Manifold Learning

Distance preservation

Master Informatique

Parcours MALIA

Université Lumière Lyon 2
Université de Lyon

Plan

1. Metrics
2. MDS
3. Sammon's nonlinear mapping
4. Isomap

Metrics

Metrics

Definition (distance or metric)

A distance over the set $\mathcal{Y}$ is an application $d : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$ satisfying the following properties ($x, y, z \in \mathcal{Y}$)

$$
\begin{aligned}
d(x, y) &\geq 0 && \text{(positive definite)} \\
d(x, y) = 0 \text{ iff } x &= y && \text{(identity)} \\
d(x, y) &= d(y, x) && \text{(symmetry)} \\
d(x, y) &\leq d(x, z) + d(z, y) && \text{(triangle inequality)}
\end{aligned}
$$

Over $\mathbb{R}^D$, the most usual ones are derived from Minkowski (or L_p) norms (see Lab x001 of PC4DS).

Similarity

Definition (similarity)

A similarity over the set $\mathcal{Y}$ is an application $s : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$ satisfying the following property ($x, y \in \mathcal{Y}$)

$$s(x, y) \leq s(x, x)$$

A way to convert a similarity s into a distance d is by setting

$$d(x, y) = (s(x, x) + s(y, y) - 2s(x, y))^2$$

MDS

Multidimensional Scaling (MDS)

- Not a single method but a family of methods
- Aim: construct a configuration of points into a target metric (e.g. euclidean) space from interpoint distances
- Metric MDS : Young and Householder (1938), Torgerson (1952), Nonmetric MDS: Shepard (1962), Kruskal (1964)
- Several application on human sciences

	Amiens	Angers	Biarritz	Bordeaux	Brest	Calais	Cherbourg	Clermond_Fd	Dijon	Grenoble	Le Havre	Lille	Lyon	Marseille	Montpellier	Nancy	Nantes	Nice	Paris	Perpignan	Reims	Rennes	Rouen	Saint-Etienne	Strasbourg	Toulouse	Tours	Vichy
Amiens	0	369	862	679	619	159	366	524	417	674	180	113	587	910	880	358	489	1067	130	1043	167	411	117	604	498	827	352	485
Angers	369	0	518	335	371	494	290	402	498	644	297	503	538	842	692	602	90	999	303	762	454	126	282	530	742	552	106	399
Biarritz	862	518	0	183	817	997	808	555	815	820	800	975	735	697	534	1001	512	879	744	445	885	618	785	705	1141	308	510	609
Bordeaux	679	335	183	0	634	814	625	369	632	655	617	792	549	662	499	818	329	844	561	460	702	435	602	515	958	250	327	442
Brest	619	371	817	634	0	714	402	752	812	996	468	723	890	1194	1048	886	305	1351	581	1094	734	246	502	900	1026	884	456	761
Calais	159	494	997	814	714	0	461	672	556	840	275	109	753	1076	1046	469	584	1233	274	1209	278	506	212	800	609	967	487	633
Cherbourg	366	290	808	625	402	461	0	647	661	874	215	470	768	1072	998	635	310	1229	340	1068	481	204	249	860	795	858	351	639
Clermond_Fd	524	402	555	369	752	672	647	0	280	286	586	613	180	463	367	472	469	620	425	464	463	507	498	145	574	389	296	59
Dijon	417	498	815	632	812	556	661	280	0	284	542	471	197	520	490	192	588	677	323	653	278	567	454	240	309	664	392	220
Grenoble	674	644	820	655	996	840	874	286	284	0	776	755	106	277	286	476	713	337	557	449	562	751	688	140	505	535	538	265
Le Havre	180	297	800	617	468	275	215	586	542	776	0	284	676	980	950	511	357	1137	211	1062	320	269	88	710	651	852	290	547
Lille	113	503	975	792	723	109	470	613	471	755	284	0	668	991	961	384	593	1148	223	1124	193	515	221	736	524	940	465	574
Lyon	587	538	735	549	890	753	768	180	197	106	676	668	0	323	293	412	607	480	470	456	475	465	588	57	434	542	432	159
Marseille	910	842	697	662	1194	1076	1072	463	520	277	980	991	323	0	163	712	911	197	793	326	798	949	892	330	757	412	736	457
Montpellier	880	692	534	499	1048	1046	998	367	490	286	950	961	293	163	0	682	743	345	763	163	768	818	862	323	727	249	647	427
Nancy	358	602	1001	818	886	469	635	472	192	476	511	384	412	712	682	0	692	869	308	845	191	659	423	460	140	856	491	412
Nantes	489	90	512	329	305	584	310	469	588	713	357	593	607	911	743	692	0	1068	393	789	544	106	372	607	832	579	196	466
Nice	1067	999	879	844	1351	1233	1229	620	677	337	1137	1148	480	197	345	869	1068	0	950	508	955	1106	1049	427	842	594	893	614
Paris	130	303	744	561	581	274	340	425	323	557	211	223	470	793	763	308	393	950	0	926	155	360	123	517	448	681	234	343
Perpignan	1043	762	445	460	1094	1209	1068	464	653	449	1062	1124	456	326	163	845	789	508	926	0	931	895	974	481	498	210	717	590
Reims	167	454	885	702	734	278	481	463	278	562	320	193	475	798	768	191	544	955	155	931	0	511	232	537	331	806	375	424
Rennes	411	126	618	435	246	506	204	507	567	751	269	515	465	949	818	659	106	1106	360	895	511	0	294	712	836	685	211	516
Rouen	117	282	785	602	502	212	249	498	454	688	88	221	588	892	862	423	372	1049	123	974	232	294	0	640	563	764	275	459
Saint-Etienne	604	530	705	515	900	800	860	145	240	140	710	736	57	330	323	460	607	427	517	481	537	712	640	0	705	545	472	145
Strasbourg	498	742	1141	958	1026	609	795	574	309	505	651	524	434	757	727	140	832	842	448	498	331	836	563	705	0	976	631	520
Toulouse	827	552	308	250	884	967	858	389	664	535	852	940	542	412	249	856	579	594	681	210	806	685	764	545	976	0	507	443
Tours	352	106	510	327	456	487	351	296	392	538	290	465	432	736	647	491	196	893	234	717	375	211	275	472	631	507	0	293
Vichy	485	399	609	442	761	633	639	59	220	265	547	574	159	457	427	412	466	614	343	590	424	516	459	145	520	443	293	0

Classical metric multidimensional scaling

Modèle : $y = \mathbf{W}x$

- $y \in \mathbb{R}^D$ observed
- $x \in \mathbb{R}^s$ latent (unobserved)
- $\mathbf{W}$ unknown $D \times s$ matrix W such that $W^T W = I$
- Call $\mathbf{Y} = [y_1, \dots y_i, \dots y_N]$ the (coordinates) data matrix ($D \times N$ constructed by concatenation of y_i's)

Definition (Gramm's matrix)

$$\mathbf{S} = \mathbf{Y}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X}$$

Spectral decomposition of Gramm matrix

Since $\mathbf{S}$ is a $N \times N$ square symmetric matrix, we have (through the SVD of $\mathbf{Y}$)

$$
\begin{aligned}
\mathbf{S} =& \mathbf{U}\Lambda\mathbf{U}^T \\
=& (\mathbf{U}\Lambda^{1/2})(\Lambda^{1/2}\mathbf{U}^T) \\
=& (\Lambda^{1/2}\mathbf{U}^T)^T(\Lambda^{1/2}\mathbf{U}^T)
\end{aligned}
$$

where $\mathbf{U}$ is a $N \times N$ orthonormal matrix and Λ is a diagonal matrix with the eigenvalues arranged in decreasing order.

Then, the configuration set of points is

$$\breve{\mathbf{X}} = \mathbf{I}_{\mathrm{s}\times\mathrm{N}}\Lambda^{1/2}\mathbf{U}^T$$

If coordinates are not available

- When coordinates are available: equivalence between from MDS & PCA
- If they are not, MDS can still be used from distance matrix $\mathbf{D}$ where the general term is $d(y_i, y_j)^2$
- Then, we obtain $\mathbf{S}$ from $\mathbf{D}$ using

$$\mathbf{S} = -\frac{1}{2}\left\{\mathbf{D} - \frac{1}{N}\mathbf{D}1_N1_N^T - \frac{1}{N}1_N1_N^T\mathbf{D} + \frac{1}{N^2}1_N1_N^T\mathbf{D}1_N1_N^T\right\}$$

The MDS algo

1. If available data consists of vectors gathered in $\mathbf{Y}$, then center them, compute the pairwise scalar products $\mathbf{S} = \mathbf{Y}^T\mathbf{Y}$, and go to step 3.
2. If available data consists of pairwise Euclidean distance, transform them into scalar products :
   - Square the distances and build $\mathbf{D}$
   - Perform the double centering of $\mathbf{D}$, this yields $\mathbf{S}$
3. Compute the eigenvalue decomposition $\mathbf{S} = \mathbf{U}\Lambda\mathbf{U}^T$
4. A $s-$dimensional representation is obtained by computing the product $\breve{\mathbf{X}} = \mathbf{I}_{\mathbf{s}\times\mathbf{N}}\Lambda^{1/2}\mathbf{U}^T$

Two benchmark manifolds

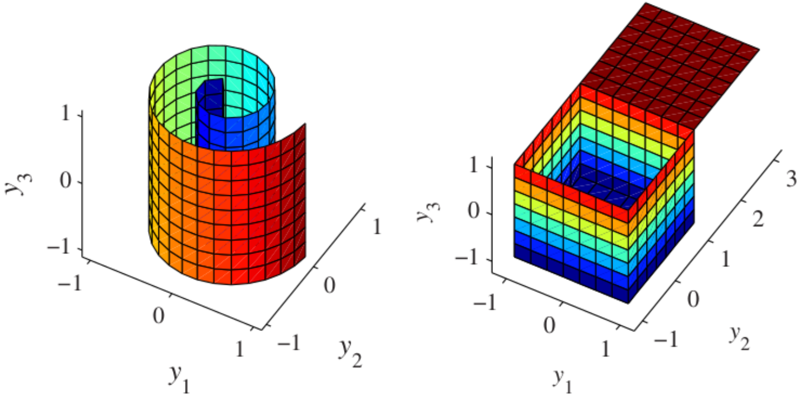
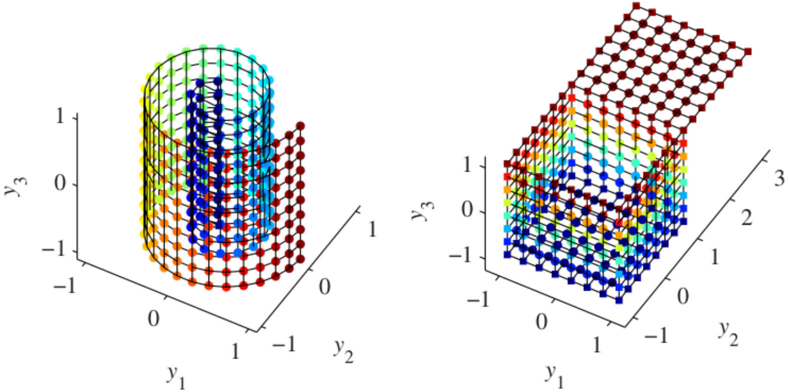


Fig. 1.3. Two benchmark manifold: the 'Swiss roll' and the 'open box'.

Sampled versions



MDS-based projection on the plan

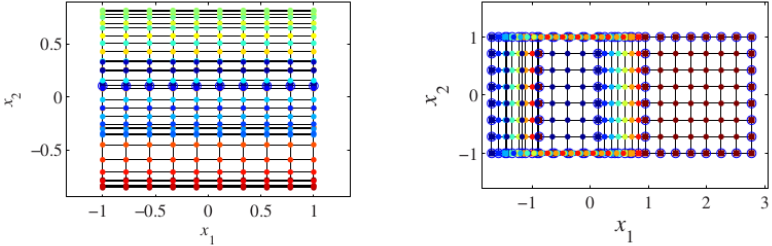


Fig. 4.2. Two-dimensional embeddings of the "Swiss roll" and "open box" data sets (Fig. 1.4), found by metric MDS.

Metric & nonparametric MDS

Metric MDS : preserve distances (instead of $< ., . >$)

Stress function (used as objective function)

$$E_{\mathrm{mMDS}} = \frac{1}{2} \sum_{i,j=1}^{N} w_{i,j} (d_y(i,j) - d_x(i,j))^2$$

Nonparametric MDS : introduce disparities

$$E_{\mathrm{nMDS}} = \sqrt{\frac{\sum_{i,j=1}^{N} w_{i,j} |f(\delta_{i,j}) - d_x(i,j)|}{c}}$$

Sammon's nonlinear mapping

Stress function

- Related to metric MDS, looks for configuration minimizing

$$E_{\mathrm{NLM}} = \frac{1}{c} \sum_{i=1, i<j}^{N} \frac{(d_y(i,j) - d_x(i,j))^2}{d_y(i,j)}$$

  where $c = \sum_{i=1,i<j}^{N} d_y(i,j)$ and d_x will be Euclidean.
- Optimization problem solved with quasi-Newton optimization
- Initialize with MDS, then use the update rule (α is call the magic number)

$$x_k(i) \leftarrow x_k(i) - \alpha \frac{\frac{\partial E_{\mathrm{NLM}}}{\partial x_k(i)}}{\left| \frac{\partial^2 E_{\mathrm{NLM}}}{\partial x_k(i)^2} \right|}$$

MDS-based projection on the plan

1. Compute all pairwise distances $d_{\mathbf{y}}(i,j)$ in the D-dimensional data space.
2. Initialize the P-dimensional coordinates of all points $\mathbf{x}(i)$, either randomly or on the hyperplane spanned by the first P principal components of the data set (after PCA or MDS).
3. Compute the right-hand side of Eq. (4.68) for the coordinates of all points $\mathbf{x}(i)$.
4. Update the coordinates of all points $\mathbf{x}(i)$.
5. Return to step 3 until the value of the stress function no longer decreases.

MDS-based projection on the plan

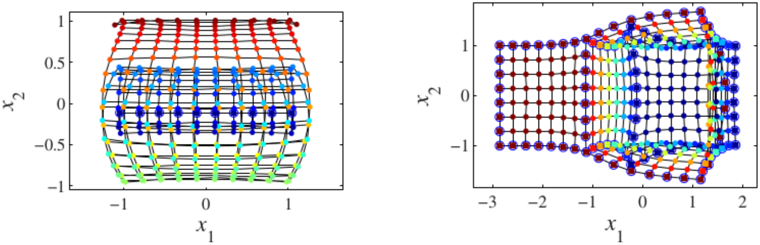


Fig. 4.4. Two-dimensional embeddings of the 'Swiss roll' and 'open box' data sets (Fig. 1.4), found by Sammon's NLM.

Isomap

How to de generalize PCA or MDS?

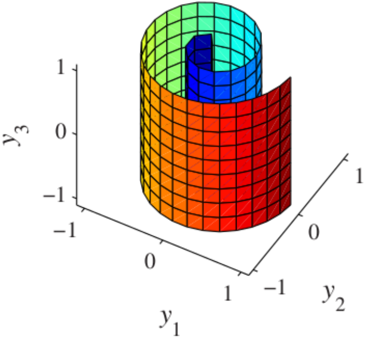
Aim

- Avoid globally *linearity*, *euclidean distances* or *normality*
  + Find a new representation where linearity is a reasonable assumption
  + a fully non parametric approach is desireable

3 approaches to be studied (partially) based on graphs

- ISOMAP : extension of MDS to geodesic
- LLE (Local linear embeddings) : conserve local geometry
- Laplacian Eigenmaps : based on spectral techniques

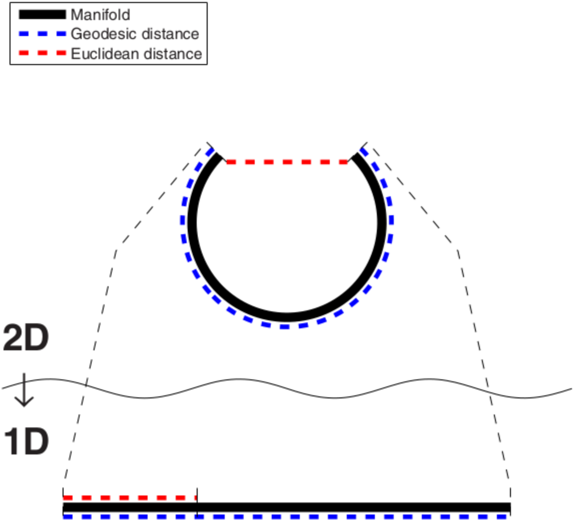
Global distance and manifolds identifiées par uns



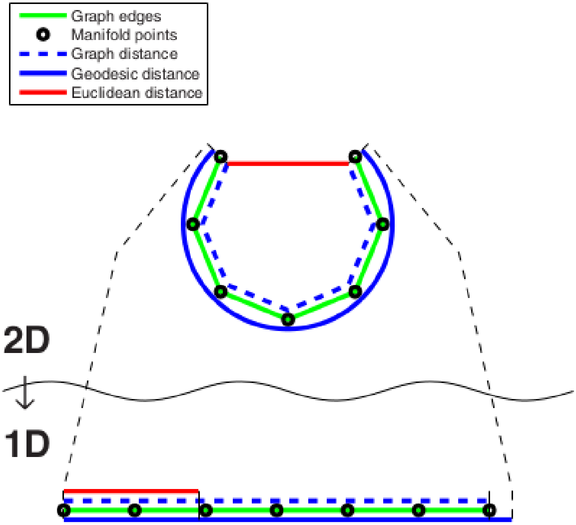
Recall:
– we need a more general notion than euclidean distance
– global vs local

A *manifold* is a space regular enough so that LOCALLY has all the properties of a linear space. Its *dimension* can be view as as the number of free coordinates needed to navigate along the manifold.

Geodesic distance



Graph distance



Isomap

KEY IDEA perform a non metric MDS only replacyng the euclidean by the geodesic distance

1. Build a graph with either the K-rule or the ϵ-rule.
2. Weight the graph by labeling each edge with its Euclidean length.
3. Compute all pairwise graph distances with Dijkstra's algorithm, square them, and store them in matrix $\mathbf{D}$.
4. Convert the matrix of distances $\mathbf{D}$ into a Gram matrix $\mathbf{S}$ by double centering.
5. Once the Gram matrix is known, compute its spectral decomposition $\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$.
6. A P-dimensional representation of $\mathbf{Y}$ is obtained by computing the product $\hat{\mathbf{X}} = \mathbf{I}_{P\times N}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T$.

Isomap on benchmark