The mathematics of UMAP

Adele Jackson

July 31, 2019

1 Introduction

UMAP (Uniform Manifold Approximation and Projection) is a new dimension reduction technique [2], currently implemented [5, 7] for both labelled and unlabelled data. Figure 1 shows a comparison of UMAP embeddings with the outputs of some other standard dimension reduction algorithms. UMAP gives similarly good outputs for visualisation as t-SNE, with a substantially better runtime (see [4]), and may capture more of the global structure of the data. The current UMAP implementation also allows UMAP to be used as a preprocessing step in a machine learning pipeline, as new data can be embedded into an existing model.

Excitingly, UMAP is based on strong and general mathematical theory, which does not depend on a Euclidean metric. As a result, UMAP can be used for datasets with a mixture of categorical and continuous features.

The main mathematical result behind UMAP is that there is an adjunction between finite fuzzy simplicial sets and finite extended pseudo-metric spaces. In this document, we will explain what this statement means and why it is useful.

Given a dataset D in $\mathbb{R}^N$, we wish to find a good representation of D in a lower-dimensional space, $\mathbb{R}^m$. We wish to develop a good loss function for evaluating a lower-dimensional representation, so we can minimise this loss. To do this, we will construct a fuzzy simplicial set from a dataset. A fuzzy simplicial set (which we will define in Section 4) is an abstract description of a topological space, with probabilities on its elements that we will interpret as giving their size. We will build a partially-defined metric space relative to each vertex, then use the adjunction to convert this family of metric spaces to a family of fuzzy simplicial sets. In the algorithm, this step corresponds to constructing the adjacency matrix A for a weighted directed graph [2, Section 3.1]. We then take a union of this family, which corresponds to constructing the symmetric matrix $B = A + A^T - A \circ A^T$ where $\circ$ is the Hadamard product. We can then compare the fuzzy simplicial set from our dataset D with that from a proposed low-dimensional representation, to find the low-dimensional representation that minimises the cross entropy loss between the two fuzzy sets.

2 Approximating the underlying manifold

Given a dataset D in $\mathbb{R}^N$, we think of the datapoints as being drawn from some Riemannian manifold[1] M, then mapped into $\mathbb{R}^N$ by some injective map $\phi : M \to \mathbb{R}^N$. See Figure 2 for an illustration of this setup.

One approach to finding a low-dimensional representation of D is to reconstruct M, then find a good map from M into $\mathbb{R}^m$. To do this, we assume that D is uniformly drawn from M, as in the example in Figure 2. (Note that parts of M might be stretched out or compressed under the map ϕ into $\mathbb{R}^N$, so this does **not** imply that the data is uniformly distributed in $\mathbb{R}^N$.) This

[1] A Riemannian manifold is a space that locally looks like Euclidean space, in which we have well-defined notions of distances, angles, and volumes. For example, the surface of a unit sphere in $\mathbb{R}^3$ is a two-dimensional Riemannian manifold.

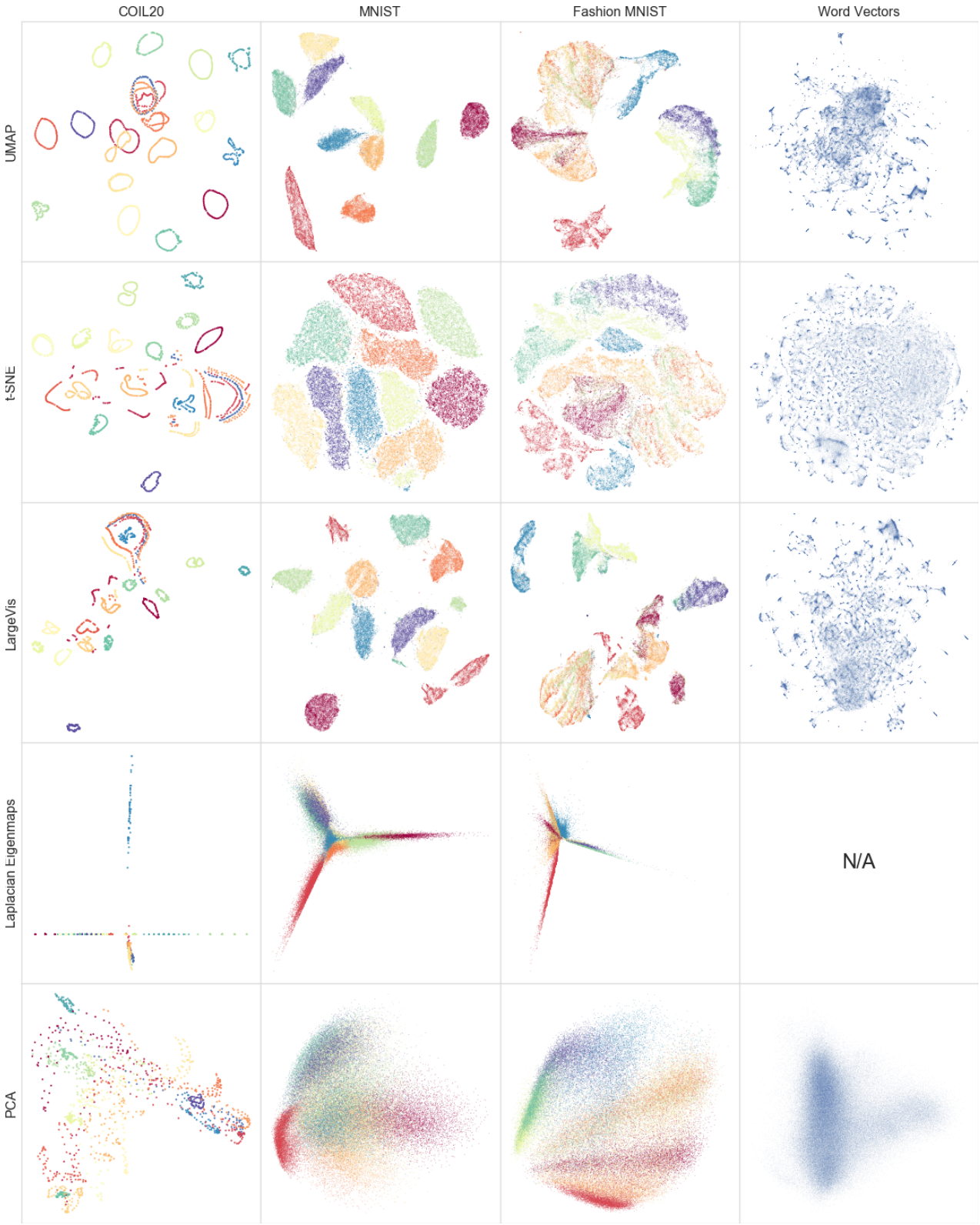
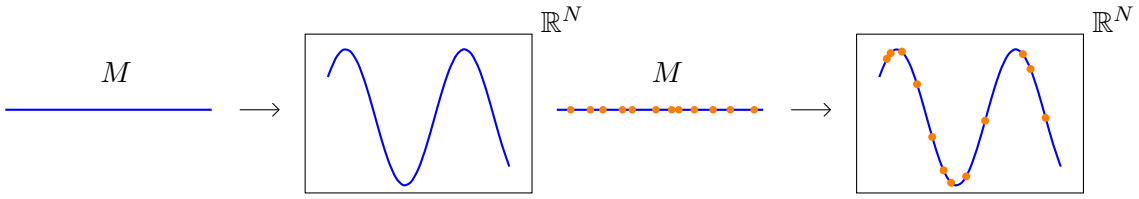
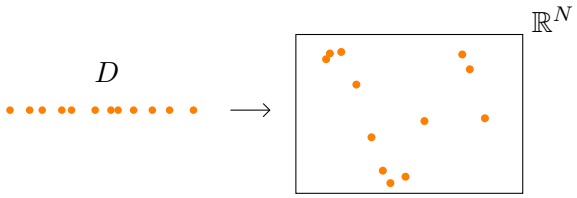


Figure 1: Embeddings from UMAP, t-SNE, LargeVis, Laplacian Eigenmaps and PCA on some standard datasets. Image taken from [2].



(a) The manifold M embedded in $\mathbb{R}^N$. (b) The dataset D is uniformly drawn from M.



(c) We see only D.

Figure 2: The dataset is drawn from a manifold embedded in $\mathbb{R}^N$.

assumption means that D approximates M well. (This is also where the 'U' in UMAP comes from.)

Note that we have two notions of distance between our datapoints: the distance from $\mathbb{R}^N$, and the "true" distance in M. To reconstruct M, we wish to reconstruct this second distance. We now formally define the mathematical object we wish to reconstruct.

Definition 1. A *metric* on a set of points X is a function $d : X \times X \to \mathbb{R}$ that satisfies the following conditions. First, the metric is non-negative: $d(x, y) \geq 0$ for all $x, y \in X$. Second, the function is symmetric: $d(x, y) = d(y, x)$ for all $x, y \in X$. Third, the function distinguishes points: $d(x, y) = 0$ if and only if $x = y$. Finally, a metric satisfies the triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

The words "metric" and "distance function" are interchangeable. A *metric space* is a the pair (X, d), where X is a set of points and d is a metric for X.

If we assume that the metric on M is locally constant, we get the following lemma. We will show that this lemma allows us to approximate the distance between two points $x, y \in D$ as measured in M, so long as x and y are close enough in $\mathbb{R}^N$, by normalising by the distance from x to its k-nearest neighbour.

Lemma ([2])**.** *Let (M, g) be a Riemannian manifold*[2] *embedded in $\mathbb{R}^n$. Let $p \in M$ be a point. Suppose that g is locally constant.*[3] *Let B be a ball in M, containing p, whose volume is $\frac{\pi^{n/2}}{\Gamma(n/2+1)}$ with respect to the metric on M. Then the distance of the shortest path in M from p to a point $q \in B$ is $\frac{1}{r} d_{\mathbb{R}^n}(p, q)$, where r is the radius of B in $\mathbb{R}^n$ and $d_{\mathbb{R}^n}(p, q)$ is the distance from p to q in $\mathbb{R}^n$.*

We approximate distances in M between nearby datapoints as follows. Fix a radius R. As the dataset D is uniformly distributed in M, the expected number of datapoints in any radius R ball in M is some constant k. Now, for $x \in D$, let $N_k(x)$ be the ball in $\mathbb{R}^N$, centred at x, that contains its k nearest neighbours. Let $\tilde{N}_k(x)$ be the preimage of $N_k(x)$ in M (this is $\phi^{-1}(N_k(x))$). If k is small enough that g is locally constant on $\tilde{N}_k(x)$, then this preimage will

[2] In this, M is the manifold and g is a two-form that gives us measures of distance, volumes and angles.

[3] To be precise, we assume that there is an open neighbourhood U with $p \in U$ on which g is locally constant, such that g is a constant diagonal matrix in the ambient coordinates.

be a ball of some radius, and the expected value of this radius will be R. By the lemma we can reconstruct distances within $N_k(x)$ by normalising by $\frac{R}{r_x}$ where r_x is the distance from x to its k-nearest neighbour As we only care about distances up to a constant factor, it suffices to normalise by r_x in each of the $N_k(x)$ for $x \in D$.

Thus, by the lemma, we can approximate distances from x to points in $N_k(x)$ as follows. Fix a value of k. Now we can calculate the distance in M, based at x_i, from x_i to x_j to be approximately

$$\frac{1}{r_i} d_{\mathbb{R}^N}(x_i, x_j),$$

where r_i is the distance from x_i to its k^{th}-nearest neighbour. To reduce the impact of noise in the distribution of datapoints on the value of r_i, in practice we take r_i to be the value such that

$$\sum_{j=1}^{k} \exp\left(\frac{-|x_i - x_{i_j}|}{r_i}\right) = \log_2(k)$$

where $\{x_{i_1}, \ldots, x_{i_k}\}$ are the k nearest neighbours of x_i.

We now modify these distances in two ways. Note that the distance in M, based at x_i, from x_j to x_k for $j, k \neq i$ is not defined. We set this to ∞ to signal this. To avoid having isolated points, we also assume the manifold is locally connected, so has no isolated points, and that there are enough datapoints that no datapoint is in its own connected component. To enforce this, let ρ_i be the distance from x_i to its nearest neighbour. Then we set

$$d_{x_i}(x_i, x_j) = \frac{1}{r_i} \max(0, d_{\mathbb{R}^N}(x_i, x_j) - \rho_i).$$

The effect of this modification is to force the distance from x_i to its nearest neighbour to be 0.

Following this process gives us an *extended pseudo-metric space* for each choice of x_i. We get an extended pseudo-metric, rather than a metric, as some of our distances are infinite and some distances between distinct points are zero.

Definition 2. An *extended pseudo-metric space* is a set X with a function $d : X \times X \to \mathbb{R} \cup \{\infty\}$ satisfying the following conditions. First, the metric is non-negative and symmetric. Second, $d(x, x) = 0$ for all x. Finally, let $x, y, z \in X$. Then either $d(x, z) = 0$, or $d(x, z) \leq d(x, y) + d(y, z)$.

The word "pseudo" refers to the weakening of the condition that $d(x, y) = 0$ if and only if $x = y$, and "extended" refers to the addition of the value ∞.

This setup presents us with a problem. The distance we get from x_i to x_j using this method will in general be different to that from x_j to x_i, as $r_j \neq r_i$. This is due to the fact that, while the x_i are uniformly drawn from the manifold, they are not all the same distance apart. We need a technique for combining a family of locally-defined finite metric spaces to get a global structure. This is where we will use the adjunction between extended pseudo-metric spaces and fuzzy simplicial sets.

Note that, although we have used the $\mathbb{R}^N$ metric to develop this theory, the idea still holds with any other metric. As a result, UMAP can be used with custom distance measures and can handle categorical data and other measures of distance between datapoints.

3 Category theory and adjunctions

We will now define what an adjunction is, and set up the background for fuzzy simplicial sets.

Category theory is a branch of mathematics that unifies common concepts across different parts of mathematics. For example, one can take the product of two vector spaces, or the product of two sets, or two groups. Using ideas from category theory, we can give one unified

definition of a "product", with all these products as special cases, and prove results about it that are valid independently of the specific context.

An adjunction is a translation between different domains of discourse. For UMAP, the adjunction will let us move from a family of metric spaces to a family of fuzzy simplicial sets, where we can take a union, in a mathematically sound way. Before we can define this translation, we will set up some fundamental category theory concepts.[4]

Definition 3. A *category* $\mathscr{C}$ is a collection of *objects*, $\mathrm{Obj}(\mathscr{C})$, and between each $X, Y \in \mathrm{Obj}(\mathscr{C})$, a collection of *morphisms* $\mathrm{Hom}_{\mathscr{C}}(X, Y)$, that satisfy the following conditions. (We write $f : X \to Y$ to mean $f \in \mathrm{Hom}_{\mathscr{C}}(X, Y)$.)

First, we can *compose* morphisms. Let $f : X \to Y$ and $g : Y \to Z$ be morphisms. Then there is a specified composite morphism $gf : X \to Z$. Furthermore, composition is *associative*. Let $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$ be morphisms. Then $h(gf) = (hg)f$.

Second, for each object X, there is a specified *identity* morphism $1_X : X \to X$. Identity morphisms act as the identity under composition. That is, for any $f : X \to Y$, $1_Y f = f 1_X = f$.

One familiar example of a category is **Vect**. This is the category whose objects are real vector spaces (of all dimensions), where the morphisms between two vector spaces V and W are the linear transformations $V \to W$. Another example is **Set**, whose objects are sets, and morphisms are functions between sets.

We can define *functors* that let us transform objects and morphisms between them in one category to those in another category.

Definition 4. Let $\mathscr{C}$ and $\mathscr{D}$ be categories. A *functor* $F : \mathscr{C} \to \mathscr{D}$ is a map from the objects and morphisms of $\mathscr{C}$ to those of $\mathscr{D}$ that respects the structure of $\mathscr{C}$ in the following ways. The functor F must preserve composition, so for all composable morphisms g and f, $F(gf) = F(g)F(f)$. It also must fix identities, so that $F(1_X) = 1_{F(X)}$.

We give some examples of functors between **Vect** and **Set**. For any category $\mathscr{C}$, we have the identity functor $\mathrm{id}_{\mathscr{C}} : \mathscr{C} \to \mathscr{C}$ that fixes all objects and morphisms.

There is a "forgetful" functor $Forget : \mathbf{Vect} \to \mathbf{Set}$ that takes a vector space to underlying set of its vectors, and a linear map to the set map on the underlying sets of the vector spaces. For example, $Forget$ takes the vector space $\mathbb{R}$ to the set $\{\lambda \,|\, \lambda \in \mathbb{R}\}$. For an example of its action on morphisms, it takes the map $f : \mathbb{R} \to \mathbb{R}$ mapping (1) to (2) to the map $Forget(f) : \{\lambda \,|\, \lambda \in \mathbb{R}\} \to \{\lambda \,|\, \lambda \in \mathbb{R}\}$ that takes λ to 2λ. Note that there are many more maps from $Forget(f)$ to $Forget(f)$ than the ones that come from linear transformations. For example, the map taking $2 \mapsto 4$ and all other $\lambda \in Forget(f)$ to 0 is a valid map of sets, but is not the image under $Forget$ of a morphism in **Vect**.

Note that there is no way to reverse $Forget$, as once we have applied it to a vector space, there is no way to reconstruct the sums or scalar multiples of vectors, or to recover a basis. We might hope to find a functor $F : \mathbf{Vect} \to \mathbf{Set}$ that we can reverse. For example, we might wish to construct a functor F that takes a vector space V to a particular basis for it. However, even if there were some way to canonically choose a basis for each vector space, this would not naturally give us a functor. This is as there is no nice morphism between sets of basis elements to send, for example, the linear transformation $v \mapsto 2 \cdot v$ to.

A particularly nice functor going in the other direction is $Free : \mathbf{Set} \to \mathbf{Vect}$. This functor takes a set $\{x_i\}_{i \in I}$ to the free vector space on its elements with real coefficients. Now we can take any morphism $f : X \to Y$ between sets to a linear transformation in a canonical way, by letting $Free(f) : Free(X) \to Free(Y)$ take the basis element $Free(x_i) \in Free(X)$ to the basis element $Free(f(x_i)) \in Free(Y)$. As a linear transformation is uniquely determined by its action on a basis for the vector space, this is well-defined.

[4] See [9] for a more detailed exposition of this material aimed at scientists. We follow [8] for these definitions.

Two functors $F : \mathscr{C} \to \mathscr{D}$ and $G : \mathscr{D} \to \mathscr{C}$ form an *adjunction* if they translate between the two categories in a compatible way. The adjunction between finite fuzzy simplicial sets and finite extended pseudo-metric spaces will allow us to take our family of finite extended pseudo-metric spaces, each defined relative to a datapoint, and convert it to a family of finite fuzzy simplicial sets, The functors F and G give us an adjunction if there is a particularly nice pair of *natural transformations* between FG and the identity on $\mathscr{D}$, and between GF and the identity on $\mathscr{C}$.

A natural transformation lets us compare two functors $F, G : \mathscr{C} \to \mathscr{D}$. One motivating example for wanting to compare functors is the natural isomorphism of a vector space with its double dual. The double dual functor takes a vector space V to its double dual V^{**} by mapping $v \in V$ to the map $f \mapsto f(v)$, which is evaluation at v. The functor takes a linear transformation $T : V \to W$ to the double dual of the linear transformation, which is defined by

$$T^{**}(\phi)(f)(v) = \phi \circ (T^*(f))(v) = \phi \circ f \circ T(v).$$

The statement that the map taking V to V^{**} is a "natural isomorphism" can be formalised by saying that there is a natural transformation from the identity functor on **Vect** to the double dual functor, that is an isomorphism on all the objects. A natural transformation is defined as follows.

Definition 5. Let $\mathscr{C}$ and $\mathscr{D}$ be categories, with $F, G : \mathscr{C} \to \mathscr{D}$ functors between them. A *natural transformation* $\alpha : F \to G$ is a morphism $\alpha_X : FX \to GX$ for each $X \in \mathrm{Obj}(\mathscr{C})$ such that for any $f : X \to Y$ in $\mathscr{C}$, $Gf \circ \alpha_X = \alpha_Y \circ Ff : FX \to GY$.

Some intuition behind this is that a natural transformation α gives maps between the images of the two functors that are compatible with all the images of the morphisms between them, without knowing anything about these morphisms.

One useful thing that we can do with natural transformations is to make a new category from any pair of categories $\mathscr{C}$ and $\mathscr{D}$. We can compose natural transformations as follows. Let F, G, H be functors $\mathscr{C} \to \mathscr{D}$, with α a natural transformation from F to G, and β one from G to H. Then there is a natural transformation $\beta \circ \alpha$ from F to H with $(\beta \circ \alpha)_X = \beta_X \circ \alpha_X$. One can check that there is a category whose objects are functors $\mathscr{C} \to \mathscr{D}$, and whose morphisms are natural transformations between them.

We can now define an adjunction. Suppose we wish to translate between two categories $\mathscr{C}$ and $\mathscr{D}$, using two functors $F : \mathscr{C} \to \mathscr{D}$ and $G : \mathscr{D} \to \mathscr{C}$. We wish to define a condition on these functors for them to be a good pair of translations. We could ask that F and G give us an equivalence of categories; i.e. that $GF = 1_{\mathscr{C}}$ and $FG = 1_{\mathscr{D}}$. However, this is generally too strong a requirement, as the interesting case is when we wish to translate between inequivalent categories (e.g. between sets and vector spaces, or topological spaces and simplicial sets). An adjunction is a weakening of this equivalence idea.

Definition 6. The functors $F : \mathscr{C} \to \mathscr{D}$ and $G : \mathscr{D} \to \mathscr{C}$ form an adjunction, notated by $F \dashv G$, if there are natural transformations $\eta : 1_{\mathscr{C}} \to GF$ and $\epsilon : FG \to 1_{\mathscr{D}}$ satisfying the following commutativity conditions:

$$\epsilon_{F(c)} \circ F(\eta_c) = \mathrm{id}_{F(c)} \text{ and}$$
$$G(\epsilon_d) \circ \eta_{G(d)} = \mathrm{id}_{G(d)}$$

for all $c \in \mathscr{C}$, $d \in \mathscr{D}$.

Note that this definition is not symmetric in F and G (hence why we use the asymmetric notation $F \dashv G$). One can show [8, Section 4] that an equivalent definition is that $F \dashv G$ if there is an isomorphism between $Hom_{\mathscr{D}}(F(c), d)$ and $Hom_{\mathscr{C}}(c, G(d))$ that is natural in both c and d.

One example of an adjunction is between the free and forgetful functors between **Vect** and **Set**. These categories are not equivalent. However, we have $Free \dashv Forget$ where these functors

are those defined earlier in this section. It is straightforward to check that there is a canonical bijection between the morphism sets.

Finally, given a category $\mathscr{C}$, we will define the opposite category $\mathscr{C}^{op}$, which we will use in the next section.

Definition 7. Let $\mathscr{C}$ be a category. The *opposite category*, $\mathscr{C}^{op}$, is the category whose objects are $\mathrm{Obj}(\mathscr{C})$, and whose morphisms are defined as follows. We set $\mathrm{Hom}_{\mathscr{C}^{op}}(X, Y) = \mathrm{Hom}_{\mathscr{C}}(Y, X)$, with composition given by $g^{op} f^{op} = (fg)^{op}$.

Note that, for each morphisms $f : X \to Y$ in $\mathscr{C}$, there is an opposite morphism $f^{op} : Y \to X$ in $\mathscr{C}^{op}$. If we imagine each morphism $f : X \to Y$ in $\mathscr{C}$ as an arrow between its domain X and codomain Y, $\mathscr{C}^{op}$ is the category where we "turn all the arrows around".

These opposite morphisms are formal maps, and there is in general not an answer to the question of where f^{op} sends some $x \in X$. In the special case where the morphisms in $\mathscr{C}$ are inclusion maps, we can interpret the opposite maps as restrictions.

4 Fuzzy simplicial sets

A simplicial set is a combinatorial way of describing a space. It generalises a simplicial complex in more naturally category theoretic language. We will define a simplicial complex, then generalise to a Delta complex and a simplicial set. Finally, we will define a fuzzy set and explain how to get a fuzzy simplicial set.

The adjunction we will define in the next section will give us a mathematically coherent translation from finite extended pseudo-metric spaces to finite fuzzy simplicial sets. In the fuzzy simplicial set world, we can take a union of two different fuzzy sets on the same underlying set of elements. Thus we can take a union of the whole family of fuzzy simplicial sets coming from the extended pseudo-metric spaces (X, d_{x_i}). We will then be able to compare the resulting fuzzy simplicial set with that from a lower-dimensional representation of the dataset.

4.1 Simplicial sets

A topological space is a set of points along with a list of all open sets in the space. Manifolds and metric spaces are both topological spaces with some more structure on them. A simplicial complex describes a topological space in a combinatorial way by building it out of simplices.[5]

Definition 8. A (geometric) *n-simplex* is the convex hull spanned by a set of $n+1$ linearly independent vertices $\{x_0, \ldots, x_n\}$ in Euclidean space.

To give an equivalent definition, a geometric n-simplex is the set

$$\{\sum_{i=0}^{n} t_i x_i \,|\, t_i \geq 0, \sum_{i=0}^{n} t_i = 1\}.$$

A 0-simplex is a single point; a 1-simplex an interval; a 2-simplex a triangle. See Figure 3 for a depiction. Note that the convex hull spanned by a n-vertex subset of the $\{x_i\}_{i=0}^{n}$ is itself an $(n-1)$-dimensional simplex. We call this a *face* of the n-simplex.

In topology, we often consider spaces up to *homeomorphism*.

Definition 9. Let X and Y be topological spaces. A map $f : X \to Y$ is a *homeomorphism* if it is a continuous bijection with a continuous inverse.

[5] This section follows the exposition in [1], which is an excellent geometrically-motivated explanation of simplicial sets and simplicial homotopy theory. See that document for more detail, motivation and examples of these definitions.

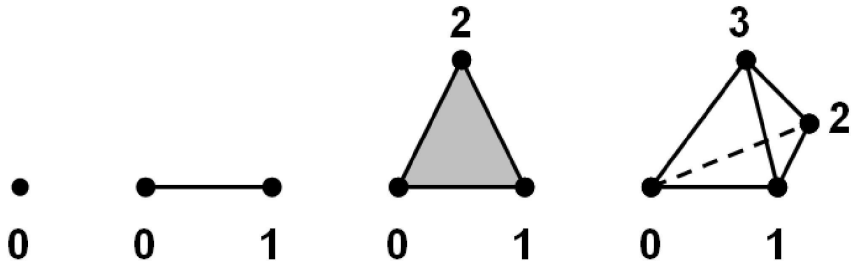


Figure 3: Examples of n-simplices for $0 \leq n \leq 3$. Image taken from [1].

For example, there is a homeomorphism between a circle and an ellipse, but not between a circle and an interval. Note that whether or not a function is a homeomorphism depends only on the open sets of the space, and is not affected by any metric or other structure on the space. Geometric n-simplices have the property that all n-simplices are homeomorphic. As a result, we can talk about an n-simplex as an abstract combinatorial object, without reference to a particular embedding in $\mathbb{R}^N$. We can describe topological spaces as simplicial complexes by decomposing them into simplices.

Definition 10. A *geometric simplicial complex* X is a collection of simplices in $\mathbb{R}^N$ such that (a) for any simplex in X, all of its faces are also in X, and (b) for any two simplices in X, their intersection is either empty or is a face of both of them.

Note that, up to homeomorphism, we can describe X by listing the vertices of each simplex. The common vertices allow us to recover which simplices share faces. We can generalise this idea as follows.

Definition 11. An (abstract) *simplicial complex* X is a series of sets X^i $(i \geq 0)$ such that the elements of X^n (the n-simplices) are $(n+1)$-element sets that satisfy the following condition: for any $\{x_i\}_{i=0}^n \in X^n$, any n-element subset of this set is in X^{n-1}.

Note that if two different simplices have common elements, this indicates that they have vertices, edges, or other sub-faces in common. For example, we can describe a square S, formed by gluing two triangles together, by

$$S^0 = \{\{a\},\{b\},\{c\},\{d\}\} \qquad S^1 = \{\{a,b\},\{a,c\},\{a,d\},\{b,c\},\{c,d\}\}$$
$$S^2 = \{\{a,b,c\},\{a,c,d\}\}.$$

We can describe the simplicial complex X depicted in Figure 4 by

$$X^0 = \{\{v_0\},\{v_1\},\{v_2\},\{v_3\},\{v_4\},\{v_5\}\}$$
$$X^1 = \{\{v_0,v_1\},\{v_0,v_2\},\{v_0,v_5\},\{v_1,v_2\},\{v_2,v_3\},\{v_2,v_4\},\{v_2,v_5\},\{v_3,v_4\}\}$$
$$X^2 = \{\{v_0,v_1,v_2\},\{v_0,v_2,v_5\},\{v_2,v_3,v_4\}\}.$$

To abstract this idea further, fix an ordering on the vertices for each simplex. We write an ordered n-simplex as $[x_0,\ldots,x_n]$. We can characterise an n-simplex in a simplicial complex by its $n+1$ face maps, where the i^{th} face map d_i sends the simplex to the face formed by removing the i^{th} vertex: $[x_0,\ldots,x_{i-1},x_{i+1},\ldots,x_n]$. Note that d_i is not a map of topological spaces: it is a formal map. If we view the simplicial complex as a category whose objects are simplices with morphisms for face inclusion, the face maps are morphisms in the opposite category.

One can check that we have the relation that, for $i \leq j$, $d_i d_j = d_{j-1} d_i$. As these face maps are enough to characterise the simplices and how they are glued together, we can generalise a simplicial complex in the following way.

Definition 12 (Delta complex)**.** A *Delta complex* X is a collection of sets X^i with, for each $n \geq 0$ and $0 \leq i \leq n$, a map $d_i : X^n \to X^{n-1}$ satisfying the relation $d_i d_j = d_{j-1} d_i$ for all $i \leq j$.

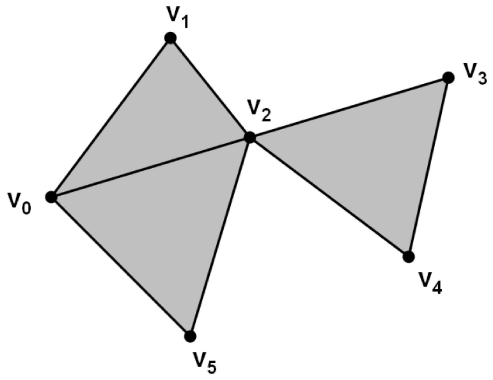


Figure 4: A simplicial complex X. Image taken from [1].

We interpret the elements of X^n as the n-simplices of the Delta complex.

If we add the restriction that the vertex set of each simplex is distinct, this is an alternate definition of an abstract simplicial complex. In a Delta complex, two distinct simplices may share the same vertex set.

Unlike simplicial complexes, we can give an equivalent definition of Delta complexes in category theoretic language.

Definition 13 (Delta complex II)**.** Let $\hat{\Delta}$ be the category whose objects are finite ordered sets $[n] := [0, 1, \ldots, n]$ and whose morphisms are strictly order-preserving maps $[m] \to [n]$. A *Delta complex* is a functor $X : \hat{\Delta}^{op} \to \mathbf{Set}$.

We can translate between this definition and the previous one as follows. Interpret the set $X([n])$ as a list of the n-simplices of the Delta complex. Note that all strictly order-preserving maps $[m] \to [n]$ factor as compositions of maps $[k] \to [k+1]$. Thus it suffices to consider the images of maps $[k+1] \to [k]$ in $\hat{\Delta}^{op}$ to describe the images of the morphisms. Suppose $f : [k] \to [k+1]$ is a morphism in $\hat{\Delta}$. This is a order-preserving injection, and can be uniquely described by the number $i \in [k+1]$ that is not in its image. We can interpret f^{op} as a restriction map that drops one number. Now its image under X takes a $(k+1)$-simplex to a k-simplex, and we interpret it as the i^{th} face map.

We are now ready to define a simplicial set.

Definition 14. Let Δ be the category whose objects are the finite ordered sets $[n]$, and whose morphisms are order-preserving maps $[m] \to [n]$. A *simplicial set* is a functor $X : \Delta^{op} \to \mathbf{Set}$.

Note that the only difference between this definition and the categorical definition of a Delta complex is that morphisms in Δ need only be weakly order-preserving.

The effect of this difference between $\hat{\Delta}$ and Δ is that in a simplicial set we allow degenerate simplices.

For example, one morphism in Δ is $f : [2] \to [1]$ mapping $0 \mapsto 0$, $1 \mapsto 1$ and $2 \mapsto 2$. Thus, for a simplicial set $X : \Delta^{op} \to \mathbf{Set}$, there is a map $X(f) : X^1 \to X^2$ that takes a 1-simplex e in X^1 to some "2-simples" in X^2. We interpret $X(f)(e)$ to be a 2-simplex (a triangle) that has been collapsed into an edge, which we call a *degenerate* simplex. A simplicial set carries the information of all possible degenerate simplices, while a Delta complex has none of them.

4.2 Fuzzy sets

A *fuzzy set* is a generalisation of a set where, rather than elements being either in the set or not, they can take a continuous range of membership values. We think of these values as probabilities.

Definition 15. A *fuzzy set* is a set of objects A and a membership function $\mu : A \to [0, 1]$.

If $\mu(a) = 1$, we view a as definitely being in the set; if $\mu(a) = 0$, it is definitely not in the set. If (A, μ) and (A, ν) are two fuzzy sets on the same underlying set of elements, we can take a union of these membership functions by interpreting them as probabilities. Using this and the adjunction from extended pseudo-metric spaces to fuzzy simplicial sets that we will define in the next section, we can combine our family of incompatible metric spaces into one fuzzy simplicial set.

The category **Fuzz** of fuzzy sets has fuzzy sets as objects, and maps of sets $f : (A, \mu) \to (B, \nu)$ such that $\nu \circ f(a) \geq \mu(a)$ as morphisms.[6]

We can now define a fuzzy simplicial set.

Definition 16. A *fuzzy simplicial set* is a functor $X : \Delta^{op} \to \mathbf{Fuzz}$. The category **sFuzz** is the category whose objects are fuzzy simplicial sets, and morphisms are natural transformations between them.

One nice property of these fuzzy simplicial sets is that the membership strength of the face of a simplex is at least the membership strength of the simplex. We say that a fuzzy simplicial set X is *finite* if it has a finite number of non-degenerate simplices, defined as follows.

Definition 17. A simplex U in $X([n])$ is *degenerate* if there is some morphism $f : [n] \to [n-1]$ in Δ such that U is in the image of $X(f^{op})$.

If U satisfies this condition, then U is a degenerate n-simplex collapsed onto $X(f^{op})^{-1}(U)$. The number of non-degenerate n-simplices of x is the number of non-degenerate elements of $X([n])$ with positive membership strength.

The category **Fin-sFuzz** is the full subcategory of **sFuzz** on the finite fuzzy simplicial sets. That is, its objects are all finite fuzzy simplicial sets, and the morphisms are all the morphisms in **sFuzz** that start and end at these objects.

5 Converting between metric spaces and fuzzy simplicial sets

We will now give the adjunction from extended pseudo-metric spaces to fuzzy simplicial sets.

Let x_i be a fixed datapoint in the dataset D. Recall that, from Section 2, using the manifold metric based at x_i, we can approximate the distance based at x_i between points $x_j, x_k \in D$ by

$$d_{x_i}(x_j, x_k) = \begin{cases} \frac{1}{r_i}(d_{\mathbb{R}^N}(x_j, x_k) - \rho_i) & \text{if } j = i \text{ or } k = i \\ \infty & \text{otherwise.} \end{cases}$$

This distance definition gives us an extended pseudo-metric space.

Let **EPMet** be the category of extended pseudo-metric spaces where the morphisms are non-expansive maps. (This condition that an edge must be taken to a shorter edge is analogous to the condition that a morphism of fuzzy sets must take an element a to an element with a higher membership strength.) Let **FinEPMet** be the sub-category of **EPMet** whose objects are finite extended pseudo-metric spaces. We restrict to this sub-category, whose spaces have only finitely many points, as our dataset D will always be finite so (D, d_{x_i}) will always gives us an object in this category.

The main theorem of [2], which is a slight modification of the main theorem of [10], gives a translation between the category of finite extended pseudo-metric spaces and the category of finite fuzzy simplicial sets.

[6] We can also define fuzzy sets categorically. Let I be the interval $[0, 1]$ whose open sets are intervals of the form $[0, a)$. Let $\mathscr{I}$ be the category whose elements are open sets in I, and whose morphisms are the inclusion maps $[0, a) \to [0, b)$ for $a \leq b$. A fuzzy set is a functor $S : \mathscr{I}^{op} \to \mathbf{Set}$ that is a sheaf and satisfies the following condition. See [11] for the definition of a sheaf in category theoretic language, where you replace the category of R-modules with **Set**. If $\rho_{b,a}$ is the inclusion map $[0, a) \to [0, b)$ for $a \leq b$, $\rho^{op}_{b,a}$ is the restriction map. Then S must satisfy the condition that $S(\rho_{b,a})$ is an injection for all $a \leq b$. This condition allows us to interpret $S([0, a))$ to be the elements of S with membership strength at least a, with the morphism $S(\rho_{b,a})$ being the inclusion from a set of elements of membership strength at least b, to that of elements of membership strength at least a.

Theorem. *There is an adjunction between* ***FinEPMet*** *and* ***Fin-sFuzz*** *given by* $FinSing$: ***FinEPMet*** $\to$ ***Fin-sFuzz*** *and* $FinReal$: ***Fin-sFuzz*** $\to$ ***FinEPMet****, with* $FinReal \dashv FinSing$.

The functors in this theorem are the following.[7] The functor $FinReal$: **Fin-sFuzz** $\to$ **FinEPMet** takes a finite fuzzy simplicial set X to the finite extended pseudo-metric space T whose elements are the vertices of X (so the set $X([0])$), and whose metric is defined as follows. Let μ be the membership strength function for the simplices of X.[8] For $x, y \in X([0]) := T$, let $\{U\}$ be the set of simplices of X (of any dimension) with both x and y as vertices. Then $d_T(x, y) = \min_U - \log(\mu(U))$.

The functor $FinSing$: **FinEPMet** $\to$ **Fin-sFuzz** is not used directly in the UMAP algorithm. We define it for completeness. Note that $FinSing$ takes an extended pseudo-metric space Y to a finite fuzzy simplicial set, which itself is a functor from $\Delta^{op} \to$ **Fuzz**. It acts as follows. The functor $FinSing(Y)$ takes $[n]$ to the fuzzy set of $(n+1)$-vertex subsets of the points of Y. If $\{x_{k_0}, \ldots, x_{k_n}\}$ is such a subset, then in $FinSing(Y)([n])$ its membership strength is

$$\mu(\{x_{k_i}\}) = \min_{i,j} e^{-d(x_{k_i}, x_{k_j})}.$$

Its action on the morphisms is in the induced action from comparing vertices.

To summarise, by Section 2 we can convert a set of datapoints D to a family of elements of **FinEPMet**, $\{(D, d_{x_i})\}_{x_i \in D}$. Using the adjoint pair of functors, we can further translate this to a family of finite fuzzy simplicial sets $FinSing(D, d_{x_i})$ for $x_i \in D$.

Now, set the *fuzzy topological representation* of D to be

$$\bigcup_{i=1}^{n} FinSing(D, d_{x_i})$$

where the union is some choice of a union of fuzzy sets. We know that each of the $FinSing(D, d_{x_i})$ has the same set of objects, which are all simplices whose vertices are in D. Now, if (A, μ) and (A, ν) are two fuzzy sets with the same underlying set of objects, one reasonable definition of the union $(A, \mu) \cup (A, \nu)$ is $(A, (\mu \cup \nu))$ where $(\mu \cup \nu)(a) = \mu(a) \perp \nu(a)$ for $\perp$ some t-conorm. The current implementation of UMAP uses $x \perp y = x + y - xy$, which is the obvious t-conorm to use if you interpret $\mu(a)$ and $\nu(a)$ as probabilities of the simplex a existing, assume these are independent between the different local metric spaces, and do not care about higher-dimensional simplices (as, for reasons of computational complexity, we will not).

6 Finding a good low-dimensional representation

We now have a method for constructing a fuzzy simplicial set from a given set of points in $\mathbb{R}^n$. Let the dataset D be in $\mathbb{R}^N$. Let E be a low-dimensional representation of our dataset D in $\mathbb{R}^m$, for $m < N$. To evaluate how good E is as a representation of D, we compare the fuzzy simplicial set X constructed from D to one constructed from E. In constructing a fuzzy simplicial set Y from E, note that we already know the metric of the underlying manifold as it is $\mathbb{R}^m$ itself. Thus, $Y = FinSing((E, d))$ where d is the Euclidean metric on $\mathbb{R}^m$.

Consider the sets of edges in X and Y as fuzzy sets. Note that they have the same underlying set of elements, which is all edges whose vertices are labelled by elements of D, and differ only in the membership strength of the simplices. We define the cross-entropy C of two fuzzy sets with the same underlying elements set, (A, μ) and (A, ν), as follows [2, Definition 10]:

$$C((A, \mu), (A, \nu)) = \sum_{a \in A} \left(\mu(a) \log\left(\frac{\mu(a)}{\nu(a)}\right) + (1 - \mu(a)) \log\left(\frac{1 - \mu(a)}{1 - \nu(a)}\right) \right).$$

[7] See [2, Section 2] for a categorical definition of these functors that makes it easier to prove they are adjoint.

[8] We do not define this precisely. See the remark after Definition 1.1 of [10] for a rigorous definition: Spivak's characteristic form is our membership strength function.

Note that, in our case, μ is fixed. We can view this formula as follows: $\mu(a)\log\left(\frac{\mu(a)}{\nu(a)}\right)$ provides the attractive force, as it is minimised if short edges in D correspond to short edges in E, since the length of the edge is small if $\nu(a)$ is large. Then $(1-\mu(a))\log\left(\frac{1-\mu(a)}{1-\nu(a)}\right)$ provides the repulsive force, as it is minimised if long edges in D correspond to long edges in E. We can then optimise the embedding using stochastic gradient descent. Note that X and Y contain many simplices of high dimension. For reasons of computational cost, the current implementation of UMAP only looks at the cross-entropy of the one-dimensional simplices in X and Y. [9]

7 Further reading

The paper presenting UMAP, [2], gives a good explanation of the algorithm and implementation separate from its mathematical foundations.

For a geometrically-motivated definition of simplicial sets, and the realisation and singular functors in the classical (non-fuzzy) context, see [1].

For an introduction to category theory aimed at non-mathematicians, see [9]. To better understand the material discussed here, you need familiarity with adjunctions (Definition 3.70).

Spivak's proof of the adjunction between the realisation and singular functors in the fuzzy set context in [10] is much more explicit than that in the UMAP paper. Both assume familiarity with adjunctions.

References

[1] G. Friedman (2008). *An elementary illustrated introduction to simplicial sets*, preprint, arXiv: 0809.4221.

[2] L. McInnes, J. Healy and J. Melville (2018). *UMAP: Uniform manifold approximation and projection for dimension reduction*, preprint, arXiv:1802.03426.

[3] L. McInnes (2018). *Topological Approaches for Unsupervised Learning*, talk at Machine Learning Prague, accessed at `https://slideslive.com/38913519/topological-approaches-for-unsupervised-learning`.

[4] L. McInnes (2018). *Performance Comparison of Dimension Reduction Implementations*, accessed at `https://umap-learn.readthedocs.io/en/latest/benchmarking.html`.

[5] L. McInnes (2019). *UMAP: Uniform manifold approximation and projection*, accessed at `https://github.com/lmcinnes/umap`.

[6] J. Melville (2018). *UMAP Examples*, accessed at `https://jlmelville.github.io/uwot/umap-examples.html`.

[7] J. Melville (2019). *UWOT: An R package implementing the UMAP dimensionality reduction method*, accessed at `https://github.com/jlmelville/uwot`.

[8] E. Riehl (2014). *Category Theory in Context*, accessed at `http://www.math.jhu.edu/~eriehl/context.pdf`.

[9] B. Fong and D. Spivak (2018). *Seven Sketches in Compositionality: An Invitation to Applied Category Theory*, accessed at `http://math.mit.edu/~dspivak/teaching/sp18/7Sketches.pdf`.

[9] For the definition of the simplices given here, the higher-dimensional simplices are entirely determined by the weights of the edge set, so we lose no information by doing this.

[10] D. Spivak. *Metric realization of fuzzy simplicial sets*, accessed at `http://math.mit.edu/~dspivak/files/metric_realization.pdf`.

[11] D. Weng. *A categorical introduction to sheaves*, accessed at `http://www.math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/WengD.pdf`.