Functorial Manifold Learning

– Research Project report –

Master of Computer Science

Minh Hoang DAO

March 10, 2023

Minh Hoang DAO
daohoang.hust@gmail.com
Student ID: 22213844

Supervisor: Prof. Vincent BARRA

Institut Supérieur d'Informatique, de Modélisation et de leurs Applications
Université Clermont Auvergne
1 Rue de la Chebarde
63178 Aubière
France

Abstract

Category theory is a general theory of mathematical structures and their relations based on simple but abstract ideas. It has been a powerful tool for mathematicians, physicians and also computer scientist by forming a general way to do researches on any fields. Because of that, this research project report aims to present basic notions of category theory, before building up more complex concepts. Later on, many ideas developed in order to construct theoretical foundation of some machine learning applications. In details, the well-known manifold learning algorithm UMAP and a functorial view on manifold learning algorithms will be mentioned and analysed. The report also provides an experiment to show the preeminence of UMAP when compared with other algorithms before concluding about the importance of category theory in machine learning and future researches.

Acknowledgements

I would like to express my sincere gratitude to all the teachers who have guided me throughout my Master's program in computer science. Their dedication and support in teaching gave me invaluable knowledge and inspiration to pursue researches.

I would like to thank my advisor, Professor Vincent Barra, for his support and guidance throughout the research process. Although the topic of the thesis was very new for a beginner studying category theory like me, his encouragement helped me to believe that I could study a new field and learn abstract mathematics.

I would also like to extend my heartfelt thanks to professor Hervé Kerivin, Mourad Baiou and David Hill for their valuable guidance and support during my coursework. Their depth of knowledge, innovative teaching methods helped me to develop a strong foundation in optimization and computer science.

Once again, I express my deepest appreciation to all the teachers who have contributed to my academic and personal growth. I will always be grateful for their guidance and support throughout this journey in Clermont-Ferrand.

Contents

1 Introduction

Category theory is a general theory of studying mathematics by constructing simple but abstract notions that was introduced by Samuel Eilenberg and Saunders Mac Lane in the middle of the 20th century in their foundational work on algebraic topology. Nowadays, category theory is used in almost all areas of mathematics, and in some areas of computer science.

The research topic of this thesis is **Functorial Manifold Learning** (FML), which studies about the same title research of Shiebler (2022). Manifold learning is also known as nonlinear dimensionality reduction, an important problem in Machine Learning. The main difference between Manifold learning and regular dimensionality reduction problem is their objects. The term "*manifold*" means locally homeomorphic to $\mathbb{R}^n$ but not globally, which means we can not use regular Euclidean distance to calculate the difference between arbitrary points in the data set as usual. But we need to create some metric on curved space that can represent data points properly, and finally make a projection to lower dimension spaces. The term "*functorial*" is from *functor* in Category theory, which can be understood as a structure preserved transformation of a system containing its objects and morphisms.

In detail, FML decomposes a manifold learning problem into composition of a hierarchical clustering algorithm and a loss function using fuzzy simplicial sets. This process is built by functors, and then we can use clustering algorithm properties to study the functoriality of manifold learning algorithms such as MDS (Abdi (2007)), IsoMap (Tenenbaum, Silva, and Langford (2000)), UMAP (McInnes et al. (2018)). All of these algorithms have their corresponding hierarchical clustering functor lie on the spectrum of created by single linkage and maximal linkage clustering algorithms. This analysis helps us to understand the functoriality of each manifold learning algorithm over the subcategories of pseudo-metric spaces and characterize them.

Along with analysing FML, this thesis also considers the categorical principles of UMAP, which is a well known algorithm for manifold learning. The algorithm relies on the theory of simplex category, with the connection between simplicial sets and topology spaces. In order to apply for metric space, a fuzzy version of simplicial sets is built also its corresponding metric realization. With the strong mathematical basis, UMAP is well grounded and has superior performance in practice.

This thesis is laid out as follows. Section 2 contains required materials about basic notions of category theory, construction of simplicial sets and geometric realization.

Section 3 presents two application of category theory in machine learning, one is the theory underlying of UMAP, one is the result of FML paper. Section 4 shows an experimental example of UMAP discusses it.

2 Preliminaries

This section provides preliminaries about category theory, with examples are referenced to several resources:

- Basic Category Theory by Leinster (2016).
- An Invitation to Applied Category Theory by Fong and Spivak (2019).
- Elementary Applied Topology by Ghrist (2014).
- nLab website: https://ncatlab.org/.

2.1 Basic notions of category theory

This subsection contains basic notions of category theory helping to build up the simplicial sets in the next one and the foundation of UMAP, FML in Section 3.

2.1.1 Categories

As in the name of category theory, categories are building blocks to start forming other objects, it is defined as follows.

Definition 2.1. *A* ***category*** $\mathscr{C}$ *consists of:*

- *a collection* $\mathrm{ob}(\mathscr{C})$ *of objects;*
- *for every two objects* $A, B \in \mathrm{ob}(\mathscr{C})$*, a collection* $\mathscr{C}(A, B)$ *of* ***morphisms*** *from* A *to* B *(also written* $\hom_{\mathscr{C}}(A, B)$ *or* $[A, B]_{\mathscr{C}}$*).*
- *for every object* $A \in \mathrm{ob}(\mathscr{C})$*, an element* $\mathbf{1}_A \in \mathscr{C}(A, A)$*, called the* ***identity*** *morphism in* A*.*
- *for every three objects* $A, B, C \in \mathrm{ob}(\mathscr{C})$ *and morphisms* $f \in \mathscr{C}(A, B)$ *and* $g \in \mathscr{C}(B, C)$*, a morphism* $g \circ f \in \mathscr{C}(A, C)$*, called the* ***composite*** *of* f *and* g *(also written* $f \mathbin{;} g$*).*

such that the following properties are satisfied:

- ***associativity***: *for any three morphisms* $f \in \mathscr{C}(A,B), g \in \mathscr{C}(B,C), h \in \mathscr{C}(C,D)$*, we have* $(h \circ g) \circ f = h \circ (g \circ f)$*.*
- ***unitality***: *for any morphism* $f \in \mathscr{C}(A,B)$*, we have* $f \circ \mathbf{1}_A = f = \mathbf{1}_B \circ f$*.*

We call **commutative diagrams** are diagrams of objects and maps, for example

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g\circ f"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

such that whenever there are two paths from an object A to C, the composed maps along all paths are equal.

Example 2.2. Categories of mathematical structures:

- **Sets**: objects are sets and morphisms are maps.
- **Grp**: objects are groups and morphisms are homomorphisms.
- **CRing**: objects are commutative rings and morphisms are homomorphisms.
- **Mon**: objects are monoids and morphisms are monoid homomorphisms.
- **Grph**: objects are graphs and morphisms are graphs morphisms.
- **Posets**: objects are partially ordered sets or posets and morphisms are order preserving functions.
- **SimpSet**: objects are simplicial sets and morphisms are simplicial maps.
- **Top**: objects are topological spaces and morphisms are continuous maps.
- **Vect**: objects are vector spaces and morphisms are linear transformations.
- **VectBund**: objects are vector bundles and morphisms are bundle morphisms.
- **Cat**: objects are small categories and morphisms are functors between them.

In the next few sections, more complex structures are built up by some above categories, such as simplex category based on **Posets**, simplicial sets as presheaves based on **Sets**, UMAP based on simplex category and **Top**.

In order to build up presheaf, we need to construct the dual category.

Construction 2.3 (Dual category). *For every category $\mathscr{C}$, we define its dual category $\mathscr{C}^{\mathrm{op}}$ having the same objects but reversed morphisms, which means $\mathscr{C}^{\mathrm{op}}(B, A) = \mathscr{C}(A, B)$ for any $A, B \in \mathrm{ob}(\mathscr{C})$.*

2.1.2 Functors

With the idea of applying developed concepts from one field in mathematics to another, functors plays an important role with the capability of transforming categories while maintaining underlying structures.

Definition 2.4. *Let $\mathscr{C}, \mathscr{D}$ be categories, a **functor** $F : \mathscr{C} \to \mathscr{D}$ is a map of objects and morphisms that assigns to each object $A \in \mathrm{ob}(\mathscr{C})$ an object $F(A) \in \mathrm{ob}(\mathscr{D})$ and to each morphism $f : A \to B$ a morphism $F(f) : F(A) \to F(B)$ such that it preserves identities and the composition of morphisms:*

- $F(\mathbf{1}_A) = \mathbf{1}_{F(A)}$ *for any* $A \in \mathscr{C}$.
- $F(g \circ f) = F(g) \circ F(f)$ *if* $g \circ f$ *is defined in* $\mathscr{C}$.

Example 2.5. Some common functors:

- identity functor: $\mathbf{1}_{\mathscr{C}}$ maps objects and morphisms of a category $\mathscr{C}$ into itself.
- forgetful functors: it removes the structure of a category $\mathscr{C}$ and return the underlying sets in **Sets**. For example, the removal of a (topological, group, order) structure from (topological spaces, groups, poset) is a forgetful functor from category (**Top**, **Grp**, **Posets**) to **Sets**.
- hom-functors: $H^A : \mathscr{C} \to \mathbf{Sets}$, let $A \in \mathrm{ob}(\mathscr{C})$ then $H^A(X) = \mathscr{C}(A, X)$, and if $f : X \to Y$ is in $\mathscr{C}$, then $H^A(f)$ is the map $\mathscr{C}(A, X) \to \mathscr{C}(A, Y)$, which sends $p \in \mathscr{C}(A, X)$ to $f \circ p \in \mathscr{C}(A, Y)$. We write $\mathscr{C}(A, f) = H^A(f)$ and $H^A = \mathscr{C}(A, -)$. The direct use of hom-functor is defining representable functors.
- order-preserving maps: when ordered sets A, B is regarded as categories, any order-preserving map between them is a functor, that is, a function $f : A \to B$ such that $a \leq a' \implies f(a) \leq f(a')$.
- singular functor and geometric realization functor create a pair of functors between categories **SimpSet** and **Top**, while the singular functor probes a

topological space with topological n-simplices and return a simplicial set, the geometric realization functor replaces n-simplices of a simplicial set with topological n-simplices in order to build up a topological space.

Definition 2.6. *Let $\mathscr{C}$ be a category, a **presheaf** on $\mathscr{C}$ is a functor $\mathscr{C}^{\mathrm{op}} \to \mathbf{Sets}$.*

Presheaves on a category $\mathscr{C}$ are something like using objects of $\mathscr{C}$ as different types of shapes and then connecting those pieces based on matching rules defined by morphisms in $\mathscr{C}$.

In the case when $\mathscr{C}$ is the poset of open sets in a topological space, presheaves are related to many notions in topology, differential geometry such as real functions, differentiable functions, vector fields, vector bundle. In general, presheaves play a very important role in modern geometry.

2.1.3 Natural transformations

In essence, functors are maps between categories and natural transformations are maps between functors. They are the main research object in category theory.

Definition 2.7. *Let $F, G : \mathscr{C} \to \mathscr{D}$ be functors, a **natural transformation** $\alpha : F \to G$ is a map which assigns any object $A \in \mathrm{ob}(\mathscr{C})$ a morphism $\alpha_A : F(A) \to G(A)$ in $\mathscr{D}$ such that for any morphism $f : A \to B$, the following diagram commutes*

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, "\alpha_A"'] & F(B) \arrow[d, "\alpha_B"] \\ G(A) \arrow[r, "G(f)"'] & G(B) \end{tikzcd}$$

The morphism $\alpha_A : F(A) \to G(A)$ in $\mathscr{D}$ is called the A-component of α. Also, we usually write as follows

$$\begin{tikzcd} \mathscr{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathscr{D} \end{tikzcd}$$

to depict the natural transformation.

Example 2.8. Some natural transformations on algebraic structures:

- Consider the total order $(\mathbb{R}, \leq)$ as a category and two translation functors $T_\mu, T_\nu : (\mathbb{R}, \leq) \to (\mathbb{R}, \leq)$ that sends $a \mapsto a + \mu$ and $a \mapsto a + \nu$ for any

$a \in \mathbb{R}$. For any morphism $f : a \to b$ which is $a \leq b$ in $\mathbb{R}$, there is a natural transformation $\alpha : T_\mu \to T_\nu$ that sends $a \mapsto a - \mu + \nu$ satisfying the commutativity $T_\nu(f)\alpha(T_\mu(a)) = T_\nu(f)\alpha(a+\mu) = T_\nu(f)(a+\nu) = b+\nu = \alpha(b+\mu) = \alpha T_\mu(f)(a+\mu) = \alpha T_\mu(f)(T_\mu(a))$. Note that this natural transformation doesn't depend on $a \in \mathbb{R}$.

- Consider the category of commutative rings **CRing**, the $n \times n$ matrices with entries in $R \in \mathrm{ob}(\mathbf{CRing})$ create a monoid $M_n(R)$, also any ring homomorphism $R \to R'$ induces a monoid homomorphism $M_n(R) \to M_n(R')$, which means $M_n : \mathbf{CRing} \to \mathbf{Mon}$ is a functor. Besides, any ring R is also a monoid $U(R)$, which means $U : \mathbf{CRing} \to \mathbf{Mon}$ is a functor.
  Notice that the determinant of any matrix $A \in M_n(R)$ is $\det_R(A)$ having its value in R. The determinant also satisfies composition and identity, so it is a monoid homomorphism $\det_R : M_n(R) \to U(R)$, and we obtain the natural transformation as follows

$$\begin{tikzcd} \mathbf{CRing} \arrow[r, bend left, "M_n"] \arrow[r, bend right, "U"'] \arrow[r, phantom, "\Downarrow \det"] & \mathbf{Mon} \end{tikzcd}$$

- In algebraic topology, many homology theories are isomorphic, and these isomorphisms are natural. Which means that there are natural transformations between various homology functors: cellular, singular, Morse, Čech,...

Because natural transformations satisfy composition and identity (a map sends a functor to itself) properties, this leads to a view of regarding natural transformations as functors between categories of initial functors. For any two categories $\mathscr{C}, \mathscr{D}$, there is a category whose objects are functors from $\mathscr{C}$ to $\mathscr{D}$ with morphisms are natural transformations. This is the **functor category** from $\mathscr{C}$ to $\mathscr{D}$ and written as $[\mathscr{C}, \mathscr{D}]$ or $\mathscr{D}^{\mathscr{C}}$.

2.1.4 Representable functors

Representability is one of the most fundamental concepts of category theory, it provides the view of objects from its category.

Definition 2.9. *A functor $X : \mathscr{C} \to$ **Sets** is called **representable** if $X \cong H^A$ for some $A \in \mathscr{C}$ and $H^A = \mathscr{C}(A, -)$ is a hom-functor. A **representation** of X is an isomorphism between H^A and X.*

Example 2.10. Representable functors of some categories:

- Consider the functor $H^1 : \mathbf{Sets} \to \mathbf{Sets}$ which takes a set X and returns all the maps from one-element set to X. These maps actually ways to extract just one element from the set X, so $H^1(X) \cong X \cong \mathbf{1_{Sets}}(X)$, which implies $\mathbf{1_{Sets}}$ is representable.

- Similar to the above functor, the forgetful functor $U : \mathbf{Top} \to \mathbf{Sets}$ which removes the topological properties and returns the underlying set is isomorphic to the functor $H^1 = \mathbf{Top}(1, -)$ extracting an element from a topological space.

- The forgetful functor $U : \mathbf{Grp} \to \mathbf{Sets}$ removes group properties of a group G and returns just the group actions as elements in a set. This functor is isomorphic to $\mathbf{Grp}(\mathbb{Z}, -)$ which takes G and gives the number of ways sending element $1 \in \mathbb{Z}$ to an element of G.

- For p is a prime number, the functor $U_p : \mathbf{Grp} \to \mathbf{Sets}$ which is defined by $U_p(G) =$ elements of G of order 1 or p probes a group G with the question "How many elements of G are cyclic with order p?", and the answer is similar to the number of maps sending the element $1 \in \mathbb{Z}_p$ to an element of G such that these maps preserve group homomorphism. This implies the isomorphism $U_p \cong \mathbf{Grp}(\mathbb{Z}_p, -)$.

- Let $\mathbf{Toph}_*$ be the category whose objects are topological spaces equipped with a basepoint and morphisms are homotopy classes of basepoint-preserving continuous maps. The $S^1 \in \mathbf{Toph}_*$ is the circle, for any $X \in \mathbf{Toph}_*$, the maps $S^1 \to X$ are the elements of the fundamental group $\pi_1(X)$. It means that the composite functor

$$\mathbf{Toph}_* \xrightarrow{\pi_1} \mathbf{Grp} \xrightarrow{U} \mathbf{Sets}$$

  is isomorphic to $\mathbf{Toph}_*(S^1, -)$ which implies that it is representables.
  The meaning behind is that the quantity of the fundamental group at the basepoint is the same as the number of twisting ways of the circle to detect one-dimensional holes in the topological space X.

In the next subsection, while constructing the simplicial sets, we will meet representable functors called n-simplices.

2.2 Category of simplicial sets

This subsection constructs the simplicial sets, which is a generalization of high-dimensional directed graph. They have their geometric realization to the corresponding topological spaces and play an important role in homotopy theory.

2.2.1 Simplex category

In short, simplicial sets are functors from dual category of simplex category to category of sets, so we firstly need to define simplex category.

Definition 2.11. *A n-**simplex** is a total ordered set*

$$[n] \coloneqq 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow \ldots \longrightarrow n-1 \longrightarrow n.$$

*The **simplex category** (denoted by $\mathbf{\Delta}$) whose objects are $[n]$ and morphisms are order-preserving functions as follows*

- ***face maps**: $\delta_i^n : [n-1] \to [n]$ is the injective map whose image leaves out $i \in [n]$ $(n > 0, 0 \le i \le n)$.*
- ***degeneracy maps**: $\sigma_i^n : [n+1] \to [n]$ is the surjective map such that $\sigma_i^n(i) = \sigma_i^n(i+1) = i$ $(n \ge 0, 0 \le i \le n)$.*

*which satisfy following **simplicial relations***

$$\delta_i^{n+1} \circ \delta_j^n = \delta_{j+1}^{n+1} \circ \delta_i^n \qquad i \le j$$

$$\sigma_j^n \circ \sigma_i^{n+1} = \sigma_i^n \circ \sigma_{j+1}^{n+1} \qquad i \le j$$

$$\sigma_j^n \circ \delta_i^{n+1} = \begin{cases} \delta_i^n \circ \sigma_{j-1}^{n-1} & i < j \\ \mathrm{Id}_n & i = j \text{ or } i = j+1 \\ \delta_{i-1}^n \circ \sigma_j^{n-1} & j+1 < i \end{cases}$$

whenever i, j are chosen so that the maps are defined.

The chain of maps in $\mathbf{\Delta}$ looks like the following diagram

$$[0] \underset{\sigma_i^0}{\overset{\delta_i^1}{\rightleftarrows}} [1] \underset{\sigma_i^1}{\overset{\delta_i^2}{\rightleftarrows}} \ldots \underset{\sigma_i^{n-2}}{\overset{\delta_i^{n-1}}{\rightleftarrows}} [n-1] \underset{\sigma_i^{n-1}}{\overset{\delta_i^n}{\rightleftarrows}} [n]$$

where each arrow is a family of maps, for example the family of δ_i^2 is as follows

$$
\begin{tikzcd}[row sep=small]
\delta_0^2 : 0 \arrow[dr] & 0 \\
1 \arrow[dr] & 1 \\
 & 2
\end{tikzcd}
\qquad
\begin{tikzcd}[row sep=small]
\delta_1^2 : 0 \arrow[r] & 0 \\
1 \arrow[dr] & 1 \\
 & 2
\end{tikzcd}
\qquad
\begin{tikzcd}[row sep=small]
\delta_2^2 : 0 \arrow[r] & 0 \\
1 \arrow[r] & 1 \\
 & 2
\end{tikzcd}
$$

and the family of σ_i^1

$$
\begin{tikzcd}[row sep=small]
\sigma_0^1 : 0 \arrow[r] & 0 \\
1 \arrow[ur] & 1 \\
2 \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}[row sep=small]
\sigma_1^1 : 0 \arrow[r] & 0 \\
1 \arrow[r] & 1 \\
2 \arrow[ur] &
\end{tikzcd}
$$

2.2.2 Simplicial sets

Definition 2.12. *A* ***simplicial set*** *X is a presheaf on $\mathbf{\Delta}$, that means it is a functor $X : \mathbf{\Delta}^{\mathrm{op}} \to \mathbf{Sets}$. In detail, the images of objects and morphisms in $\mathbf{\Delta}$ in* **Sets** *defined as follows*

- *for each $[n]$, $X([n])$ (or X_n) is the set of n-simplices.*
- *for each injective map $\delta_i^n : [n-1] \to [n]$, a function $d_i^n := X(\delta_i^n) : X_n \to X_{n-1}$ is the i-th face map on n-simplices ($n > 0, 0 \leq i \leq n$).*
- *for each surjective map $\sigma_i^n : [n+1] \to [n]$, a function $s_i^n := X(\sigma_i^n) : X_n \to X_{n+1}$ is the i-th degeneracy map on the n-simplices ($n > 0, 0 \leq i \leq n$).*

such that these maps preserve primitive compositions of simplicial relations, which is called ***simplicial identities****.*

As discussed, we can think of presheaves on a category $\mathscr{C}$ are something like using objects of $\mathscr{C}$ as different types of shapes and then connecting those pieces based on matching rules defined by morphisms in $\mathscr{C}$.

For this situation with shapes being n-simplices, a simplicial set X is constructed by n-simplices in some way that preserves the face maps when we look at boundary of simplices, also the degeneracy maps when we try to connect points to form bigger simplices contained in X.

Consider presheaves of the form $\mathbf{\Delta}(-, [n])$ (or Δ^n for short) that takes a total ordered set $[m]$ and returns the set of order-preserving maps from $[m]$ to $[n]$ in

$\mathbf{\Delta}$. Obviously, they are representables presheaves, these special maps are actually **Yoneda embeddings** and have a important property.

Lemma 2.13 (**Yoneda lemma**). *Let $\mathscr{C}$ be a category, $[\mathscr{C}^{\text{op}}, \mathbf{Sets}]$ is the functor category out of the dual category of $\mathscr{C}$ into* **Sets**. *For any $A \in \text{ob}(\mathscr{C})$, there is a Yoneda embedding functor $\mathscr{C}(-, A)$ which also is a representable presheaf. For any $X \in [\mathscr{C}^{\text{op}}, \mathbf{Sets}]$ is a presheaf, there is a isomorphism*

$$\hom_{[\mathscr{C}^{\text{op}}, \mathbf{Sets}]}(\mathscr{C}(-, A), X) \simeq X(A)$$

between the set of presheaf homomorphisms from the representable presheaf $\mathscr{C}(-, A)$ to X and the value of X at A.

The presheaf homomorphisms above actually are natural transformations. One interpretation for the meaning of Yoneda lemma is that instead of studying about property of an object X in view A, we can study how X interacts with the generalized object of A that can answer how to study A in all cases. Or by a casual example, to know how good a person X studies A, just see at all the ways X performs on studying A.

In the case where $\mathscr{C} = \mathbf{\Delta}$, the functor category $[\mathbf{\Delta}^{\text{op}}, \mathbf{Sets}]$ is called the simplex category **SimpSet** whose objects are simplicial sets and morphisms are simplicial maps. Applying the Yoneda lemma for $\mathbf{\Delta}$ obtains

$$\hom_{[\mathbf{\Delta}^{\text{op}}, \mathbf{Sets}]}(\Delta^n, X) \simeq X_n$$

that means there is a natural correspondence between n-simplices of X and morphisms $\Delta^n \to X$. The density theorem stating that a simplicial set X is a colimit of its simplices, or unofficially written as

$$\operatorname*{colim}_{x \in X_n} \Delta^n \simeq X$$

which can be understood that X is created by the smallest union of all of its simplices.

2.2.3 Geometric realization

Geometric realization is the operation that builds from a simplicial set X a topological space $|X|$. It takes each abstract n-simplex of X_n and transforms it into a standard topological n-simplex and then gluing all simplices along their boundaries to a combined topological space with respect to face and degeneracy maps of X telling how those simplices are connected to each other.

More formally, geometric realization is a functor $|\cdot| : \mathbf{\Delta} \to \mathbf{Top}$ that sends $[n]$ to the standard topological n-simplex $|\Delta^n| \subset \mathbb{R}^{n+1}$ defined as

$$|\Delta^n| \coloneqq \left\{ (t_0, ..., t_n) \in \mathbb{R}^{n+1} \middle| \sum_{i=0}^{n} t_i = 1,\ t_i \geq 0 \right\}.$$

and whose topology is the subspace topology inducing from the canonical topology in $\mathbb{R}^{n+1}$. And now we can construct the realization of entire X (denoted by $|X|$) by using its colimit representation

$$|X| = \operatorname*{colim}_{x \in X_n} |\Delta^n|$$

In contrast, there is a functor called singular functor that sends any topological space Y to its corresponding simplicial set $S(Y)$ as follows

$$S(Y) : [n] \mapsto \hom_{\mathbf{Top}}(|\Delta^n|, Y)$$

for each object $[n] \in \mathrm{ob}(\mathbf{\Delta})$. In classical homotopy theory, the geometric realization functor $|\cdot|$ and the singular functor S form an adjunction, which can be understood that they are dual to each other as the following diagram

$$\begin{tikzcd} \mathbf{SimpSet} \arrow[r, bend left, "|\cdot|"] & \mathbf{Top} \arrow[l, bend left, "S"] \end{tikzcd}$$

and satisfy the isomorphism

$$\hom_{\mathbf{SimpSet}}(X, S(Y)) \cong \hom_{\mathbf{Top}}(|X|, Y).$$

3 Machine learning models using category theory

3.1 UMAP

This subsection analyses the theory underlying of manifold learning (MfL) algorithm UMAP (McInnes et al. (2018)), which is well-constructed using category theory language.

In MfL problem, we usually fit the initial dataset into some kind of metric spaces which are induced by some methods. The main reason is that MfL problem want to reduce the dimension of the dataset, the evaluation functions usually base on the difference of pairwise distances of points between the initial space and embedded space. So the metric space should be constructed to quantify this objective.

As discussed in 2.2.3, two functors $|\cdot|$ and S form an adjunction between two categories $\mathbf{\Delta}$ and **Top**. The category **Top** just only cares about topological spaces, not those equipped with a metric. In order to use constructed materials, Spivak (2012) transmuted simplicial sets into fuzzy ones and defined metric realization functor mapping them to *uber-metric space*. Using those materials, UMAP defined a related metric space, but for now we construct the fuzzy simplicial sets at first.

3.1.1 Fuzzy simplicial sets

Let I denote the unit interval $(0, 1] \subseteq \mathbb{R}$ and whose topology given by intervals $[0, a)$ where $a \in (0, 1]$. The category of open sets with inclusions as morphisms has a Grothendieck topology, given in the usual way.

Definition 3.1. *A* ***fuzzy set*** *is a presheaf* $X \in [I^{\mathrm{op}}, \mathbf{Sets}]$ *such that for any morphism* $a \leq b$ *in* $(0, 1]$*, the map* $X(a \leq b)$ *is a injection.*

Since the Grothendieck topology on I is trivial, all the presheaves are sheaves which satisfy the sheaf condition. The sheaf condition is the uniqueness of gluing all sections which are fuzzy set on the trivial I. The category of sheaves on I is denoted by $\mathbf{Shv}(I)$ and also called a **topos**.

Definition 3.2. *The category* **Fuzz** *of fuzzy sets is the full subcategory of* $\mathbf{Shv}(I)$ *spanned by the fuzzy sets.*

We only care about the category spanned by the fuzzy sets which have injective maps on **Sets**. The reason is that the injection condition preserves the monotonicity of fuzziness on poset I.

Definition 3.3. *The category of* ***fuzzy simplicial sets*** *denoted by* **sFuzz** *is the category with objects given by functors from* $\mathbf{\Delta}^{\mathrm{op}}$ *to* **Fuzz**, *and morphisms given by natural transformation.*

In the category of small categories **Cat**, we have a universal property of exponential object called **currying** as follows

$$[\mathscr{A} \times \mathscr{B}, \mathscr{C}] \cong \big[\mathscr{A}, [\mathscr{B}, \mathscr{C}]\big].$$

For $\mathscr{A} = \mathbf{\Delta}^{\mathrm{op}}$, $\mathscr{B} = I^{\mathrm{op}}$ and $\mathscr{C} = \mathbf{Sets}$, we obtain

$$[\mathbf{\Delta}^{\mathrm{op}} \times I^{\mathrm{op}}, \mathbf{Sets}] \cong \big[\mathbf{\Delta}^{\mathrm{op}}, [I^{\mathrm{op}}, \mathbf{Sets}]\big] = [\mathbf{\Delta}^{\mathrm{op}}, \mathbf{Fuzz}] = \mathbf{sFuzz}.$$

Alternatively, a fuzzy simplicial set can be viewed as a presheaf over $\mathbf{\Delta} \times I$ which has the product topology with trivial topology of $\mathbf{\Delta}$. We denote the representable sheaf of the object $([n], [0, a))$ as $\Delta^n_{<a}$. With these objects, we can make a connection to metric spaces, as we did for simplicial sets and topological spaces.

3.1.2 Metric realization

Definition 3.4. *An extended-pseudo-metric space* (X, d) *is a set* X *and a map* $d : X \times X \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ *such that*

- $d(x, y) \geq 0$, *and* $x = y$ *implies* $d(x, y) = 0$,
- $d(x, y) = d(y, x)$,
- $d(x, z) \geq d(x, y) + d(y, z)$ *or* $d(x, z) = \infty$.

The category of extended-pseudo-metric spaces is denoted by **EPMet** *whose objects are extended-pseudo-metric space and whose morphisms are non-expansive maps. The subcategory of finite extended-pseudo-metric spaces is* **FinEPMet**.

The word "extended" means allowing the infinity distances, and "pseudo" means that $d(x, y)$ is not necessarily equal to 0 if $x \neq y$.

Now we construct a pair of adjoint functors *Real* and *Sing* between **sFuzz** and **EPMet** as we did for $\mathbf{\Delta}$ and **Top**. The standard topological fuzzy simplices $\Delta^n_{<a}$ is defined by *Real* functor as follows

$$Real(\Delta^n_{<a}) := \left\{ (t_0, ..., t_n) \in \mathbb{R}^{n+1} \,\middle|\, \sum_{i=0}^{n} t_i = -\log(a), t_i \geq 0 \right\}$$

which is similar to geometric realization functor $|\cdot|$. A morphism $\Delta^n_{<a} \to \Delta^m_{<b}$ exists only if $a \leq b$ and is determined by a morphism $\sigma : [n] \to [m]$ in $\mathbf{\Delta}$. The action of *Real* on that morphism is given by the map

$$(x_0, ..., x_n) \mapsto \frac{\log(b)}{\log(a)} \left(\sum_{i_0 \in \sigma^{-1}(0)} x_{i_0}, ..., \sum_{i_0 \in \sigma^{-1}(m)} x_{i_0} \right).$$

We also can define the metric realization for entire X using colimit as follows

$$Real(X) \coloneqq \operatorname*{colim}_{\Delta^n_{<a} \to X} Real(\Delta^n_{<a}).$$

In contrast, the right adjoint functor *Sing* defined for an extended-pseudo-metric space Y sends it to the corresponding fuzzy simplicial set $Sing(Y)$ as follows

$$Sing(Y) : \big([n], [0, a)\big) \mapsto \hom_{\mathbf{EPMet}}(Real(\Delta^n_{<a}), Y)$$

for each object $\big([n], [0, a)\big) \in \mathrm{ob}(\mathbf{\Delta} \times I)$. The two functors *Real* and *Sing* form an adjunction as in the following diagram

$$\begin{tikzcd} \mathbf{sFuzz} \arrow[r, bend left, "Real"] & \mathbf{EPMet} \arrow[l, bend left, "Sing"] \end{tikzcd}$$

In the case of finite dimension, we need to define two new adjoint functors $FinReal$ and $FinSing$ for the category of bounded fuzzy simplicial sets **Fin-sFuzz** and the category of finite extended-pseudo-metric spaces **FinEPMet**.

Definition 3.5. *The functor $FinReal$: **Fin-sFuzz** $\to$ **FinEPMet** defined by*

$$FinReal(\Delta^n_{<a}) \coloneqq \big(\{x_1, ..., x_n\}, d_a\big),$$

where

$$d_a(x_i, x_j) = \begin{cases} -\log(a) & i \neq j \\ 0 & \textit{otherwise} \end{cases}$$

and then defining

$$FinReal(X) \coloneqq \operatorname*{colim}_{\Delta^n_{<a} \to X} FinReal(\Delta^n_{<a}).$$

The action of $FinReal$ on a morphism $\Delta^n_{<a} \to \Delta^m_{<b}$ is given by

$$\big(\{x_1, ..., x_n\}, d_a\big) \mapsto \big(\{x_{\sigma(1)}, ..., x_{\sigma(n)}\}, d_b\big)$$

As a result, the functor $FinSing$ can be constructed as how we defined $Sing$.

Definition 3.6. *The functor* $FinSing : \mathbf{FinEPMet} \to \mathbf{Fin\text{-}sFuzz}$ *defined by*

$$FinSing(Y) : \big([n], [0, a)\big) \mapsto \hom_{\mathbf{FinEPMet}}(Real(\Delta^n_{<a}), Y).$$

And we have the theorem about these two functors.

Theorem 1. *The functors* $FinReal : \mathbf{Fin\text{-}sFuzz} \to \mathbf{FinEPMet}$ *and* $FinSing : \mathbf{FinEPMet} \to \mathbf{Fin\text{-}sFuzz}$ *form an adjunction with* $FinReal$ *the left adjoint and* $FinSing$ *the right adjoint.*

The above theorem gives the following diagram

$$\begin{tikzcd}
\mathbf{Fin\text{-}sFuzz} \arrow[r, bend left, "FinReal"] & \mathbf{FinEPMet} \arrow[l, bend left, "FinSing"]
\end{tikzcd}$$

Put together all materials above, we can define the fuzzy topological representation of a dataset.

Definition 3.7. *Let* $X = \{X_1, ..., X_N\}$ *be a dataset in* $\mathbb{R}^n$. *Let* $\big\{(X, d_i)\big\}_{i=1...N}$ *be a family of extended-pseudo-metric spaces of* X *such that*

$$d_i(X_j, X_k) = \begin{cases} d_{\mathcal{M}}(X_j, X_k) - \rho & if \;\; i = j \;\; or \;\; i = k, \\ \infty & otherwise, \end{cases}$$

where ρ *is the distance to the nearest neighbour of* X_i *and* $d_{\mathcal{M}}$ *is the geodesic distance on the manifold* $\mathcal{M}$. *The fuzzy topological representation of* X *is*

$$\bigcup_{i=1}^{N} FinSing\big((X, d_i)\big).$$

The fuzzy topological representation of X is actually merging all metric spaces with respect to local areas of each X_i in order to form the global representation of the manifold $\mathcal{M}$.

3.1.3 Optimizing a low dimensional representation

To perform dimension reduction of the manifold $\mathcal{M}$, we just need to find a lower-dimensional fuzzy topological representation of $\mathcal{M}$ that matches the initial representation with respect to some predefined objectives.

In the paper of UMAP, authors considered only the 1-skeleton of fuzzy simplicial sets, which is the family of $\Delta^1_{<a}$, this represents each fuzzy simplicial sets as a fuzzy graph containing fuzzy edges. The objective is chosen as the fuzzy set cross entropy, which evaluates on fuzzy sets. So we need to convert a fuzzy topological representation to a reference set A and its corresponding membership function $\mu : A \to [0,1]$, this can be done by setting $A = \cup_{a\in(0,1]} S([0,a))$ and $\mu(x) = \sup \big\{a \in (0,1] \big| x \in S([0,a))\big\}$ for each sheaf representation S.

Definition 3.8. *The cross entropy C of two fuzzy sets (A,μ) and (A,ν) is defined as*

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

With this loss function, we can use traditional gradient-based methods to find the optimal lower-dimensional fuzzy topological representation of the manifold $\mathcal{M}$. The steps of the UMAP algorithm can be described as follows

1. From the dataset $X = \{X_1, ..., X_N\} \in \mathbb{R}^n$, using some methods to calculate local metrics of X_i which induces geodesic distances on an assumed manifold $\mathcal{M}$ in order to obtain the family $\big\{(X,d_i)\big\}_{i=1...N}$ denoted by Y.

2. Using functor $FinSing$ on each (X,d_i) with discretized interval I to form a subsheaf $FinSing\big((X,d_i)\big)$ with 1-skeletons $\Delta^1_{<a}$ only. This subsheaf gives information about fuzzy edges of X_i to the others in the fuzzy graph.

3. Combining all subsheaves to form the fuzzy topological representation of X as $S \coloneqq \bigcup_{i=1}^N FinSing\big((X,d_i)\big)$.

4. Calculating the reference set A with its membership function μ.

5. Initializing a random (or using some methods) lower-dimensional representation of dataset X called $X' = \{X'_1, ..., X'_N\} \in \mathbb{R}^m$ with $m < n$. Repeating steps 1 to 4 for X' to obtain the membership function ν.

6. Using gradient descent methods to calculate the change of ν as $\Delta\nu$ and use it to modify its fuzzy topological representation S'. Repeating this step to iteratively update S'.

3.2 Functorial Manifold Learning

This part analyses the paper Functorial Manifold Learning (Shiebler (2022)) classifying MfL algorithms based on their functoriality, which is similar to classifying clustering algorithms in the paper of Carlsson and Mémoli (2013).

Firstly, we will define basic notions of simplicial complexes and needed categories.

3.2.1 Hierarchical clustering functor

In machine learning, there is a class of hierarchical clustering algorithms, their main idea is clustering data in many resolutions. The FML paper and several other researches used hierarchical clustering functors to model multiscale clusters of data. Therefore, this part will construct the hierarchical clustering functor step by step.

Definition 3.9. *A* ***finite simplicial complex*** *is a family of finite sets that is closed under taking subsets. A* ***flag complex*** *is a finite simplicial complex generated by cliques in a graph.*

In the clustering problem, each cluster can be regarded as a cover of a subset of the initial set X, that leads to the following definition.

Definition 3.10. *Given a set X, a* ***non-nested flag cover*** *of $\mathcal{C}_X$ of X is a cover of X such that:*

- *if $A, B \in \mathcal{C}_X$ and $A \subseteq B$ then $A = B$,*
- *the simplicial complex over entire X with faces are all finite subsets of the sets in $\mathcal{C}_X$ is a flag complex.*

In the above definition, the first condition is the *non-nested* property and the second one is that we can locally cover X by cliques.

Definition 3.11. *The category* **Cov** *is defined as follows*

- *objects:* $(X, \mathcal{C}_X)$ *where* $\mathcal{C}_X$ *is a non-nested flag cover of* X.
- *morphisms:* $f : (X, \mathcal{C}_X) \to (Y, \mathcal{C}_Y)$ *if* $f(S) \subseteq S'$ *for any* $S \in \mathcal{C}_X$ *and some* $S' \in \mathcal{C}_Y$.

The category **Cov** is the category of non-nested flag covers of, X with morphisms between covers being "coarser topology" maps.

As we did in UMAP, here we also consider the poset $I = (0, 1]$ and define the fuzzy covers.

Definition 3.12. *The category* **FCov** *is defined as follows*

- *objects: functors* $F_X : I^{\mathrm{op}} \to \mathbf{Cov}$ *such that* $F_X(a \leq a')$ *is just the identity on the underlying set.*
- *morphisms: natural transformations between functors.*

In UMAP, we already define the category of extended-pseudo-metric spaces **EPMet**, similarly we can define the category of pseudo-metric spaces as **PMet**.

Consider a subcategory **D** of **PMet**, a **flat D-clustering functor** is a functor $C : \mathbf{D} \to \mathbf{Cov}$ regarded as clustering a family of pseudo-metric spaces in **D** and the clusters will separate gradually if the distances in spaces in **D** increase.

In order to extend clustering functor to a hierarchical one, we define a functor $H : \mathbf{D} \to \mathbf{FCov}$, the universal property for this functor implies

$$[\mathbf{PMet}, \mathbf{FCov}] = \big[\mathbf{PMet}, [I^{\mathrm{op}}, \mathbf{Cov}]\big] \cong [\mathbf{PMet} \times I^{\mathrm{op}}, \mathbf{Cov}].$$

As a result, for each value $a \in I^{\mathrm{op}}$, the functor $H(-)(a) : \mathbf{PMet} \to \mathbf{Cov}$ is a flat **D**-clustering functor, and the clusters are also more separated when a increases. Therefore, when we evaluate H over I^{op}, we obtain clusters in different resolutions, that's how we construct the **hierarchical clustering functor** H.

In summary, categories and functors are depicted in the following diagram

$$\begin{tikzcd} & \mathbf{PMet} \arrow[dl, "C"'] \arrow[d, "H"] & \\ \mathbf{Cov} & \mathbf{FCov} & I^{\mathrm{op}} \arrow[ll, Rightarrow] \end{tikzcd}$$

where the category **FCov** lying on two arrows (functors) means that **FCov** is a functor category and H is a functor from **PMet** to **FCov**.

3.2.2 Manifold learning functor

A manifold learning algorithm is actually finding an embedding of n data points in a m-dimensional space ($m \in \mathbb{N}$), which is represented by a matrix $A \in \mathbb{R}^{n\times m}$. After that, we want to evaluate embeddings A by a loss function l having $l(A) \in \mathbb{R}$, so that it is an optimization problem minimizing l over a vector space.

Moreover, we usually pay attention to the pairwise loss functions, that means $l(A) : \sum_{i,j\in 1\ldots n} l_{ij}(\|A_i - A_j\|)$. We call these kinds of problems **pairwise embedding optimization problems** and denote one as $(n, m, \{l_{ij}\})$.

Formally, a **manifold learning problem** is defined as a function that maps the pseudo-metric space (X, d_X) to a pairwise embedding optimization problem of the form $(|X|, m, \{l_{ij}\})$. A reasonable example of loss function is $l_{ij}(\|A_i - A_j\|) = (d_X(x_i, x_j) - \|A_i - A_j\|)^2$ which penalizes the pairwise difference of distances between the input space and the embedded space.

Besides, when the input space is mapped to an isometric space, the algorithm should remain the result, and this is called **isometry-invariant**. The following proposition states that isometry-invariant manifold learning problems factor through hierarchical clustering algorithms.

Proposition 3.13. *There exists a manifold learning problem $L \circ H$ that has the same solution space as any isometry-invariant manifold learning problem M on any input (X, d_X) with*

- *H is a hierarchical overlapping clustering algorithm.*
- *L is a function mapping the output of H to an embedding optimization problem.*

For the proof of this proposition, the author shows that there exists at least one $L \circ H$ being $(M \circ \mathit{Real}) \circ \mathcal{ML}$ where $\mathcal{ML}$ is the maximal linkage hierarchical overlapping clustering algorithm. The functor *Real* is a metric realization functor that maps a simplicial complex to a metric space, it is quite similar to the functor *Real* that maps a simplicial set to a metric space defined in UMAP.
In order to build the function L above, we need to define several related categories and functors.

Definition 3.14. *The preorder* $\mathbf{L}$ *has objects tuples* $\big(n, \{l_{ij}\}\big)$ *where* $l_{ij} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ *is a real-valued function that is equal to 0 for any* $i, j > n$ *and has* $\big(n, \{l_{ij}\}\big) \leq \big(n', \{l'_{ij}\}\big)$ *iff.* $l_{ij}(x) \leq l'_{ij}(x)$ *for all* $x \in \mathbb{R}_{\geq 0}$.

In particular, a loss function l_{ij} can be decomposed as the sum of a monotonically increasing function with a monotonically decreasing function, for example

$$l_{ij}(\delta) = \big(d_X(x_i, x_j) - \delta\big)^2 = \big(\delta^2 + d_X(x_i, x_j)^2\big) + \big(-2\delta d_X(x_i, x_j)\big)$$

That means, we can build the category of loss functions as the following pullback

$$\begin{tikzcd} \mathbf{Loss} \arrow[r, dashed] \arrow[d, dashed] & \mathbf{L} \arrow[d, "U"] \\ \mathbf{L}^{\mathrm{op}} \arrow[r, "U"] & \mathbb{N} \end{tikzcd}$$

where U is the forgetful functor. Therefore, **Loss** is a subcategory of $\mathbf{L}^{\mathrm{op}} \times \mathbf{L}$ and has objects as tuples $\big(n, \{c_{ij}, e_{ij}\}\big)$ where $\big(n, \{c_{ij}, e_{ij}\}\big) \leq \big(n', \{c'_{ij}, e'_{ij}\}\big)$ iff. for any $x \in \mathbb{R}_{\geq 0}$ we have $c'_{ij}(x) \leq c_{ij}(x)$ and $e_{ij}(x) \leq e'_{ij}(x)$. For $m \in \mathbb{N}$, each object of **Loss** corresponds to $\big(n, m, \{l_{ij}\}\big)$ where $l_{ij}(\delta) = c_{ij}(\delta) + e_{ij}(\delta)$.

As we constructed fuzzy covers as functors from I^{op} to **Cov**, we also want to evaluate the loss function over each $a \in I^{\mathrm{op}}$ presenting the partial loss of a-resolution clusters. For that reason, we define the category **FLoss** as functors $F : I^{\mathrm{op}} \to \mathbf{Loss}$ with $F(a) = \big(n, \{c_{F(a)_{ij}}, e_{F(a)_{ij}}\}\big)$. Also, we define a functor $Flatten : \mathbf{FLoss} \to \mathbf{Loss}$ that maps $F(a) \mapsto \big(n, \{c_{ij}, e_{ij}\}\big)$ where:

$$c_{ij}(x) = \int_{a \in I} c_{F(a)_{ij}}(x) da \qquad \text{and} \qquad e_{ij}(x) = \int_{a \in I} e_{F(a)_{ij}}(x) da$$

Therefore, each functor $F \in \mathbf{FLoss}$ corresponds to the pairwise embedding optimization problem $\big(n, m, \{l_{F_{ij}}\}\big)$, hence the sum over all i, j is F-loss as follows

$$\mathbf{l}_F(A) = \sum_{i,j \in 1\ldots n} l_{F_{ij}}\big(\|A_i - A_j\|\big) = \sum_{i,j \in 1\ldots n} \int_{a \in I} c_{F(a)_{ij}}\big(\|A_i - A_j\|\big) + e_{F(a)_{ij}}\big(\|A_i - A_j\|\big) da$$

Definition 3.15. *Given the subcategories* $\mathbf{D} \subseteq \mathbf{PMet}, \mathbf{D}' \subseteq \mathbf{FCov}$*, the composition* $L \circ H : \mathbf{D} \to \mathbf{FLoss}$ *forms a* $\mathbf{D}$***-manifold learning functor*** *if* $H : \mathbf{D} \to \mathbf{D}'$ *is a hierarchical* $\mathbf{D}$*-clustering functor and* $L : \mathbf{D}' \to \mathbf{Loss}$ *is a functor mapping a fuzzy non-nested flag cover over* X *to some* $F_X \in \mathbf{FLoss}$.

Each manifold learning functor M corresponds to a manifold learning problem that maps (X, d_X) to $\big(|X|, m, \mathbf{I}_{M(X,d_X)}\big)$. When M is decomposed as $M = L \circ H$, we have the following summary diagram

$$\begin{tikzcd}
 & \mathbf{PMet} \arrow[dl, "C"'] \arrow[d, "H"] \arrow[rr, "M"] & & \mathbf{FLoss} \arrow[r, bend left, "Flatten"] & \mathbf{Loss} \arrow[r, dashed] \arrow[d, dashed] & \mathbf{L} \arrow[d, "U"] \\
\mathbf{Cov} & \mathbf{FCov} \arrow[urr, "L"] & I^{\mathrm{op}} \arrow[ll, Rightarrow] \arrow[urr, Rightarrow] & & \mathbf{L}^{\mathrm{op}} \arrow[r, "U"] & \mathbb{N}
\end{tikzcd}$$

We denote $\mathcal{SL}, \mathcal{ML}$ as the single and maximal linkage hierarchical overlapping clustering algorithms mapping pseudo-metric space (X, d_X) to the fuzzy non-nested flag cover (X, C_{X_a}) where C_{X_a} is the connected components of

- the $-\log(a)$-Vietoris-Rips complex of (X, d_X) for $\mathcal{SL}$.
- the sets of the relation R_a such that $x_1 R x_2$ if $d_X(x_1, x_2) \leq -\log(a)$ for $\mathcal{ML}$.

We have the following proposition stating the spectrum over the category of manifold learning functors that is bounded by $\mathcal{ML}$ and $\mathcal{SL}$.

Proposition 3.16. *Given* $\mathbf{D}$ *is a subcategory of* $\mathbf{PMet}$ *such that* $\mathbf{PMet}_{bij} \subseteq D$, $L \circ H$ *is a* $\mathbf{D}$*-manifold learning functor and for all* $a \in I$*, the functor* $H(-)(a) : \mathbf{D} \to \mathbf{Cov}$ *has a clustering parameter* $\delta_{H,a}$*. Then for* $a \in I$ *and* $(X, d_X) \in \mathbf{D}$ *we have maps:*

$$(L \circ \mathcal{ML})(X, d_X)(e^{-\delta_{H,a}}) \leq (L \circ H)(X, d_X)(a) \leq (L \circ \mathcal{SL})(X, d_X)(e^{-\delta_{H,a}})$$

that are natural in a *and* (X, d_X)*.*

The proof of this proposition is based on a result stating that in the category of non-trivial hierarchical clustering functors, there exist natural transformations with inclusion maps as components from $\mathcal{ML}(-)(e^{-\delta_{H,a}})$ to H and from H to $\mathcal{SL}(-)(e^{-\delta_{H,a}})$. That means the $\mathcal{ML}$ is the limit and $\mathcal{SL}$ is the colimit, this duality is a consequence of $\mathcal{ML}, \mathcal{SL}$ definition where one is defined by the maximum distance to merge clusters and the other is formed by linking small distances to merge clusters respectively.

This proposition actually states that every manifold learning functor maps (X, d_X) to a loss being no more interconnected than the loss that does not distinguish points within the same connected component of the Vietoris-Rips complex ($\mathcal{SL}$) and no less interconnected than the loss that regards pairs of points1 independently ($\mathcal{ML}$).

3.2.3 Characterization of some algorithms

Based on the functoriality over a subcategory $\mathbf{D} \subseteq \mathbf{PMet}$, we can characterize manifold learning algorithms. We denote $\mathbf{PMet}_{isom}$ as subcategory of $\mathbf{PMet}$ with morphisms being isometry maps, similarly for $\mathbf{PMet}_{sur}$ with surjective maps. We now list a few examples of some popular manifold learning algorithms with the decomposition $L \circ H$ of them:

- Metric Multidimensional Scaling (Abdi (2007)) - $\mathbf{PMet}_{sur}$-manifold learning functor: FML paper defined a functor $MDS : \mathbf{FCov}_{sur} \to \mathbf{FLoss}$ with formula related to original paper and proved that the Metric Multidimensional Scaling embedding optimization problem is the image of functor $MDS \circ \mathcal{ML}$.
- IsoMap (Tenenbaum, Silva, and Langford (2000)) - $\mathbf{PMet}_{sur}$-manifold learning functor: FML paper defined the $\mathbf{PMet}$-clustering functor $IsoCluster_\sigma$ and concluded the algorithm as $MDS \circ IsoCluster_\sigma$.
- UMAP (McInnes et al. (2018)) - $\mathbf{PMet}_{isom}$-manifold learning functor: the hierarchical clustering functor H is decomposed into 4 functors and called $FuzzySimplex$. FML paper also defines a loss functor FCE and concluded that UMAP has a manifold learning functor $FCE \circ FuzzySimplex$.

Alongside to characterize manifold learning algorithms to give better understand about them using category theory, the FML paper also proposed that we can create new algorithms by combining different algorithms with their hierarchical clustering functors H and loss functors L. To demonstrate, the paper proposed $MDS \circ \mathcal{ML}$ as an example for the DNA recombination task and showed the significant improvement.

4 Experiments

In order to demonstrate the efficiency of UMAP, this section presents a comparison of UMAP, t-SNE, PCA, MDS, IsoMap on the **COIL-20** dataset embedded in 2-dimensional space. The dataset contains a set of 1440 greyscale images consisting of 20 objects as in Figure 1 under 72 different rotations spanning 360 degrees. Each image is a 128x128 image which we treat as a single 16384 dimensional vector for the purposes of computing distance between images. The 72 rotation images of an object in COIL-20 is depicted in Figure 2.

Figure 1: 20 objects in the COIL-20 dataset.

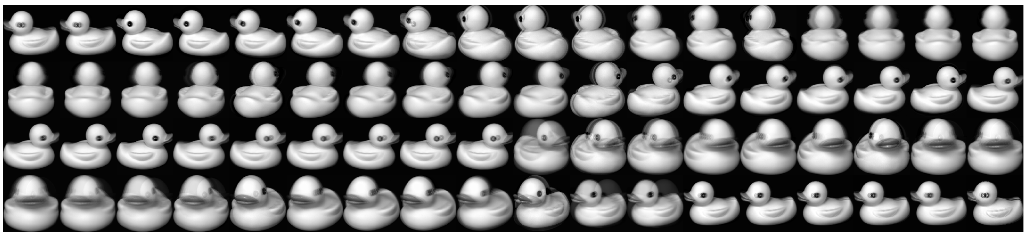

Figure 2: 72 images of an object in COIL-20.

Consequently, each set of 72 images of an object should create a loop in the high dimensional space. Therefore, images have strong connections to their neighbourhood ones within each set. As a consequence, we expect manifold learning algorithms to preserve this property in 2D embedded space while recognisably separating different objects.

The result of UMAP is shown in Figure 3. As we can see, UMAP impressively preserves loop structures of the dataset while separating most of the classes thanks to its well construction based on fuzzy simplicial sets, which relies on strong theoretical foundation of algebraic topology.

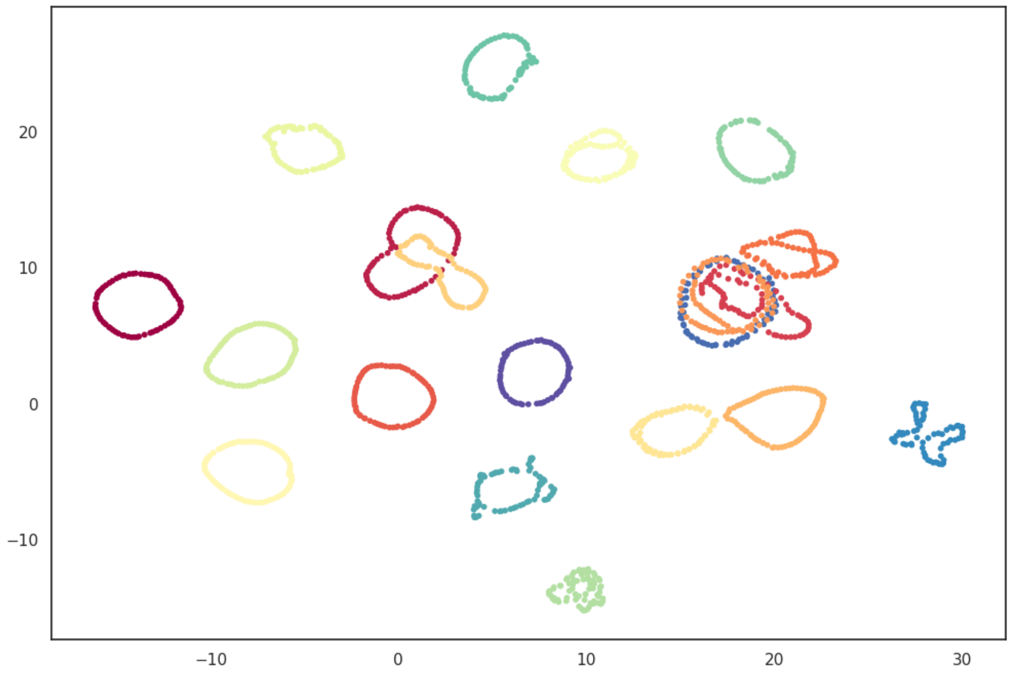

Figure 3: UMAP result on COIL-20.

In comparison with result of 4 other algorithms in Figure 4, t-SNE also has a decent performance in separating classes but still remains some classes not having loop structure or being partial tore off. PCA shows a chaotic mixture of classes, owing to its linear transformation when projecting a dataset to lower dimensional space, so that only partial global structure is preserved and local properties are lost dramatically. MDS is designed to well preserve global structure, but it also breaks many loops and distorts some. IsoMap also reveals many loops in the dataset, but these loops highly overlaps each other and are not so clean.

5 Conclusion

The category theory is an important tool with its strength in building up modern mathematics. Although it starts with simple terms and turns into abstraction very quickly, but it is very useful to generalize many ideas of different fields.

In the few past years, many research papers in machine learning used the language of category theory to construct and analyse their works. With the advantage of strong theoretical basis, category theory helps researchers having a logical and general view on their research fields. Besides, category theory provides a clear way for researchers to prove and reason about their result.

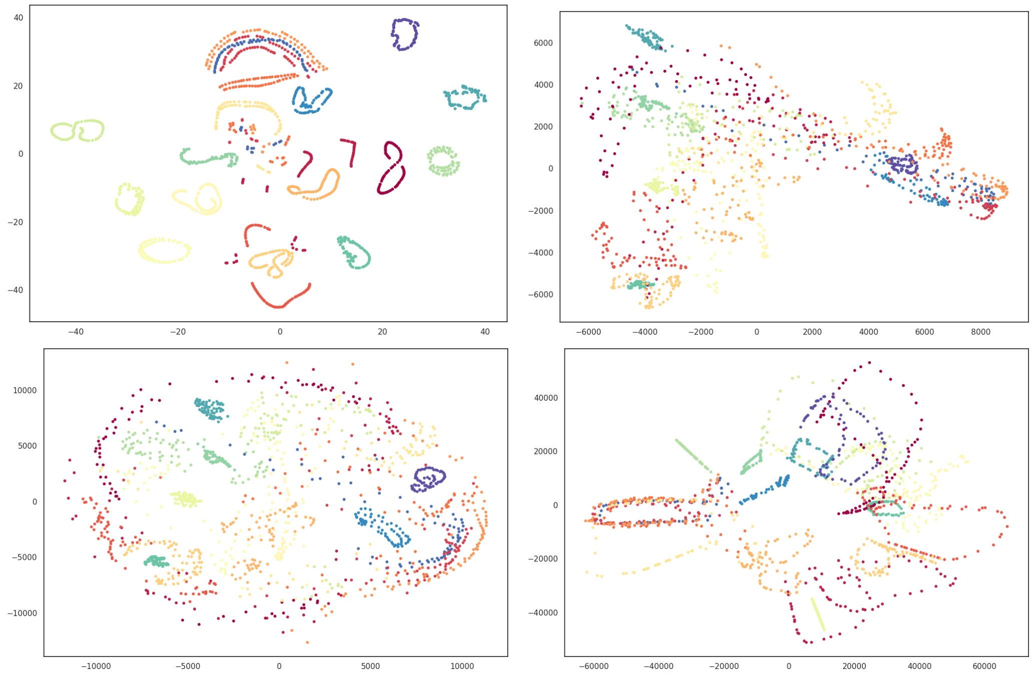

Figure 4: (Left to right, top to bottom: t-SNE, PCA, MDS, IsoMap.)

In the scope of a research project, this report achieved the predefined objectives including:

- Giving basic notions about category theory and presenting examples in order to have a good understanding, especially construction of presheaf, simplicial sets, realization functors and meaning of the Yoneda lemma.

- Analysing some applications of category theory in machine learning, including very famous manifold learning tools UMAP and a functorial view of manifold learning problems in the Functorial Manifold Learning paper.

- Experimenting UMAP algorithm to learn about its preeminence when compared to other dimensionality reduction algorithms.

In the future, machine learning in general will become a fundamental application research field, so that its theoretical basis should be very well-constructed. This will open a wide door for mathematicians and related field researchers to collaborate, therefore category theory will become a powerful tool to reshape and unite the ways we learn about the theory of machine learning.

References

Abdi, H. 2007. "Metric multidimensional scaling (MDS): analyzing distance matrices." *Encyclopedia of Measurement and Statistics.*

Carlsson, G., and F. Mémoli. 2013. "Classifying clustering schemes." *Foundations of Computational Mathematics.*

Fong, B., and D. I. Spivak. 2019. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality.* Cambridge University Press.

Ghrist, R. 2014. "Elementary Applied Topology." *CreateSpace Independent Publishing Platform.*

Leinster, T. 2016. "Basic Category Theory." *Cambridge University Press.*

McInnes, L., J. Healy, N. Saul, and L. Großberger. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *Journal of Open Source Software* 3 (29): 861.

Shiebler, D. 2022. "Functorial Manifold Learning." *Proceedings of the Fourth International Conference on Applied Category Theory,* 1–13.

Spivak, D. I. 2012. "Metric realization of fuzzy simplicial sets." *Self published notes.*

Tenenbaum, J. B., V. D. Silva, and J. C. Langford. 2000. "A global geometric framework for nonlinear dimensionality reduction." *Science* 290 (5500): 2319–2323.