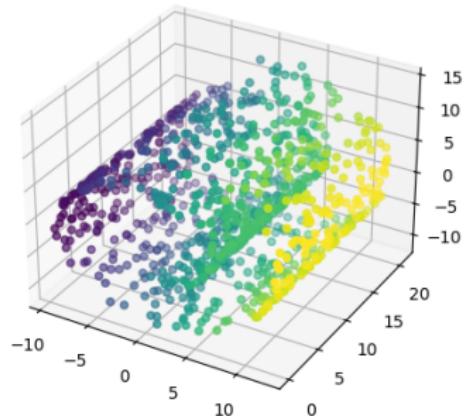
# Explosion dimensionnelle

- Données modernes : images, sons, textes → espaces  $\mathbb{R}^d$  avec  $d \gg 1$ .
- **Malédiction de la dimension :**
  - Volume croît exponentiellement avec  $d$ .
  - Distances deviennent peu discriminantes.
- Besoin de méthodes exploitant la structure « cachée ».

## Exemple : MNIST

- Chaque image :  $28 \times 28 = 784$  pixels  $\rightarrow$  point dans  $\mathbb{R}^{784}$ .
- Mais les chiffres manuscrits vivent sur une structure de dimension **intrinsèque**  $k \ll 784$ .
- Question : comment trouver  $k$  et représenter les données sur  $\mathbb{R}^k$  ?

# Exemple visuel : Swiss Roll



Un nuage de points en 3D mais structurellement 2D.

## Hypothèse de variété

Les données  $x_i \in \mathbb{R}^d$  sont concentrées sur une variété  $\mathcal{M}$  de dimension  $k \ll d$ .

$$x_i \in \mathcal{M} \subset \mathbb{R}^d, \quad \dim(\mathcal{M}) = k$$

# Espace métrique

Un espace métrique est une paire  $(X, d)$  avec :

$$d(x, y) \geq 0 \quad (\text{positivité})$$

$$d(x, y) = 0 \iff x = y \quad (\text{séparation})$$

$$d(x, y) = d(y, x) \quad (\text{symétrie})$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad (\text{inégalité triangulaire})$$

# Distances usuelles

- Norme  $\ell^2$ :  $d(x, y) = \|x - y\|_2 = \sqrt{\sum_i (x_i - y_i)^2}$ .
- Norme  $\ell^1$ :  $d(x, y) = \sum_i |x_i - y_i|$ .
- Cosinus:  $d(x, y) = 1 - \frac{\langle x, y \rangle}{\|x\| \|y\|}$ .
- Distances adaptées aux graphes et variétés.

# Distance cosinus et distance angulaire

Pour deux vecteurs  $x, y \in \mathbb{R}^d$  non nuls :

$$\cos \theta = \frac{x \cdot y}{\|x\| \|y\|}.$$

- **Distance cosinus classique :**

$$d_{\text{cos}}(x, y) = 1 - \cos \theta = 1 - \frac{x \cdot y}{\|x\| \|y\|}.$$

- Varie entre 0 et 2. - Très utilisée en NLP et apprentissage automatique. - Pas une vraie métrique : triangle inequality peut échouer.

- **Distance basée sur l'angle :**

$$d_{\text{angle}}(x, y) = \arccos \left( \frac{x \cdot y}{\|x\| \|y\|} \right).$$

- Angle réel entre les vecteurs. - Vraie métrique : satisfait la triangle inequality. - Correspond à distance géodésique sur la sphère unité  $S^{d-1}$ .

- Une variété  $\mathcal{M}$  est un espace qui ressemble localement à  $\mathbb{R}^k$ .
- Exemple : sphère  $S^2$  dans  $\mathbb{R}^3$ .

## Définition simplifiée

Pour chaque  $x \in \mathcal{M}$ , il existe un voisinage  $U$  et une bijection  $\varphi : U \rightarrow V \subset \mathbb{R}^k$ .

# Distance géodésique

Sur une variété  $\mathcal{M}$ , la distance naturelle est la longueur du plus court chemin  $\gamma$  contenu dans  $\mathcal{M}$  :

$$d_{\mathcal{M}}(x, y) = \inf_{\gamma: [0, 1] \rightarrow \mathcal{M}} \int_0^1 \|\dot{\gamma}(t)\| dt$$

Exemple : distance sur la sphère = angle central multiplié par le rayon.

# Variétés de données : définition détaillée

**Définition informelle :** Une variété  $\mathcal{M}$  de dimension  $k$  est un sous-ensemble de  $\mathbb{R}^d$  qui, localement autour de chaque point, ressemble à  $\mathbb{R}^k$ .

- Pour chaque point  $x \in \mathcal{M}$ , il existe un voisinage  $U$  et une application bijective **carte**  $\varphi : U \rightarrow V \subset \mathbb{R}^k$  telle que  $\varphi$  et  $\varphi^{-1}$  soient continues (ou différentiables pour les variétés différentiables).
- $k$  est la **dimension intrinsèque** de la variété.
- Exemple : le cercle  $S^1 \subset \mathbb{R}^2$  est une variété 1D, la sphère  $S^2 \subset \mathbb{R}^3$  est une variété 2D.

**Notation :**  $\mathcal{M}^k \subset \mathbb{R}^d$  indique une variété de dimension  $k$  dans  $\mathbb{R}^d$ .

# Propriétés fondamentales d'une variété

- **Localement Euclidienne** : autour de chaque point, les distances et topologie se comportent comme dans  $\mathbb{R}^k$ .
- **Dimension intrinsèque** : le nombre minimal de coordonnées nécessaires pour paramétriser la variété.
- **Continuité et différentiabilité** : cartes et applications de transition doivent être continues ou différentiables.
- **Géodésiques** : la distance naturelle sur la variété est la longueur du plus court chemin restant dans la variété.
- **Courbure** : mesure locale de la déviation par rapport à un espace plat  $\mathbb{R}^k$ .

# Exemples de variétés courantes

- Cercle  $S^1$  dans  $\mathbb{R}^2$  (dimension 1).
- Sphère  $S^2$  dans  $\mathbb{R}^3$  (dimension 2).
- Cylindre dans  $\mathbb{R}^3$  (dimension 2).
- Variétés de matrices de rang fixe :  $\{X \in \mathbb{R}^{m \times n} : \text{rank}(X) = r\}$ .
- Espace des rotations  $SO(3)$  (dimension 3), utilisé en robotique et vision.

# Carte et atlas

- **Carte** : fonction  $\varphi : U \subset M \rightarrow V \subset \mathbb{R}^k$  qui localement paramétrise la variété.
- **Atlas** : collection de cartes  $\{(U_i, \varphi_i)\}$  recouvrant toute la variété.
- **Exemple** : sphère  $S^2$ , on peut utiliser coordonnées sphériques différentes pour couvrir les pôles et l'équateur.
- Les cartes permettent de définir des notions de dérivée, intégrale et courbure sur la variété.

# Distances et géodésiques sur une variété

- La distance euclidienne  $\|x - y\|$  dans  $\mathbb{R}^d$  ne respecte pas toujours la structure intrinsèque de la variété.
- La **distance géodésique**  $d_M(x, y)$  est la longueur du chemin le plus court restant dans la variété.
- Exemples :
  - Cercle  $S^1$ : distance géodésique = arc le plus court entre deux points.
  - Sphère  $S^2$ : distance géodésique = longueur de l'arc de grand cercle.
- Les distances géodésiques sont fondamentales pour la réduction de dimension non linéaire et le clustering sur variétés.

# Distance géodésique sur la sphère : arc de grand cercle

- Le plus court chemin sur la sphère est l'**arc de grand cercle** reliant  $A$  et  $B$ .
- Longueur de l'arc :  $L = R \cdot \theta$ , avec  $\theta$  en radians.
- Angle au centre :  $\theta = \arccos\left(\frac{x \cdot y}{R^2}\right)$  pour  $x, y \in S^2$ .
- Pour une sphère unité ( $R = 1$ ) :  $d_{S^2}(x, y) = \arccos(x \cdot y)$ .

# Applications des variétés en science des données

- Réduction de dimension non linéaire : Isomap, LLE, Diffusion Maps.
- Détection de structure intrinsèque dans des données haute dimension.
- Modélisation de données sur des espaces non Euclidiens : rotations, formes, graphes.
- Méthodes de machine learning adaptées aux données de faible dimension intrinsèque.

# Laplacien de graphe

Pour un graphe  $G = (V, E)$  avec poids  $w_{ij}$  :

$$D_{ii} = \sum_j w_{ij},$$

$$L = D - W.$$

Propriétés :

- $L$  est semi-défini positif.
- $L$  approxime le Laplacien sur la variété sous-jacente.

## Exemple : chaleur sur un graphe

$$\frac{du}{dt} = -Lu$$

$L$  joue le rôle de dérivée seconde discrète. Solution :  $u(t) = e^{-tL}u(0)$ .

# Résumé

## Idée clé

Les données haute dimension vivent souvent sur une variété de faible dimension. La géométrie des données permet d'exploiter cette structure via distances, graphes et spectre.

## Prochain module

Réduction de dimension non linéaire (Isomap, Diffusion Maps, etc.).

# Introduction à la réduction de dimension

But : trouver  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  qui conserve la géométrie.

- PCA : directions de variance maximale.
- Isomap : distances géodésiques.
- Diffusion maps : diffusion de chaleur.

# PCA en formule

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$$

vecteurs propres de  $\Sigma \Rightarrow$  directions principales.

# Analyse en Composantes Principales (PCA) : Introduction

- PCA est une méthode linéaire de réduction de dimension.
- Objectif : trouver les directions principales (axes) qui capturent le plus de variance dans les données.
- Idée : projeter les données dans un sous-espace de dimension  $k \ll d$  en minimisant la perte d'information.
- Utile pour visualisation, compression, prétraitement de machine learning.

# Formulation mathématique de la PCA

- Soient  $X = [x_1, \dots, x_n]^\top \in \mathbb{R}^{n \times d}$  les données centrées ( $\bar{x} = 0$ ).
- Matrice de covariance :

$$\Sigma = \frac{1}{n} X^\top X \in \mathbb{R}^{d \times d}.$$

- Cherchons vecteurs propres  $v_j$  et valeurs propres  $\lambda_j$  :

$$\Sigma v_j = \lambda_j v_j, \quad j = 1, \dots, d.$$

- Les  $v_j$  sont les directions principales (composantes principales).

# Projection sur les composantes principales

- Les  $k$  premières composantes principales correspondent aux  $k$  plus grandes valeurs propres  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ .
- Projection :

$$y_i = V_k^\top x_i, \quad V_k = [v_1, \dots, v_k] \in \mathbb{R}^{d \times k}.$$

- Reconstruction approchée :

$$\hat{x}_i = V_k y_i = V_k V_k^\top x_i.$$

- Erreur de reconstruction : minimisée par la PCA.

# Propriété d'optimalité

La PCA est l'approximation linéaire optimale au sens des moindres carrés :

$$\min_{V_k \in \mathbb{R}^{d \times k}, V_k^\top V_k = I_k} \sum_{i=1}^n \|x_i - V_k V_k^\top x_i\|^2.$$

La solution est donnée par les  $k$  vecteurs propres associés aux  $k$  plus grandes valeurs propres de  $\Sigma$ .

# Variance expliquée

- Variance totale :  $\text{Tr}(\Sigma) = \sum_{j=1}^d \lambda_j$ .
- Variance capturée par les  $k$  premières composantes :  $\sum_{j=1}^k \lambda_j$ .
- Fraction de variance expliquée :

$$\text{FVE}(k) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^d \lambda_j}.$$

- Permet de choisir le nombre optimal  $k$  pour la réduction de dimension.

# Exemple numérique

- Données  $X \in \mathbb{R}^{100 \times 5}$  simulées.
- Calculer  $\Sigma = X^\top X / 100$ .
- Calcul des valeurs propres et vecteurs propres.
- Projection sur  $k = 2$  premières composantes principales pour visualisation.
- Observer la concentration des données le long des directions principales.

# PCA et SVD

- Alternative : utiliser la décomposition en valeurs singulières (SVD) :  
 $X = U\Sigma V^\top$ .
- Composantes principales : colonnes de  $V$ .
- Valeurs propres : carrés des valeurs singulières divisées par  $n$ .
- Utile pour  $d \gg n$  ou pour stabilité numérique.

# Exercice PCA

Soit un ensemble de données centrées  $X \in \mathbb{R}^{10 \times 3}$  :

- Calculer la matrice de covariance  $\Sigma$ .
- Déterminer les vecteurs propres et valeurs propres.
- Projeter les données sur les 2 premières composantes principales.

# Corrigé exercice PCA

- Étape 1 :  $\Sigma = X^\top X / 10$ .
- Étape 2 : calcul des valeurs propres  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  et vecteurs propres  $v_1, v_2, v_3$ .
- Étape 3 : projection  $y_i = [v_1, v_2]^\top x_i$ .
- Étape 4 : visualisation et analyse de la variance expliquée.

# Isomap

- ① Construire graphe  $k$ -NN.
- ② Approximer distances géodésiques par plus courts chemins.
- ③ Appliquer MDS (multidimensional scaling).

# Isomap : Introduction

- Isomap (Isometric Mapping) est une méthode non-linéaire de réduction de dimension.
- Objectif : préserver les distances géodésiques entre les points d'une variété plongée dans un espace de haute dimension.
- Extension de MDS (Multidimensional Scaling) aux variétés non-linéaires.

# Idée clé d'Isomap

- ① Construire un graphe des plus proches voisins ( $k$ -NN ou  $\epsilon$ -voisinage).
- ② Approximer les distances géodésiques par des plus courts chemins dans le graphe.
- ③ Appliquer le MDS classique avec ces distances géodésiques approximées.

## Étape 1 : Graphe de voisinage

- Pour chaque point  $x_i$ , on relie ses  $k$  plus proches voisins (ou tous les voisins dans une boule de rayon  $\epsilon$ ).
- Arêtes pondérées par la distance euclidienne :

$$w_{ij} = \|x_i - x_j\|_2.$$

- Graphe  $G = (V, E)$  approximant la structure locale de la variété.

## Étape 2 : Distances géodésiques

- La distance géodésique entre  $x_i$  et  $x_j$  est approximée par la longueur du plus court chemin dans  $G$  :

$$d_G(i, j) = \min_{chemins(i \rightarrow j)} \sum_{(u, v) \in chemin} w_{uv}.$$

- Utilisation d'algorithmes classiques : Dijkstra ou Floyd–Warshall.
- Matrice des distances géodésiques  $D_G = (d_G(i, j))_{i,j}$ .

## Étape 3 : Application de MDS

- Objectif : trouver une configuration de points  $Y_1, \dots, Y_n \in \mathbb{R}^d$  telle que

$$\|Y_i - Y_j\|^2 \approx d_G(i, j)^2.$$

- On applique le MDS classique (Scaling multidimensionnel) :

- On construit la matrice des distances au carré  $D_G^{(2)} = (d_G(i, j)^2)_{ij}$ .
- On centre cette matrice :

$$B = -\frac{1}{2}HD_G^{(2)}H, \quad H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top.$$

- $B$  est une approximation de la matrice de Gram  $YY^\top$ .
- On diagonalise  $B = V\Lambda V^\top$ .
- Les coordonnées en dimension  $d$  sont données par :

$$Y = V_d \Lambda_d^{1/2},$$

où  $V_d$  contiennent les  $d$  vecteurs propres principaux et  $\Lambda_d$  les  $d$  plus grandes valeurs propres.

- Ainsi, on obtient une immersion en  $d$  dimensions qui préserve au mieux les distances géodésiques.

## Étape 3 : Application de MDS

- On applique le MDS classique sur  $D_G$  pour obtenir une représentation en basse dimension.
- Centrage double :

$$B = -\frac{1}{2}HD_G^{(2)}H, \quad H = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top.$$

- Décomposition spectrale :

$$B = V\Lambda V^\top.$$

- Coordonnées en dimension  $d$  :

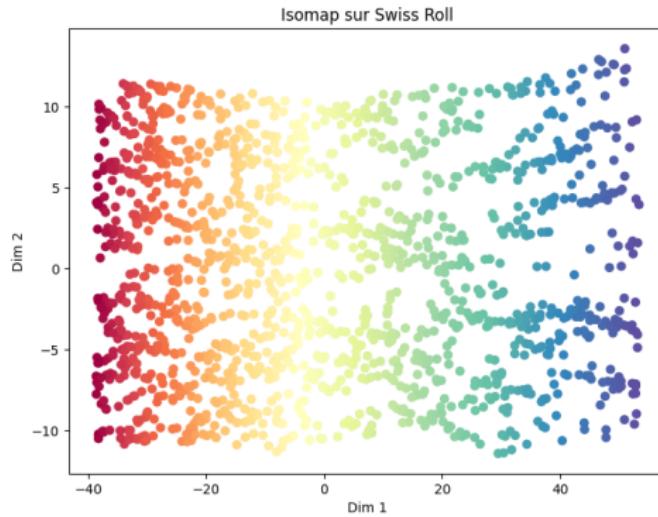
$$Y = V_d\Lambda_d^{1/2}.$$

# Propriétés d'Isomap

- Préserve la géométrie globale de la variété.
- Gère des données fortement non-linéaires (spirale, Swiss roll).
- Consistance théorique : converge vers la métrique de la variété quand  $n \rightarrow \infty$ .
- Sensible aux choix de  $k$  ou  $\epsilon$ .

## Exemple : Swiss Roll

- Données 3D enroulées en spirale (« rouleau suisse »).
- PCA : incapacité à « dérouler » la structure.
- Isomap : reconstitue correctement la structure 2D sous-jacente.



# Exercice Isomap

- Générer un jeu de données 3D de type Swiss roll.
- Construire le graphe  $k$ -NN avec  $k = 10$ .
- Calculer les plus courts chemins (algorithme de Dijkstra).
- Appliquer MDS sur la matrice de distances.
- Visualiser la représentation 2D obtenue.

# Corrigé Exercice Isomap

- Étape 1 : génération des points  $(x, y, z)$ .
- Étape 2 : construction du graphe  $k$ -NN.
- Étape 3 : distances géodésiques via Dijkstra.
- Étape 4 : matrice  $B$  et décomposition spectrale.
- Résultat : une carte 2D déroulant le Swiss roll.

# Résumé intuitif d'Isomap

- But : représenter les données dans un espace de dimension réduite  $d$  (typiquement 2 ou 3).
- On calcule d'abord les distances géodésiques  $d_G(i,j)$  entre points dans l'espace original (via le graphe de voisinage).
- Puis on cherche des points  $Y_i \in \mathbb{R}^d$  tels que :

$$\|Y_i - Y_j\| \approx d_G(i,j).$$

- Ainsi :
  - Les distances locales sont respectées.
  - La géométrie intrinsèque de la variété est préservée.
  - On obtient une carte en basse dimension reflétant la structure réelle des données.
- Différence clé avec PCA : Isomap ne se base pas uniquement sur les distances euclidiennes globales, mais sur les distances intrinsèques (géodésiques).

# Diffusion Maps

- Construire matrice de transition  $P = D^{-1}W$ .
- Valeurs propres  $\lambda_1, \lambda_2, \dots$  et vecteurs propres  $\psi_i$ .
- Représentation des données :

$$x \mapsto (\lambda_1^t \psi_1(x), \dots, \lambda_k^t \psi_k(x))$$

# Exemple de visualisation

