ISOMAP

Advanced Institute for Artificial Intelligence – AI2

https://advancedinstitute.ai

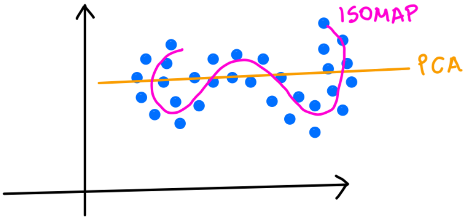
Introdução

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 construimos 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 $\mathcal{X} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$, em que $\boldsymbol{x}_i \in \mathbb{R}^n$ e $y \in \mathcal{Y}$ tal que $\mathcal{Y} = \{\omega_1, \omega_2, \ldots, \omega_c\}$ 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 $\boldsymbol{D} \in \mathbb{R}^{m \times m}$ entre as amostras de $\mathcal{X}$. Desta forma, D_{ij} representa o menor caminho entre as amostras $\boldsymbol{x}_i$ e $\boldsymbol{x}_j$.
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 $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ 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, $\boldsymbol{v}_i = \boldsymbol{x}_i$.

Temos que um caminho $P^*_{\boldsymbol{v}_1, \boldsymbol{v}_m}$ é dito ser ótimo se o seu custo C é o menor possível, ou seja:

$$C(P^*_{\boldsymbol{v}_1, \boldsymbol{v}_m}) = \sum_{i=1}^{m-1} w(\boldsymbol{v}_i, \boldsymbol{v}_{i+1}) = w(\boldsymbol{v}_2, \boldsymbol{v}_2) + w(\boldsymbol{v}_2, \boldsymbol{v}_3) + \ldots + w(\boldsymbol{v}_{n-1}, \boldsymbol{v}_m), \quad (1)$$

em que $C(P^*_{\boldsymbol{v}_1, \boldsymbol{v}_m})$ se aproxima da distância geodésica entre $\boldsymbol{v}_1$ e $\boldsymbol{v}_m$.

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 seguinte definições:

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

Vejamos, agora, o algoritmo do Dijkstra.

Dijkstra($\mathcal{G}$, $\boldsymbol{s}$)

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

  $\lambda(\boldsymbol{v}) \leftarrow +\infty$

  $\pi(\boldsymbol{v}) \leftarrow nil$

$\lambda(\boldsymbol{s}) \leftarrow 0$

$\pi(\boldsymbol{s}) \leftarrow nil$

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

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

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

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

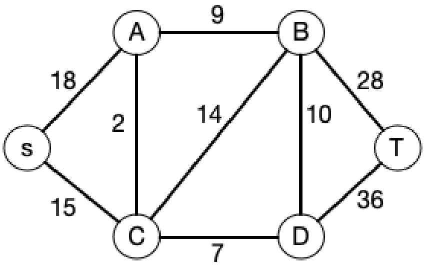
  for each $\boldsymbol{v} \in \mathcal{N}(\boldsymbol{u})$ **do**

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

    then

      $\lambda(\boldsymbol{v}) \leftarrow \lambda(\boldsymbol{u}) + w(\boldsymbol{u}, \boldsymbol{v})$

      $\pi(\boldsymbol{v}) \leftarrow \boldsymbol{u}$



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 $\boldsymbol{x}_i$ e $\boldsymbol{x}_j$ é dada por:

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

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

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

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

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

Subproblema 1: encontrar $\boldsymbol{B}$ a partir de $\boldsymbol{D}$.

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

$$\boldsymbol{\mu} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{x}_i = 0. \tag{3}$$

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:

$$D_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_i - 2\boldsymbol{x}_i^T \boldsymbol{x}_j + \boldsymbol{x}_j^T \boldsymbol{x}_j. \tag{4}$$

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

$$
\begin{aligned}
M_s = \frac{1}{m}\sum_{i=1}^{m} D_{is} &= \frac{1}{m}\sum_{i=1}^{m} \overbrace{(\boldsymbol{x}_i^T\boldsymbol{x}_i - 2\boldsymbol{x}_i^T\boldsymbol{x}_s + \boldsymbol{x}_s^T\boldsymbol{x}_s)}^{D_{is}} \\
&= \frac{1}{m}\left(\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i - 2\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_s + \sum_{i=1}^{m}\boldsymbol{x}_s^T\boldsymbol{x}_s\right)
\end{aligned}
$$

constante

$$
\begin{aligned}
&= \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i - \frac{2}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_s + \frac{1}{m}\cancel{\sum_{i=1}^{m}}\boldsymbol{x}_s^T\boldsymbol{x}_s \qquad (5)
\end{aligned}
$$

0 Equação 3

$$
\begin{aligned}
&= \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i - 2\boldsymbol{x}_s\cancel{\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i} + \frac{1}{m}m(\boldsymbol{x}_s^T\boldsymbol{x}_s) \\
&= \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i + \boldsymbol{x}_s^T\boldsymbol{x}_s.
\end{aligned}
$$

Analogamente, podemos calcular a média da linha r da matriz $\boldsymbol{D}$ da seguinte maneira:

$$
\begin{aligned}
M_r = \frac{1}{m}\sum_{j=1}^{m} D_{rj} &= \frac{1}{m}\sum_{j=1}^{m} \overbrace{(\boldsymbol{x}_r^T\boldsymbol{x}_r - 2\boldsymbol{x}_r^T\boldsymbol{x}_j + \boldsymbol{x}_j^T\boldsymbol{x}_j)}^{D_{rj}} \\
&= \frac{1}{m}\left(\sum_{j=1}^{m}\boldsymbol{x}_r^T\boldsymbol{x}_r - 2\sum_{j=1}^{m}\boldsymbol{x}_r^T\boldsymbol{x}_j + \sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j\right) \\
&= \frac{1}{m}\cancelto{\text{constante}}{\sum_{j=1}^{m}}\boldsymbol{x}_r^T\boldsymbol{x}_r - \frac{2}{m}\sum_{j=1}^{m}\boldsymbol{x}_r^T\boldsymbol{x}_j + \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j \qquad (6) \\
&= \frac{1}{m}m(\boldsymbol{x}_r^T\boldsymbol{x}_r) - 2\boldsymbol{x}_r\cancelto{0 \text{ Equação 3}}{\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_j} + \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j = \boldsymbol{x}_r^T\boldsymbol{x}_r + \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j.
\end{aligned}
$$

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

$$
\begin{aligned}
\bar{M} &= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} D_{ij} = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\boldsymbol{x}_i^T\boldsymbol{x}_i - 2\boldsymbol{x}_i^T\boldsymbol{x}_j + \boldsymbol{x}_j^T\boldsymbol{x}_j\right) \\
&= \frac{1}{m^2}\sum_{i=1}^{m}\cancelto{m}{\sum_{j=1}^{m}}\boldsymbol{x}_i^T\boldsymbol{x}_i - \frac{2}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_j + \frac{1}{m^2}\cancelto{m}{\sum_{i=1}^{m}}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j \\
&= \frac{1}{m^2}m\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i - \frac{2}{m^2}\sum_{i=1}^{m}\boldsymbol{x}_i^T\sum_{j=1}^{m}\boldsymbol{x}_j + \frac{1}{m^2}m\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j \\
&= \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i - \cancelto{0\text{ Equação 3}}{\frac{2}{m}\sum_{i=1}^{m}}\boldsymbol{x}_i^T\frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j + \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j \\
&= \underbrace{\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i + \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{x}_j^T\boldsymbol{x}_j}_{\text{termos iguais}} = \frac{2}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i.
\end{aligned}
\tag{7}
$$

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

$$B_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j = -\frac{1}{2} \left(D_{ij} - \boldsymbol{x}_i^T \boldsymbol{x}_i - \boldsymbol{x}^T \boldsymbol{x}_j \right). \tag{8}$$

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

$$-\boldsymbol{x}_i^T \boldsymbol{x}_i = \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{x}_j^T \boldsymbol{x}_j - \frac{1}{m} \sum_{j=1}^{m} D_{ij}. \tag{9}$$

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

$$-\boldsymbol{x}_j^T \boldsymbol{x}_j = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{x}_i^T \boldsymbol{x}_i - \frac{1}{m} \sum_{i=1}^{m} D_{ij}. \tag{10}$$

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

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

Da Equação 7, podemos escrever:

$$\frac{2}{m}\sum_{i=1}^{m}\boldsymbol{x}_i^T\boldsymbol{x}_i = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}D_{ij}. \tag{12}$$

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

Assim sendo, podemos reescrever a Equação 8:

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

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

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

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

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

Podemos mostrar, ainda, que:

$$\boldsymbol{B} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{H}, \tag{15}$$

em que

$$\boldsymbol{H} = \boldsymbol{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^T, \tag{16}$$

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

Temos, então, que:

$$\boldsymbol{U} = \mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \tag{17}$$

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

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

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

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

$$\boldsymbol{B} = \hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T, \tag{19}$$

em que $\hat{\boldsymbol{X}} \in \mathbb{R}^{m \times p}$ é a matriz dos pontos em $\mathbb{R}^p$. A matriz $\boldsymbol{B}$ possui certas propriedades, isto é, é simétrica, positiva semi-definida e possui rank p, o que significa que possui p autovalores não negativos. A sua decomposição espectral é dada como segue:

$$\boldsymbol{B} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T, \tag{20}$$

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

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

$$\boldsymbol{B} = \boldsymbol{V}'\boldsymbol{\Lambda}'\boldsymbol{V}'^T, \tag{21}$$

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

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

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

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

$$\hat{\boldsymbol{X}} = \boldsymbol{V}'\boldsymbol{\Lambda}'^{1/2}. \tag{23}$$

Vejamos, agora, o algoritmo MDS.

1. Construir o grafo k-nn e aplicar o algoritmo de Dijkstra m vezes mudando o nó fonte para calcular a matriz $\boldsymbol{D}$.
2. Faça $\boldsymbol{A} = \frac{1}{2}\boldsymbol{D}$.
3. Faça $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{m}\boldsymbol{1}\boldsymbol{1}^T$.
4. Calcule $\boldsymbol{B} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{H}$.
5. Encontrar os autovalores e autovetores de $\boldsymbol{B}$.
6. Tomar os p autovetores associados aos p maiores autovalores de $\boldsymbol{B}$ para montar $\boldsymbol{V}'$ e $\boldsymbol{\Lambda}'$.
7. Calcular $\hat{\boldsymbol{X}} = \boldsymbol{V}'\boldsymbol{\Lambda}'^{1/2}$.

Vejamos algumas limitações do ISOMAP:

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

Um agradecimento especial ao **Prof. Alexandre Levada** do Centro de Ciências Exatas e de Tecnologia, Departamento de Computação, Universidade Federal de São Carlos, pelas notas de aula.