UMAP for the working mathematician

Rick Jardine

University of Western Ontario

December 14, 2021

UMAP: "Uniform Manifold Approximation and Projection for Dimension Reduction"

Healy-McInnes (2018),
Healy-McInnes-Melville (2020) [3]

Outline

- Spivak's extended pseudo-metric spaces (ep-metric spaces)
- TDA constructions in ep-metric spaces
- weighted graphs
- classical dimension reduction (PCA)
- optimisation on low dimensional graph: fuzzy sets, cross entropy

ep-metric spaces (Spivak, 2009)

An extended pseudo-metric space (**ep-metric space**) (X, D) is a set X and a function $D : X \times X \to [0, \infty]$ such that

1) $D(x, x) = 0$,
2) $D(x, y) = D(y, x)$,
3) $D(x, z) \leq D(x, y) + D(y, z)$.

- Can have distinct x, y such that $D(x, y) = 0$ ("pseudo").
- Can have u, v such that $D(u, v) = \infty$ ("extended").

Example: Every metric space (Y, d) is an ep-metric space:

$$Y \times Y \xrightarrow{d} [0, \infty) \subset [0, \infty].$$

A **morphism** $f : (X, d_X) \to (Y, d_Y)$ of ep-metric spaces is a function $f : X \to Y$ such that

$d_Y(f(x), f(y)) \leq d_X(x, y)$ (compresses distance, "non-expanding").

$ep - \underline{\text{Met}}$ is the category of ep-metric spaces and their morphisms.

Cocompleteness

(X, d) an ep-metric space and $p : X \to Y$ a surjective function.

For $x, y \in Y$ set

$$D(x, y) = \inf_P \, [\sum_{i=0}^{k} d(x_i, y_i)]$$

"Polygonal path" P : pairs (x_i, y_i), $0 \leq i \leq k$, in X with $x = p(x_0)$, $p(y_i) = p(x_{i+1})$, $y = p(y_k)$.

Quotient map $p : (X, d) \to (Y, D)$ satisfies universal property.

Lemma 1.

$ep - \underline{\text{Met}}$ is cocomplete (has all small colimits).

Underlying set is colimit in sets: $\sqcup_i \, (X_i, d_i)$ has $D(x, y) = d_i(x, y)$ if x, y in some X_i, $D(x, y) = \infty$ otherwise.

Coequalizers (or pushouts) given by quotient construction.

Vietoris-Rips complex

(X, d) a finite **ep-metric space**, $d : X \times X \to [0, \infty]$.

$s \in [0, \infty]$:

$P_s(X)$ is the poset of all subsets $\sigma \subset X$ such that $d(x, y) \leq s$ for all $x, y \in \sigma$.

$P_s(X)$ defines an abstract simplicial complex $V_s(X)$ – Vietoris-Rips complex.

Nerve $BP_s(X)$ is the **barycentric subdivision** of $V_s(X)$.

There is a natural weak equivalence

$$\gamma : BP_s(X) \xrightarrow{\simeq} V_s(X)$$

defined by the last vertex map, subject to a total ordering on X.

NB: Poset const. $BP_s(X)$ **does not** use a total ordering on X.

Comments

Write $V_s(X) = BP_s(X)$, or $V_s(X,d) = BP_s(X,d)$.

1) The systems $s \mapsto H_p V_s(X,d)$ (coefficients in a field) define **persistent homology** for (X,d).

2) The sets $\pi_0 V_s(X,d)$ are **clusters** for (X,d).

Hierarchy $\Gamma(X,d)$: tree with vertices $(s,[x])$ with $s \leq t$, $[x] \in \pi_0 V_s(X)$. Source of hierarchical clustering algorithms.

3) One cares most about

$$\varinjlim_{s<\infty} V_s(X,d),$$

which is a disjoint union of **global components**, each of which consists of simplices having edges of finite length.

Each global component is contractible, filtered by distance d.

UMAP complex

(X, d) = finite data set, or finite ep-metric space ... but very big.

1) Choose **neighbourhood set** N_x for each $x \in X$.
Set $U_x = \{x\} \sqcup N_x \subset X$.

e.g. $N_x = k$ nearest neighbours (if you can find them — there are algorithms).

2) Set $(U_x, D_x) = \vee_{y \in N_x} (\{x, y\}, d_y)$ (wedge of rays) in $ep-\text{Met}$.
$d_y(x, y) > 0$ is a **weight** — original distance, or adjustment.

3) Extend to an ep-metric D_x on X by setting $D_x(y, z) = \infty$ if either y or z is outside of U_x.

4) We have inclusions $X \subset V(X, D_x)$, $x \in X$ (X discrete). Form the iterated pushout

$$V(X, N) = \vee_{x \in X} V(X, D_x).$$

$V(X, N)$ is "the" **UMAP complex** — a diagram of simp. sets defined on $[0, \infty]$

"Fuzzy graph" for ep-metric space (X, d)

$\Gamma_s(X)$ has elements of X as vertices, and edges $\{x, y\}$, 2-elt subsets of X , with $d(x, y) < s$.

The functor $\Gamma_*(X)$: $s \mapsto \Gamma_s(X)$ is the "fuzzy graph" for (X, d) (functor on $[0, \infty]$).

$\Gamma_*(X)$ is the non-degenerate part of $\text{sk}_1 B\check{C}(X)$.

$\check{C}(X)$ is **Čech groupoid**: X = objects, and there is a morphism $x \to y$ iff $d_X(x, y) < \infty$. $B\check{C}(X)_s$: $d_X(x, y) \leq s$.

Functor $\check{C}(X)_s \to \pi BP_s(X)$ (fund. groupoid) takes a morphism $x \to y$ to the composite $\{x\} \to \{x, y\} \leftarrow \{y\}$

Theorem 2.

There are natural bijections

$$\pi_0\Gamma_s(X) \cong \pi_0 B\check{C}(X)_s \xrightarrow{\cong} \pi_0\pi BP_s(X) \cong \pi_0 BP_s(X).$$

The fuzzy graph $\Gamma_*(X)$ computes clusters for (X, d).

Excision

Given (X, d_X) and (X, D_x) as above, form the wedge sum

$$(X, D) = \vee_{x \in X} (X, D_x)$$

in ep-metric spaces. (X, D) is the "*realization*" of $V(X, N)$.

Theorem 3 (Excision for π_0).

$V(X, N) = \vee_x V(X, D_x) \to V(X, D)$ *induces a bijection*

$$\pi_0 V(X, N)_s \xrightarrow{\cong} \pi_0 V(X, D)_s, 0 \leq s < \infty.$$

Paths in both are hops through nearest neighbours.

Corollary 4.

There are bijections

$$\pi_0 \Gamma(X, D)_s \cong \pi_0 V(X, D)_s \cong \pi_0 V(X, N)_s, \; 0 \leq s < \infty.$$

Dimension reduction step (PCA)

We have the weighted graph $\Gamma(X, D)$, where (X, D) is a finite ep-metric space, $|X| = N$.

Assume (X, D) is connected: $D(x, y) < \infty$ for all $x, y \in X$.

The elements x_i define standard basis vectors $e_i \in \mathbb{R}^N$.

Idea: Find function $p : X \to \mathbb{R}^d$ with d small, such that $d(p(x_i), p(x_j))$ is "close" to weight $w_{i,j} = D(x_i, x_j)$ for all i, j.

- Weight matrix $W = (w_{i,j})$ for $i \neq j$,
- Diagonal matrix $D = (d_{i,i})$ with $d_{i,i} = \sum_{i \neq j} \ w_{i,j}$.
- $L = D - W$ is the **weighted Laplacian** for the graph $\Gamma(X)$.

Find orthonormal basis of eigenvectors $\{w_1, \dots, w_N\}$ for L, and choose eigenvectors $\{w_1, \dots, w_d\}$ with associated eigenvalues λ_i satisfying $\lambda_1 \geq \lambda_2 \geq \dots$. Usually, $d = 2, 3$.

$e_i = \alpha_{i,1} w_1 + \cdots + \alpha_{i,N} w_N$. Set $p(x_i) = \alpha_{i,1} w_1 + \cdots + \alpha_{i,d} w_d$.

What we have, so far

1) Start with a data set X in a metric space.

2) Choose nearest neighbours N_x, $x \in X$, and and use these to construct an ep-metric space structure (X, D), with corresponding weighted graph $\Gamma(X, D)$.

3) Use classical dimension reduction method to find low dimensional graph $Y \subset \mathbb{R}^d$ with vertices $y_i = p(x_i)$ with actual distances $d(y_i, y_j)$ approximating the weights $D(x_i, x_j)$.

Next step: improve on the choice of points $y_i \in \mathbb{R}^d$ to better approximate the weights.

This is done with fuzzy set methods.

Fuzzy sets

Classical definition: A **fuzzy set** is a function $p : X \to [0, 1]$,

Barr, 1986 [1]: A **fuzzy set** with coefficients in a **locale** L is a function $\phi : X \to L$.

Examples: $L = [0, \infty], [0, \infty]^{op}, [0, 1], [0, 1]^{op}$.

A **morphism** $\phi \to \psi$ of fuzzy sets is a function $f : X \to Y$ such that in the picture

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {\scriptstyle\phi}\searrow \quad \swarrow{\scriptstyle\psi} & \\ & L & \end{array}$$

$\phi(x) \leq \psi(f(x))$ for all $x \in X$. i.e. diagram hypy. commutes. Fuzz(L) is corr. category.

Barr: A fuzzy set $X \to L$ is a sheaf (of monomorphisms) on L_+. (L with new inital element $+$).

Example: (X, d) finite ep-metric space. $V(X, d)$ is a simplicial sheaf on $[0, \infty]^{op}_+$, or simplicial fuzzy set.

Complement (classical)

The function $s \mapsto 1 - s =: e(s)$ defines a poset isomorphism (idempotent) $e : [0,1] \to [0,1]^{op}$, called the **dual**.

Classical: If $p : X \to [0,1]$ is a fuzzy subset of X, its **complement** p^c is defined by $p^c(x) = 1 - p(x) = e(p(x))$.

Example: $A \subset X$ has a step function $p_A : X \to [0,1]$ and $e \cdot p_A = p_{X-A}$.

We have poset isomorphisms

$$\phi : [0,1] \overset{\cong}{\leftrightarrows} [0,\infty]^{op} : \psi$$

with $\phi(s) = -\log(s)$ and $\psi(t) = e^{-t}$ (order reversing).

ϕ is Shannon's **information function** (Shannon, 1948)

$$\begin{array}{ccc} [0,\infty] & \xrightarrow{\psi} & [0,1]^{op} \\ {\scriptstyle e}\downarrow & & \downarrow{\scriptstyle e} \\ [0,\infty]^{op} & \xrightarrow[\psi]{} & [0,1] \end{array} \qquad (1)$$

defines a duality isomorphism (idempotent) $e : [0,\infty] \to [0,\infty]^{op}$.

$$x^* := e(x) = -\log(1 - e^{-x}) = x - \log(e^x - 1),\ \ x \in [0,\infty].$$

Facts: $x \mapsto x^*$ is continuous. x and x^* are close if x is near $\log(2)$.

Cross entropy (standard)

Suppose that $u, v : X \to [0,1]$ are fuzzy sets.

The **cross entropy** (Bhandari 1993 [2]) from v to u is defined by

$$C(u,v) = \sum_{x \in X} (u(x) \log(\frac{u(x)}{v(x)}) + (1-u(x)) \log(\frac{1-u(x)}{1-v(x)}))$$

$C(u,v)$ is a sum of elements of the form

$$F(s,t) = s \cdot \log(\frac{s}{t}) + (1-s) \cdot \log(\frac{1-s}{1-t}) \geq 0.$$

with $s, t \in [0,1]$ (Kullback-Liebler divergence, 1951).

u (hence s) is fixed. $C(u,v)$ is a (non-symmetric) measure of the distance of v from u.

Cross entropy (new)

Suppose $s = e^{-y}$ and $t = e^{-x}$.

$$F(s,t) = F(e^{-y}, e^{-x}) = e^{-y}(x-y) + e^{-y^*}(x^* - y^*).$$

$|F(e^{-y}, e^{-x})|$ depends on $|x-y|$, $|x^* - y^*|$ in $[0,\infty]$.

Given $u, v : X \to [0,\infty]$, set

$$\begin{aligned} C(u,v) &= C(\psi \cdot u, \psi \cdot v) \\ &= \sum_{x \in X} [e^{-u(x)}(v(x) - u(x)) + e^{-u(x)^*}(v(x)^* - u(x)^*)]. \end{aligned}$$

For $u, v : X \to [0,\infty]$, $C(u,v)$ is the **cross entropy** from v to u.

Optimisation

Start with $\Gamma(X)$, run PCA to get a graph (Y, E) with vertices $y_i = p(x_i)$, and edges $[y_i, y_j]$ with weights $d(y_i, y_j)$ given by distance.

Adjust vertices of (Y, E) by minimizing cross entropy:

$$w_X : X \times X \to [0, \infty],\ (x, y) \mapsto D(x, y),$$
$$w_Y : X \times X \to [0, \infty],\ (x, y) \mapsto d(p(x), p(y)).$$

$$C(w_X, w_Y) = \sum_{x,y} [e^{-w_X(x,y)}(w_X(x, y) - w_Y(x, y)) + e^{-w_X(x,y)^*}(w_X(x, y)^* - w_Y(x, y)^*)]$$

Suggestion: Minimise $C(w_X, w_Y)^2$ by moving the y_i in directions of "negative slope" (stochastic gradient descent).

Endgame

One finishes with a graph (Y, E), $X = Y \subset \mathbb{R}^d$, which is "optimally close" to (X, D).

We're not done:

Write $(Y, E)_s$ for the subgraph of (Y, E) with vertices $Y = X \subset \mathbb{R}^d$ and edges $[x, y] \in E$ with $d(x, y) \leq s$.

- We have clusters $\pi_0(Y, E)_s$ and a hierarchy $\Gamma(Y, E)$.

- $\Gamma(Y, E)$ is a tree with objects $(s, [x])$ with $[x] \in \pi_0(Y, E)_s$, and morphisms $(s, [x]) \to (t, [x])$ with $s \leq t$.

References

Michael Barr.
Fuzzy set theory and topos theory.
Canad. Math. Bull., 29(4):501–508, 1986.

Dinabandhu Bhandari and Nikhil R. Pal.
Some new information measures for fuzzy sets.
Inform. Sci., 67(3):209–228, 1993.

Leland McInnes, John Healy, and James Melville.
Umap: Uniform Manifold Approximation and Projection for Dimension Reduction.
Preprint, arXiv: 1802.03426 [stat.ML], 2020.

D.I. Spivak.
Metric realization of fuzzy simplicial sets.
Preprint, 2009.

Graphs from simplicial sets

Z = simplicial set. There is a π_0-isomorphism

$$N\,\mathrm{sk}_1\,Z = \mathrm{sk}_0\,Z \cup (\sqcup_{\sigma:\Delta^1\to Z,\ d_0\sigma\neq d_1\sigma}\ \Delta^1) \to Z.$$

$\Gamma(Z)$ is the unoriented graph with vertices Z_0 with edges given by boundaries $\{d_0\sigma, d_1\sigma\}$ for $\sigma : \Delta^1 \to Z$ with $d_0\sigma \neq d_1\sigma$.

$$\pi_0\Gamma(Z) = Z_0/\sim,$$

where $x \sim y$ if $\{x, y\}$ is an edge.

There are bijections

$$\pi_0\Gamma(Z) \xleftarrow{\cong} \pi_0 N\,\mathrm{sk}_1\,Z \xrightarrow{\cong} \pi_0 Z,$$

natural wrt inclusions $Z \subset Z'$ of simplicial sets.