

ISOMAP

Advanced Institute for Artificial Intelligence – AI2

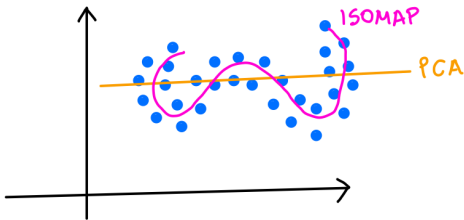
<https://advancedinstitute.ai>

A técnica *Isometric Feature Mapping* (ISOMAP) é um algoritmo para redução não linear de dimensionalidade que baseia-se no arcabouço de **aprendizado de variedades** (*manifold learning*). A principal diferença desse método para outros do tipo PCA (KPCA) e LDA/MDA (KFDA) é que aqui construímos uma relação de adjacência entre as amostras.

Ideia geral: construir um grafo unindo os vizinhos mais próximos, calcular os menores caminhos entre cada par de vértices e encontrar um mapeamento para o plano que preserve essas distâncias.

Hipótese: caminhos mínimos em um grafo podem aproximar bem as distâncias geodésicas em espaços não Euclidianos (variedades).

Aprendizado de métricas: métodos lineares falham em aprender uma medida de distância adequada na presença de não linearidades nos dados.



O algoritmo do ISOMAP pode ser resumidos em três passos:

- 1 Induzir o grafo k -nn a partir do conjunto de dados $\mathcal{X} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, em que $\mathbf{x}_i \in \mathbb{R}^n$ e $y \in \mathcal{Y}$ tal que $\mathcal{Y} = \{\omega_1, \omega_2, \dots, \omega_c\}$ e $k < n$. Para amostras que não são vizinhas, utilizamos o valor $+\infty$ como distância entre elas.
- 2 Criar a matriz de distâncias ponto a ponto $\mathbf{D} \in \mathbb{R}^{m \times m}$ entre as amostras de \mathcal{X} . Desta forma, D_{ij} representa o menor caminho entre as amostras \mathbf{x}_i e \mathbf{x}_j .
- 3 Encontrar o conjunto de pontos $\hat{\mathcal{X}}$ no espaço \mathbb{R}^p tal que as distâncias originais sejam "preservadas". Esta etapa é realizada pela técnica MDS (*Multidimensional Scaling*).

Trata-se, portanto, de uma abordagem **global**, pois utiliza todas as amostras para calcular as distâncias. A técnica é baseada em um teorema que diz que caminhos mínimos em grafos k -nn são boas aproximações para as distâncias geodésicas (distâncias que respeitam as variedades).

Como podemos calcular essas distâncias geodésicas? Seja $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$ um grafo tal que $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$ e \mathcal{E} representam os conjuntos de vértices e arestas, respectivamente, e $w : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^+$ uma função que calcula a distância entre dois nós. Note que os nós do grafos são compostos pelas amostras em \mathcal{X} , ou seja, $v_i = x_i$.

Temos que um caminho P_{v_1, v_m}^* é dito ser ótimo se o seu custo C é o menor possível, ou seja:

$$C(P_{v_1, v_m}^*) = \sum_{i=1}^{m-1} w(v_i, v_{i+1}) = w(v_1, v_2) + w(v_2, v_3) + \dots + w(v_{m-1}, v_m), \quad (1)$$

em que $C(P_{v_1, v_m}^*)$ se aproxima da distância geodésica entre v_1 e v_m .

Uma técnica bastante conhecida para calcular caminhos mínimos em grafos é chamada de **Algoritmo de Dijkstra**. Existe um teorema que mostra que este algoritmo sempre calcula a distância geodésica entre um nó **fonte** e demais nós do grafo. Vejamos as seguintes definições:

- $\lambda(v)$: menor custo até o momento de alguma amostra s até v .
- $\pi(v)$: predecessor de v no caminho de custo ótimo.
- Q : fila de prioridades dos vértices.

Vejamos, agora, o algoritmo do Dijkstra.

Dijkstra(\mathcal{G}, s)

for each $v \in \mathcal{V}$ **do**

$\lambda(v) \leftarrow +\infty$

$\pi(v) \leftarrow nil$

$\lambda(s) \leftarrow 0$

$\pi(s) \leftarrow nil$

$\mathcal{Q} \leftarrow \mathcal{V}, \mathcal{S} \leftarrow \emptyset$

while $\mathcal{Q} \neq \emptyset$ **do**

$u \leftarrow Remove(\mathcal{Q})$

$\mathcal{S} \leftarrow \mathcal{S} \cup u$

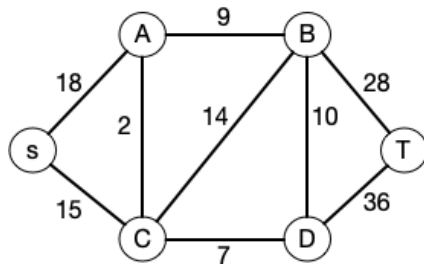
for each $v \in \mathcal{N}(u)$ **do**

if $\lambda(v) > \lambda(u) + w(u, v)$

then

$\lambda(v) \leftarrow \lambda(u) + w(u, v)$

$\pi(v) \leftarrow u$



Como fazemos o mapeamento para o espaço \mathbb{R}^k (k é escolhido pelo usuário)? Assim, dada a nossa matriz de distâncias geodésicas calculadas pelo algoritmo de Dijkstra, o objetivo é encontrar um novo mapeamento das amostras de tal forma que suas distâncias originais sejam preservadas. A solução para este problema é dada pela técnica MDS. Lembrando que a distância D_{ij} entre dois vetores \mathbf{x}_i e \mathbf{x}_j é dada por:

$$D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j). \quad (2)$$

Note que a matriz de distâncias \mathbf{D} já foi calculada anteriormente.

O método MDS baseia-se na solução de dois subproblemas:

- 1 Encontrar a matriz $\mathbf{B} \in \mathbb{R}^{m \times m}$ a partir de \mathbf{D} e
- 2 Recuperar (mapear) as coordenadas dos novos pontos a partir de \mathbf{B} .

Note que \mathbf{B} corresponde à matriz de produtos internos entre todas as amostras, ou seja, $B_{ij} = \mathbf{x}_i^T \mathbf{x}_j$. Vejamos, então, como resolver cada um dos subproblemas.

Subproblema 1: encontrar B a partir de D .

Hipótese: a média $\mu \in \mathbb{R}^n$ dos dados é nula, ou seja:

$$\mu = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i = 0. \quad (3)$$

Assumimos essa hipótese pois, caso contrário, teremos infinitas soluções para o problema (diferentes versões do espaço que estão transladadas, por exemplo). Ademais, aplicando a operação distributiva na Equação 2, temos que:

$$D_{ij} = \mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j. \quad (4)$$

A partir da matriz \mathbf{D} , podemos calcular a média de uma coluna s da seguinte maneira:

$$\begin{aligned}
 M_s &= \frac{1}{m} \sum_{i=1}^m D_{is} = \frac{1}{m} \sum_{i=1}^m \overbrace{(\mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_s + \mathbf{x}_s^T \mathbf{x}_s)}^{D_{is}} \\
 &= \frac{1}{m} \left(\sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - 2 \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_s + \sum_{i=1}^m \mathbf{x}_s^T \mathbf{x}_s \right) \\
 &= \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_s + \frac{1}{m} \sum_{i=1}^m \mathbf{x}_s^T \mathbf{x}_s \quad \text{constante} \\
 &= \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_s^T \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i + \frac{1}{m} m (\mathbf{x}_s^T \mathbf{x}_s) \quad \text{0 Equação 3} \\
 &= \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_s^T \mathbf{x}_s.
 \end{aligned} \tag{5}$$

Analogamente, podemos calcular a média da linha r da matriz D da seguinte maneira:

$$\begin{aligned}
 M_r &= \frac{1}{m} \sum_{j=1}^m D_{rj} = \frac{1}{m} \sum_{j=1}^m \overbrace{(\mathbf{x}_r^T \mathbf{x}_r - 2\mathbf{x}_r^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j)}^{D_{rj}} \\
 &= \frac{1}{m} \left(\sum_{j=1}^m \mathbf{x}_r^T \mathbf{x}_r - 2 \sum_{j=1}^m \mathbf{x}_r^T \mathbf{x}_j + \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \right) \\
 &= \frac{1}{m} \sum_{j=1}^m \mathbf{x}_r^T \mathbf{x}_r - \frac{2}{m} \sum_{j=1}^m \mathbf{x}_r^T \mathbf{x}_j + \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \quad (6) \\
 &= \frac{1}{m} m(\mathbf{x}_r^T \mathbf{x}_r) - 2\mathbf{x}_r \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i + \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j = \mathbf{x}_r^T \mathbf{x}_r + \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j.
 \end{aligned}$$

Finalmente, podemos calcular a média \bar{M} dos elementos da matriz D da seguinte forma:

$$\begin{aligned}
\bar{M} &= \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m D_{ij} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m (\mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j) \\
&= \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{2}{m^2} \sum_{i=1}^m \sum_{j=1}^m \mathbf{x}_i^T \mathbf{x}_j + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \\
&= \frac{1}{m^2} m \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{2}{m^2} \sum_{i=1}^m \mathbf{x}_i^T \sum_{j=1}^m \mathbf{x}_j + \frac{1}{m^2} m \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \\
&= \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j + \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j \\
&= \underbrace{\frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i + \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j}_{\text{termos iguais}} = \frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i.
\end{aligned} \tag{7}$$

0 Equação 3

Podemos definir o elemento B_{ij} da matriz de produtos internos \mathbf{B} da seguinte forma por meio da Equação 4:

$$B_{ij} = \mathbf{x}_i^T \mathbf{x}_j = -\frac{1}{2} (D_{ij} - \mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}^T \mathbf{x}_j). \quad (8)$$

Da Equação 6, podemos isolar o termo $-\mathbf{x}_r^T \mathbf{x}_r$. Tornando $r = i$, temos que:

$$-\mathbf{x}_i^T \mathbf{x}_i = \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j - \frac{1}{m} \sum_{j=1}^m D_{ij}. \quad (9)$$

Da Equação 5, podemos isolar o termo $-\mathbf{x}_s^T \mathbf{x}_s$. Tornando $s = j$, temos que:

$$-\mathbf{x}_j^T \mathbf{x}_j = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{1}{m} \sum_{i=1}^m D_{ij}. \quad (10)$$

Somando-se as Equações 9 e 10, temos que:

$$\begin{aligned} -\mathbf{x}_i^T \mathbf{x}_i + (-\mathbf{x}_j^T \mathbf{x}_j) &= \frac{1}{m} \sum_{j=1}^m \mathbf{x}_j^T \mathbf{x}_j - \frac{1}{m} \sum_{j=1}^m D_{ij} + \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i - \frac{1}{m} \sum_{i=1}^m D_{ij} \\ &= -\frac{1}{m} \sum_{j=1}^m D_{ij} - \frac{1}{m} \sum_{i=1}^m D_{ij} + \frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i \end{aligned} \quad (11)$$

Da Equação 7, podemos escrever:

$$\frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m D_{ij}. \quad (12)$$

Lembrando que o nosso objetivo é reescrever os termos b_{ij} da matriz de produtos internos \mathbf{B} utilizando D_{ij} .

Assim sendo, podemos reescrever a Equação 8:

$$\begin{aligned}
 B_{ij} = \mathbf{x}_i^T \mathbf{x}_j &= -\frac{1}{2} \left(D_{ij} - \overbrace{\mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}^T \mathbf{x}_j}^{\text{Equação 11}} \right) \\
 &= -\frac{1}{2} \left(D_{ij} - \frac{1}{m} \sum_{j=1}^m D_{ij} - \frac{1}{m} \sum_{i=1}^m D_{ij} + \underbrace{\frac{2}{m} \sum_{i=1}^m \mathbf{x}_i^T \mathbf{x}_i}_{\text{Equação 12}} \right) \\
 &= -\frac{1}{2} \left(D_{ij} - \frac{1}{m} \sum_{j=1}^m D_{ij} - \frac{1}{m} \sum_{i=1}^m D_{ij} + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m D_{ij} \right)
 \end{aligned} \tag{13}$$

Aplicando-se a operação distributiva na Equação 13, podemos criar a matriz $\mathbf{A} \in \mathbb{R}^{m \times m}$ e, fazendo $A_{ij} = -\frac{1}{2}D_{ij}$, temos as seguintes definições:

- $\bar{A}_{i.} = \frac{1}{m} \sum_{j=1}^m A_{ij}$ (média na linha i)
- $\bar{A}_{.j} = \frac{1}{m} \sum_{i=1}^m A_{ij}$ (média na coluna j)
- $\bar{M} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m A_{ij}$ (média da matriz \mathbf{A})

Desta forma, podemos reescrever a Equação 13 da seguinte forma:

$$\begin{aligned} B_{ij} &= -\frac{1}{2} \left(D_{ij} - \frac{1}{m} \sum_{j=1}^m D_{ij} - \frac{1}{m} \sum_{i=1}^m D_{ij} + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m D_{ij} \right) \\ &= A_{ij} - \bar{A}_{i.} - \bar{A}_{.j} + \bar{M}. \end{aligned} \quad (14)$$

Podemos mostrar, ainda, que:

$$\mathbf{B} = \mathbf{H} \mathbf{A} \mathbf{H}, \quad (15)$$

em que

$$\mathbf{H} = \mathbf{I} - \frac{1}{m} \mathbf{1} \mathbf{1}^T, \quad (16)$$

tal que $\mathbf{I} \in \mathbb{R}^{m \times m}$ é a matriz identidade e $\mathbf{1} \in \mathbb{R}^m$ é um vetor com todos os elementos iguais à 1.

Temos, então, que:

$$\mathbf{U} = \mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \quad (17)$$

tal que $\mathbf{U} \in \mathbb{R}^{m \times m}$. Note que a Equação 15 nada mais é do que a forma matricial da Equação 14, ou seja:

$$\begin{aligned} \mathbf{B} &= \mathbf{H}\mathbf{A}\mathbf{H} = \left(\mathbf{I} - \frac{1}{m}\mathbf{U}\right) \mathbf{A} \left(\mathbf{I} - \frac{1}{m}\mathbf{U}\right) = \left(\mathbf{A} - \frac{1}{m}\mathbf{U}\mathbf{A}\right) \left(\mathbf{I} - \frac{1}{m}\mathbf{U}\right) \\ &= \mathbf{A} - \mathbf{A}\frac{1}{m}\mathbf{U} - \frac{1}{m}\mathbf{U}\mathbf{A} + \frac{1}{m^2}\mathbf{U}\mathbf{A}\mathbf{U}. \end{aligned} \quad (18)$$

Conseguimos, então, resolver o subproblema 1, ou seja, encontrar a matriz \mathbf{B} . Falta, agora, resolvermos o subproblema 2, ou seja, recuperar as coordenadas $\hat{\mathbf{x}}_i \in \mathbb{R}^p$ a partir da matriz \mathbf{B} , $\forall \mathbf{x}_i \in \mathcal{X}$.

Como \mathbf{B} é a matriz dos produtos internos, podemos reescrevê-la da seguinte maneira:

$$\mathbf{B} = \mathbf{X}\mathbf{X}^T, \quad (19)$$

em que $\mathbf{X} \in \mathbb{R}^{m \times n}$ é a matriz dos pontos em \mathbb{R}^n . A matriz \mathbf{B} é simétrica, positiva semi-definida e possui rank $p = n$, possuindo p autovalores não negativos (linhas/colunas linearmente independentes). A sua decomposição espectral é dada como segue:

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad (20)$$

em que $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ é a matriz diagonal dos autovalores e $\mathbf{V} \in \mathbb{R}^{m \times m}$ é a matriz dos autovetores.

Sem perda de generalidade, podemos considerar $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Considerando apenas os p maiores autovalores, podemos reescrever a Equação 20 da seguinte forma:

$$B = V' \Lambda' V'^T, \quad (21)$$

em que $\Lambda' \in \mathbb{R}^{p \times p}$ é a matriz diagonal dos autovalores e $V' \in \mathbb{R}^{m \times p}$ denota a matriz dos autovetores. Da Equação 19, temos que:

$$B = \hat{X} \hat{X}^T = V' \Lambda' V'^T = \underbrace{V' \Lambda'^{1/2}}_{\hat{X}} \overbrace{\Lambda'^{1/2} V'^T}^{\hat{X}^T}, \quad (22)$$

em que $\Lambda'^{1/2}$ equivale à aplicação da raiz quadrada em todos os elementos de Λ' .

Desta forma, a matriz dos elementos projetados pode ser obtida da seguinte forma:

$$\hat{\mathbf{X}} = \mathbf{V}'\mathbf{\Lambda}'^{1/2}. \quad (23)$$

Vejamos, agora, o algoritmo MDS.

- ❶ Construir o grafo k -nn e aplicar o algoritmo de Dijkstra m vezes mudando o nó fonte para calcular a matriz \mathbf{D} .
- ❷ Faça $\mathbf{A} = -\frac{1}{2}\mathbf{D}$.
- ❸ Faça $\mathbf{H} = \mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^T$.
- ❹ Calcule $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$.
- ❺ Encontrar os autovalores e autovetores de \mathbf{B} .
- ❻ Tomar os p autovetores associados aos p maiores autovalores de \mathbf{B} para montar \mathbf{V}' e $\mathbf{\Lambda}'$.
- ❼ Calcular $\hat{\mathbf{X}} = \mathbf{V}'\mathbf{\Lambda}'^{1/2}$.

Vejamos algumas limitações do ISOMAP:

- Não é supervisionado.
- Em dados não convexos, ou seja, quando temos "buracos" nas variedades, o ISOMAP pode não funcionar adequadamente.