## MDS (Multidimensional Scaling).

Given $N$ objects s.t. a distance function is defined on them. (In particular, $\mathcal{X}=\{x_1,\cdots,x_N\}\subseteq\mathbb{R}^D$.)

$\boxed{\delta_{ij}} :=$ distance b/w object $i$ and object $j$.

$\boxed{\Delta} := (\delta_{ij})_{N\times N}$ is called the $\boxed{\text{dissimilarity matrix}}$.

**Goal**: Given $\Delta$. Try to find $\mathcal{Y}=\{y_1,\cdots,y_N\}\subseteq\mathbb{R}^d$

s.t. $\|y_i-y_j\|\approx\delta_{ij}$, $\forall\, i,j=1,\cdots,N$.

Formally, we can achieve this via solving the following optimization problem:

$$\min_{\substack{y_1,\cdots,y_N\\ \in\mathbb{R}^d}} \sum_{i<j}\left(\|y_i-y_j\|-\delta_{ij}\right)^2. \quad (*)$$

**Rmk**:

① MDS is a dimension reduction method when our input is $\mathcal{X}=\{x_1,\cdots,x_N\}\subseteq\mathbb{R}^D$ and we choose $d\ll D$.

② In general, the input is just the dissimilarity matrix $\Delta$.

③ MDS is a "global" method since it tries to minimize $(*)$, which is the total error made when choosing $\mathcal{Y}$.

④ MDS can be solved using optimization algorithms such as gradient descent.

## Isomap (Isometric map)

**Algorithm**:

**Inputs**: ① $\mathcal{X}=\{x_1,\cdots,x_N\}\subseteq\mathbb{R}^D$.
② $d\ll D$. ③ $\varepsilon>0$ or $K\in\mathbb{N}$

**Output**: $\mathcal{Y}=\{y_1,\cdots,y_N\}\subseteq\mathbb{R}^d$.

**Procedure**:

1° Determine the neighbors of each pt by
 (i) $\varepsilon$-neighborhoods, or
 (ii) $K$ nearest neighborhoods. (KNN).

2° Construct a neighborhood graph $G$:
Vertex set: $V=\mathcal{X}=\{x_1,\cdots,x_N\}$.
The edge set $E$ is determined by
 (i) $\overline{x_ix_j}\in E \iff \|x_i-x_j\|<\varepsilon$, or
 (ii) $\overline{x_ix_j}\in E \iff x_i$ is a KNN of $x_j$ or $x_j$ " " " of $x_i$.
 (iii) $\overline{x_ix_j}\in E$ has weight $\underbrace{}_{w_{ij}}$ equal to $\|x_i-x_j\|$.

3° Compute shortest path between any $x_i, x_j$:
The graph distance b/w $x_i,x_j\in\mathcal{X}$ is defined by $d_G(x_i,x_j):=\min\{L(\gamma)\mid \gamma$ is a path in $G$ from $x_i$ to $x_j\}$
where, if $\gamma=x_{i_1}-x_{i_2}-\cdots-x_{i_K}$,
then $L(\gamma):=\sum_{j=1}^{K-1} w_{i_ji_{j+1}}$.

**Rmk**:
Commonly, people use Dijkstra's algorithm to compute $d_G(x_i,x_j)$.

4° Denote $\Delta_G(\mathcal{X})=[d_G(x_i,x_j)]$.

Apply MDS on $\Delta_G(\mathcal{X})$ to obtain a low dimensional pt cloud
$\mathcal{Y}=\{y_1,\cdots,y_N\}\subseteq\mathbb{R}^d$ (w/ $\|y_i-y_j\|\approx d_G(x_i,x_j)$ $\forall i,j$).

Throughout this note,

$M$ : cpt connected $d$-dim. submfd of $\mathbb{R}^n$

(boundary and boundary corners are permitted).

**Def**

For $x, y \in M$, define

$\boxed{d_M(x,y)} := \inf\{L(\gamma) \mid \gamma : \text{admissible from } x \text{ to } y.\}$

**Def**

$\mathcal{X} = \{x_i\} \subseteq M$, a finite set.

$G$ : a graph on $\mathcal{X}$.

Define, $\forall\, x, y \in \mathcal{X}$,

$\boxed{d_G(x,y)} := \min\limits_{P} (\|x_0 - x_1\| + \cdots + \|x_{p-1} - x_p\|)$ and

$\boxed{d_S(x,y)} := \min\limits_{P} (d_M(x_0,x_1) + \cdots + d_M(x_{p-1},x_p))$,

where $P = (x_0, x_1, \cdots, x_p)$ is a path on $G$

w/ $x_0 = x$ and $x_p = y$.

**Goal:** We want to prove $d_G \approx d_M$.

**Approach:** $d_G \approx d_S$ and $d_S \approx d_M$.

**Prop 1**

$d_M(x,y) \le d_S(x,y)$ and

$d_G(x,y) \le d_S(x,y)$, $\forall\, x, y \in \mathcal{X}$.

<Pf>

The first is obvious.

The second comes from the fact that straight lines are shortest in $\mathbb{R}^n$. #

**Thm 2** ($d_M \approx d_S$).

$\varepsilon, \delta > 0$ w/ $4\delta < \varepsilon$. Suppose

① $G$ contains all edges $xy$ w/ $d_M(x,y) \le \varepsilon$

② $\forall\, m \in M$, $\exists\, x_i \in \mathcal{X}$ s.t. $d_M(x_i, m) \le \delta$.

Then, $\forall\, x, y \in \mathcal{X}$,

$$d_M(x,y) \le d_S(x,y) \le (1 + 4\delta/\varepsilon)\, d_M(x,y).$$

**Rmk:**

Condition ② is called the $\boxed{\delta\text{-sampling condition}}$.

<Pf>

"$\le$" of LHS follows from Prop 1.

Now consider "$\le$" on the RHS.

It suffices to prove the following claim.

**claim:** $\forall$ piecewise $C^\infty$ curve $\gamma$ from $x$ to $y$, w/ $\ell = L(\gamma)$, $\exists$ path $P = (x_0, \cdots, x_p)$ on $G$ from $x$ to $y$ s.t. $d_M(x_0,x_1) + \cdots + d_M(x_{p-1},x_p) \le (1+4\delta/\varepsilon)\ell$

(then the result follows from taking inf. among all such $\gamma$).

<Pf of claim>.

**case1** $\ell \le \varepsilon - 2\delta$. (Then $d_M(x,y) \le \ell \le \varepsilon$).

By condition ①, $xy \in G$.

$\Rightarrow d_S(x,y) \le d_M(x,y) \le \ell \le (1 + 4\delta/\varepsilon)\,\ell$.

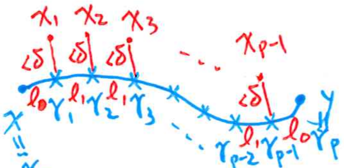
$(x,y)$ is a path in $G$.

**case2** $\ell > \varepsilon - 2\delta$.

By division algorithm,

$\ell = (\varepsilon - 2\delta)q + r$, w/ $q \in \mathbb{N} \cup \{0\}$ and $r \in [\varepsilon - 2\delta, 2(\varepsilon - 2\delta)]$

$= \ell_0 + \underbrace{(\ell_1 + \cdots + \ell_1)}_{q = p-2} + \ell_0$, w/ $\ell_1 = \varepsilon - 2\delta$ and $\ell_0 = r/2 \in [\frac{\varepsilon - 2\delta}{2}, \varepsilon - 2\delta]$




Divide $\gamma$ according to this length subdivision as in the picture. (Denote the pts by $\gamma_1, \cdots, \gamma_{p-1}$).

By condition ②, for $1 \le i \le p-1$, $\exists\, x_i \in \mathcal{X}$ s.t. $d_M(x_i, \gamma_i) \le \delta$.

Then $d_M(x_i, x_{i+1}) \le \delta + \ell_1 + \delta = \varepsilon = \ell_1 \cdot \frac{\varepsilon}{\varepsilon - 2\delta}$.

$d_M(x, x_1) \le \ell_0 + \delta$ and $d_M(x_{p-1}, y) \le \delta + \ell_0$.

Notice that $\frac{1}{2} \le \frac{\ell_0}{\varepsilon - 2\delta} \Rightarrow \delta \le \ell_0 \left(\frac{2\delta}{\varepsilon - 2\delta}\right)$

$\Rightarrow l_0 + \delta \le l_0 \left(\frac{\varepsilon}{\varepsilon - 2\delta}\right)$.

Thus $d_M(x, x_1) \le l_0 \cdot \left(\frac{\varepsilon}{\varepsilon-2\delta}\right)$ and

$$d_M(x_{p-1}, y) \le l_0 \cdot \left(\frac{\varepsilon}{\varepsilon - 2\delta}\right).$$

Hence,

$$d_M(x, x_1) + d_M(x_1, x_2) + \cdots + d_M(x_{p-1}, y)$$

$$\le \left(\frac{\varepsilon}{\varepsilon-2\delta}\right)(l_0 + l_1 + \cdots + l_1 + l_0) \le \frac{\varepsilon}{\varepsilon - 2\delta} \cdot l.$$

Notice that $\frac{1}{1-t} < 1 + 2t,\ \forall t \in (0, 1/2)$.

$$\frac{\varepsilon}{\varepsilon - 2\delta} = \frac{1}{1 - (2\delta/\varepsilon)}.$$

$2\delta/\varepsilon \in (0, 1/2)$ (due to $4\delta < \varepsilon$).

Thus $\frac{\varepsilon}{\varepsilon - 2\delta} \cdot l < (1 + 2 \cdot (2\delta/\varepsilon)) \cdot l = \left(1 + \frac{4\delta}{\varepsilon}\right) l$

Therefore, the claim follows. #

---

**Def**

The $\boxed{\text{minimum radius of curvature}}$ $\boxed{r_0 = r_0(M)}$ is defined by $\frac{1}{r_0} = \max_{\gamma, t} \{\|\ddot{\gamma}(t)\|\}$, where

$\gamma$ varies over all unit speed geod. and $t$ is in the domain of $\gamma$.

**Rmk:** (Intuition)

Geod. in $M$ "curl around less tightly" than circles of radii $r_0(M)$.

**Def**

The $\boxed{\text{minimum branch separation}}$ $\boxed{s_0 = s_0(M)}$ is

$$s_0 := \sup \{ s \mid d_M(x,y) \le \pi r_0 \ \ \forall x, y \in M \text{ w/ } \|x - y\| < s \}$$

**Rmk:**

Existence and positivity of $r_0$ and $s_0$ are guaranteed by compactness of $M$.

(For $s_0$, we may need Lebesgue number lemma).

Now, we use $r_0$ and $s_0$ to give a technical lemma.

$\boxed{\textbf{Lemma 3}}$

$\gamma$: a geod. in $M$ from $x$ to $y$.

$l := L(\gamma) \le \pi r_0$.

Then $2r_0 \sin(l/2r_0) \le \|x - y\| \le l$.

**Rmk:**

① By calculus, $\sin(t) \ge t - t^3/6,\ \forall t \ge 0$.

Thus, we can weaken the ineq. as:

$$\left(1 - \frac{l^2}{24 r_0^2}\right) l \le \|x-y\| \le l.$$

Thus, for small $l$, $l \approx \|x - y\|$.


② For $l \le \pi r_0$, $l/2r_0 \le \frac{\pi}{2}$.

Treating $t := l/2r_0$, $t \in (0, \frac{\pi}{2})$ and $\sin t \le \frac{2}{\pi} t$.

$$\Rightarrow 2r_0 \sin(l/2r_0) = 2 \cdot \frac{l}{2t} \sin t \ge 2 \cdot \frac{l}{2t} \cdot \frac{2}{\pi} t = (2/\pi) l.$$

Thus, we may weaken the ineq. as

$$(2/\pi) l \le \|x - y\| \le l.$$

⟨Proof of L3 is postponed in the appendix⟩.

$\boxed{\textbf{Cor 4} \text{ (Lemma for } d_S \approx d_G)}$

$\lambda > 0$.

$x_i, x_{i+1} \in M$.

Assume (i) $\|x_i - x_{i+1}\| < s_0$,

(ii) $\|x_i - x_{i+1}\| \le (2/\pi) r_0 \sqrt{24\lambda}$, and

(iii) $\exists$ geod. on $M$ of length $d_M(x_i, x_{i+1})$ connecting $x_i$ and $x_{i+1}$.

Then

$$(1 - \lambda) d_M(x_i, x_{i+1}) \le \|x_i - x_{i+1}\| \le d_M(x_i, x_{i+1}).$$

⟨Pf⟩

By (i) and def of $s_0$, $d_M(x_i, x_{i+1}) \le \pi r_0$.

By (iii), we can apply L3.

By ② in the Rmk above, $(2/\pi) l \le \|x_i - x_{i+1}\|$.

By (ii), we thus have $l \le r_0 \sqrt{24\lambda}$.

$\Rightarrow 1 - \lambda \le 1 - l^2/24 r_0^2$. By Rmk ① above, we are done. #

Def
$M$ is called **geodesically convex** if every $x, y \in M$ can be connected by a geod. of length $d_M(x,y)$.

**Main Thm A.**

$M$: cpt submfd of $\mathbb{R}^n$. (possibly w/ boundary)

$\mathcal{X} = \{x_i\}$ : finite set of data pts in $M$.

$G$ : a graph on $\mathcal{X}$.

$0 < \lambda_1, \lambda_2 < 1$.

$\varepsilon_{min}, \varepsilon_{max} > 0$, $\delta > 0$.

Assume

(i) $G$ contains all edges $xy$ w/ $\|x-y\| \le \varepsilon_{min}$.

(ii) All edges of $G$ have length $\|x-y\| \le \varepsilon_{max}$.

(iii) $\mathcal{X}$ satisfies $\delta$-sampling condition.

(iv) $M$ is geod. convex.

and

(v) $\varepsilon_{max} < s_0$, where $s_0$ : minimum branch sep. of $M$.

(vi) $\varepsilon_{max} \le (2/\pi)\, r_0 \sqrt{24\lambda_1}$,

where $r_0$ : minimum radius of curvature.

(vii) $\delta \le \lambda_2 \varepsilon_{min}/4$. (hence $4\delta < \varepsilon_{min}$).

Then

$$(1-\lambda_1)\, d_M(x,y) \le d_G(x,y) \le (1+\lambda_2)\, d_M(x,y),$$

$$\forall\, x, y \in \mathcal{X}.$$

<Pf>

Condition (i), (iii), (vii) are for applying Thm 2.

Condition (ii), (v), (vi), (iv) are for applying Cor 4.

Thus, $d_M \le d_S \le (1 + 4\delta/\varepsilon_{min})\, d_M \le (1+\lambda_2)\, d_M$

and $(1-\lambda_1)\, d_S \le d_G \le d_S$. (Thm 2)

Hence, $(1-\lambda_1)\, d_M \le (1-\lambda_1)\, d_S \le d_G \le d_S \le (1+\lambda_2)\, d_M$. #

(Thm 2) (Cor 4)