Category Theory for Machine Learning

Thomas Walker

December 2023

Abstract

Machine learning is a rapidly advancing field, with new techniques emerging frequently which showcase advances in performance and capabilities. Many have referred to the collection of machine learning architectures as a zoo, there is a wealth of varieties each with its strengths and weaknesses. Consequently, there has been a growing desire to construct a unifying machine learning theory that encapsulates all these different architectures. Ultimately, category theory aims to introduce rigour into machine learning research and provide a framework for future progress. Recently, researchers have been attracted to abstract algebra to formalise many notions within machine learning. In these pages, we discuss how category theory is being applied to machine learning. Category theory concerns itself with describing the relationship between a collection of objects. It does not try to understand the individual object but rather deals with the collection on the meta-level. It provides theories on how these collections of objects interact, then it is up to the individual to contextualise these ideas within a particular setting they are interested in. This later part has been the concern of some machine learning researchers who have started to describe concepts in machine learning using the powerful framework provided by category theory.

Contents

1 Theory

1.1 Categories

Definition 1.1.1. *A category is an object that contains the following data.*

- *A collection of objects* $(X, Y, Z, \dots)$.
- *A collection of morphisms* $(f, g, h, \dots)$. *Where each morphism has a domain,* X, *and codomain,* Y, *being objects. That is,* $f : X \to Y$.

These collections of data must come with the following.

- *Each object has an identity morphism,* $1_X : X \to X$.
- *There is a notion of composition of morphisms,* $f : X \to Y, g : Y \to Z \longrightarrow gf : X \to Z$.

For a category, we assume the following set of axioms.

1. For $f : X \to Y$ a morphism we have $1_Y f = f = f 1_X$.
2. Composition is associative. That is, for f, g, h morphisms we have $h(gf) = (hg)f$.

Remark 1.1.2.

- *The collection of morphisms between objects* X *and* Y *in the category* $\mathcal{C}$ *is often denoted in different ways, such as,* $\mathcal{C}(X, Y)$ *or* $\mathrm{Hom}(X, Y)$.
- *One ought to be careful with the semantics they use to describe a category. For instance, as the objects of categories may themselves be sets, referring to the collections of objects as a set introduces complications such as Russel's paradox. Hence, the data pertaining to a category is often referred to as a collection or universe.*

Definition 1.1.3. *An isomorphism in a category is a morphism* $f : X \to Y$ *for which there is a* $g : Y \to X$ *such that* $gf = 1_X$ *and* $fg = 1_Y$. *In this case,* X *and* Y *are said to be isomorphic, written* $X \cong Y$.

Definition 1.1.4. *An endomorphism is a morphism whose codomain equals its domain.*

Definition 1.1.5. *An automorphism is an endomorphism that is also an isomorphism.*

Definition 1.1.6. *A subcategory* $\mathcal{D}$ *of a category* $\mathcal{C}$ *is the restriction of* $\mathcal{C}$ *to a sub-collection of objects and morphisms, with all the necessary axioms still holding.*

Definition 1.1.7. *A groupoid is a category in which every morphism is an isomorphism*

Example 1.1.8.

1. *The category* Set *is a category whose objects are sets and whose morphisms are functions between those sets, with compositionality being the usual composition of functions.*
2. *The category* Vect_k *is the category of vector spaces over a field* k, *with the morphisms being linear transformations.*

1.2 Duality

Note that if the direction of the morphisms in a category are reversed, that is the domain and codomain are interchanged, then all of the necessary axioms can be made to hold. Therefore, in category theory, notions have a dual that manifests upon reversing the direction of the morphisms.

Definition 1.2.1. *Let $\mathcal{C}$ be a category. Its opposite category $\mathcal{C}^{op}$ has the same objects as $\mathcal{C}$. However, each morphism f^{op} in $\mathcal{C}^{op}$ has a corresponding morphism f in $\mathcal{C}$ whose domain is the codomain of f^{op}, and whose codomain is the domain of f^{op}. For f^{op}, g^{op} morphisms in $\mathcal{C}^{op}$, we let*

$$g^{op} \cdot f^{op} = (f \cdot g)^{op}.$$

That is, we define composition in $\mathcal{C}^{op}$ using the composition of $\mathcal{C}$.

Remark 1.2.2. *Note that the identity of X in $\mathcal{C}^{op}$ is 1_X^{op}.*

1.3 Monomorphisms and Epimorphisms

Definition 1.3.1. *Let $f : X \to Y$ be a morphism in a category $\mathcal{C}$.*

1. *We call f a monomorphism if for any parallel morphisms $h, k : W \rightrightarrows X$, $fh = fk$ implies that $h = k$.*
2. *We call f an epimorphism if for any parallel morphisms $h, k : Y \rightrightarrows Z$, $hf = kf$ implies that $h = k$.*

Remark 1.3.2.

1. *Note that A monomorphism in $\mathcal{C}$ becomes an epimorphism in $\mathcal{C}^{op}$ and vice-versa.*
2. *A monomorphism is denoted with $\rightarrowtail$ and $\twoheadrightarrow$ denotes an epimorphism.*

Proposition 1.3.3. *Let $f : X \to Y$ be a morphism in a category $\mathcal{C}$.*

1. *Then f is a monomorphism in $\mathcal{C}$ if and only if for all morphisms $g : W \to X$ in $\mathcal{C}$, the post-composition with f, that is $f \circ g$, defines an injection.*
2. *Then f is an epimorphism in $\mathcal{C}$ if and only if to all morphisms $g : Y \to Z$ in $\mathcal{C}$, the pre-composition with f, that is $g \circ f$, defines an injection.*

1.4 Functors

Definition 1.4.1. *Let $F : \mathcal{C} \to \mathcal{D}$ be a map between categories $\mathcal{C}$ and $\mathcal{D}$ with the following structure.*

1. *An object X in $\mathcal{C}$ is mapped to an object FX in $\mathcal{D}$.*
2. *A morphism $f : X \to Y$ is mapped to a morphism $Ff : FX \to FY$ in $\mathcal{D}$. The map F is a covariant functor if the following Funtoriality axioms are satisfied.*
   (a) *For any composable pair $f, g \in \mathcal{C}$, $Fg \cdot Ff = F(g \cdot f)$.*
   (b) *For each object X in $\mathcal{C}$, $F1_X = 1_{FX}$.*

Definition 1.4.2. *A contravariant functor F from $\mathcal{C}$ to $\mathcal{D}$ is a covariant functor $F : \mathcal{C}^{op} \to \mathcal{D}$.*

Henceforth, we will use the term functor in the sense of covariant functors.

Remark 1.4.3. *The Funtoriality axioms for a contravariant functor are the following.*

1. *For any composable pair $f, g \in \mathcal{C}$, $Ff \cdot Fg = F(g \cdot f)$.*
2. *For each object X in $\mathcal{C}$, $F1_X = 1_{FX}$.*

Lemma 1.4.4. *Functors preserve isomorphisms.*

Example 1.4.5. *Let G be the group defined on a set X with the binary operator $\cdot$. The category $\mathrm{B}G$ consists of the single object X, and morphisms given by the action of the elements on X. The identity morphism for X is given by the identity of the group, e. One can check that a group homomorphism $\varphi : G \to H$, defines a functor $F_\varphi : \mathrm{B}G \to \mathrm{B}H$.*

Definition 1.4.6. *For categories $\mathcal{C}$ and $\mathcal{D}$, let $\mathcal{C} \times \mathcal{D}$ denote the category constructed as follows.*

- *The objects of $\mathcal{C} \times \mathcal{D}$ are ordered pairs (X, Y) where X is an object of $\mathcal{C}$ and Y is an object of $\mathcal{D}$.*
- *The morphisms of $\mathcal{C} \times \mathcal{D}$ are ordered pairs $(f, g) : (X, Y) \to (X', Y')$, where $f : X \to X'$ is a morphism of $\mathcal{C}$ and $g : Y \to Y'$ is a morphism of $\mathcal{D}$.*

Composition in $\mathcal{C} \times \mathcal{D}$ and identities are defined component-wise.

Definition 1.4.7. *A functor $F : \mathcal{C} \to \mathcal{D}$ is an isomorphism if there exists a functor $G : \mathcal{D} \to \mathcal{C}$ such that $G \circ F$ and $F \circ G$ are equal to the identity functors on $\mathcal{C}$ and $\mathcal{D}$ respectively.*

1.5 Natural Transformations

Definition 1.5.1. *Given categories $\mathcal{C}$ and $\mathcal{D}$ and functors $F, G : \mathcal{C} \to \mathcal{D}$. A natural transformation α from F to G consists of morphisms $\alpha_X : FX \to GX$ for each object X in $\mathcal{C}$ such that for any object X' in $\mathcal{C}$ and $f \in \mathcal{C}(X, X')$we have*

$$Gf \circ \alpha_X = \alpha_{X'} \circ Ff.$$

Using Definition 1.5.1 we can construct a category of functors, with the functors being the objects and the natural transformations being the morphisms.

Definition 1.5.2. *If for every object X in $\mathcal{C}$ the morphism α_X is an isomorphism, then we call $\alpha : F \to G$ a natural equivalence. In such a case we write $F \sim G$.*

Definition 1.5.3. *We call $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ equivalences of categories if $FG \sim \mathrm{Id}_{\mathcal{C}}$ and $GF \sim \mathrm{Id}_{\mathcal{C}}$.*

1.6 Products

For objects X, Y, Z in a category $\mathcal{C}$, we say that Z is the product of X and Y if there are maps $\pi_1 : Z \to X$ $\pi_2 : Z \to Y$, called the projections, such that given any other pair of maps $f_1 : A \to X$ and $f_2 : A \to Y$there is a unique map

$$f : A \to Z$$

such that

$$\pi_1 \circ f = f_1$$

and

$$\pi_2 \circ f = f_2.$$

That is we can first map $X \to Z$, take the projections in Z and still retrieve the projections of X, Y in A. Due to the existence of the morphisms $\pi_1 : Z \to X$ and $\pi_2 : Z \to Y$, the object C is said to have a universal property. Of course, this notion can be extrapolated to an arbitrary collection of objects.

Example 1.6.1. *In the category* Set *the product of the objects X_1 and X_2 is the usual Cartesian product of sets,*

$$X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_1,\ x_2 \in X_2\}.$$

In this case, $\pi_i : X_1 \times X_2 \to X_i$ is given by $(x_1, x_2) \mapsto x_i$ for $i = 1, 2$.

1.7 Monoidal Categories

Definition 1.7.1. *A monoid is a set, $\mathcal{V}$, along with an associative binary operation, $\otimes$, that has an identity element, I, in the set. A monoid is commutative if the binary relation is also commutative. We denote such a monoid as $(\mathcal{V}, \otimes, I)$.*

A monoidal category is a category with a product, which we will denote $\otimes$.

Definition 1.7.2. *A monoidal category $\mathcal{C}$ is a category equipped with the following structure.*

1. *A function $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$.*
2. *A unit object, I.*
3. *A natural isomorphism $a : (\cdot \otimes \cdot) \otimes \cdot \to \cdot \otimes (\cdot \otimes \cdot)$.*
4. *A natural isomorphism $\lambda : (I \otimes \cdot) \to \cdot$.*
5. *A natural isomorphism $\rho : \cdot \otimes I \to \cdot$.*

Remark 1.7.3.

1. *The isomorphisms in Definition 1.7.2 formalise the notion of associativity, and the property of the unit object.*
2. *If each of the above isomorphisms in Definition 1.7.2 are the identity morphisms then the monoidal category is called strict.*

1.8 Enriched Categories

In the categories we have seen thus far, each object has a collection of morphisms mapping it from itself to another object. In an enriched category, the set of morphisms acts as objects and themselves form a category. Recall that for objects X, Y in a category $\mathcal{C}$ we use $\mathcal{C}(X, Y)$ to denote the set of all morphisms $X \to Y$. The collection $\mathcal{C}(X, Y)$ is also referred to as the hom-set. An enriched category $\mathcal{K}$ has the $\mathcal{C}(X, Y)$ as objects. If $\mathcal{C}$ is monoidal the notion of composition is given by,

$$\circ : \mathcal{C}(Y, Z) \otimes \mathcal{C}(X, Y) \to \mathcal{C}(X, Z).$$

Definition 1.8.1. *A preorder is a set, $\mathcal{V}$, together with a reflexive, transitive relation, $\leq$. On the preorder $(\mathcal{V}, \leq)$ we can construct the category* Preorder *with the following structure.*

1. *Objects are elements of the set $\mathcal{V}$.*
2. *A morphism from elements X and Y exists if and only if $X \leq Y$.*

Definition 1.8.2. *A commutative monoidal preorder $(\mathcal{V}, \leq, \otimes, I)$ is a preorder $(\mathcal{V}, \leq)$ and a commutative monoid $(\mathcal{V}, \otimes, I)$ satisfying $X \otimes Y \leq X' \otimes Y'$ whenever $X \leq X'$ and $Y \leq Y'$.*

Definition 1.8.3. *Let $(\mathcal{V}, \leq, \otimes, 1)$ be a commutative monoidal preorder and let $\mathcal{C}$ denote the* Preorder *category constructed from $(\mathcal{V}, \leq, \otimes, 1)$. Then a $\mathcal{V}$-enriched category, $\mathcal{K}$, is the enriched category defined using the structure of* Preorder*. For objects X and Y of $\mathcal{C}$, there is an object $\mathcal{C}(X, Y)$ of $\mathcal{K}$ called a $\mathcal{V}$-hom object. Such hom-objects satisfy*

$$1 \leq \mathcal{C}(X, X)$$

and

$$\mathcal{C}(Y, X) \otimes \mathcal{C}(X, Z) \leq \mathcal{C}(Y, Z)$$

for all objects X, Y and Z.

Example 1.8.4. *The interval $[0, 1]$ equipped with multiplication and the usual $\leq$ relation is a commutative monoidal preorder. A $[0, 1]$-enriched category consists of a set of objects $\mathcal{C}$ and a $[0, 1]$-valued function $(x, y) \mapsto \mathcal{C}(x, y)$ defined for every $x, y \in \mathcal{C}$.*

1.9 Enriched Copresheaves

A presheaf is a functor $F : \mathcal{C}^{\mathsf{op}} \to$ Set. A copresheaf is the dual of a presheaf, that is a functor $G : \mathcal{C} \to$ Set. Throughout we use Set to denote the category of sets.

Definition 1.9.1. *Let $\mathcal{C}$ and $\mathcal{D}$ be categories enriched over a commutative monoidal preorder $(\mathcal{V}, \otimes, \leq, I)$. An enriched functor $\mathcal{C} \to \mathcal{D}$ is a map $f : \mathcal{C} \to \mathcal{D}$ satisfying*

$$\mathcal{C}(X, Y) \leq \mathcal{D}(fX, fY)$$

for all objects X and Y is $\mathcal{C}$.

Definition 1.9.2. *Let $\mathcal{C}$ be a category enriched over a closed commutative monoidal preorder $\mathcal{V}$. An enriched copresheaf is a map $f : \mathcal{C} \to \mathcal{V}$ satisfying $\mathcal{C}(X, Y) \leq \mathcal{V}(fX, fY)$ for all objects $X, Y \in \mathcal{C}$.*

Lemma 1.9.3. *If $\mathcal{C}$ is a category enriched over $[0, 1]$ then the category $\hat{\mathcal{C}} : [0, 1]^{\mathcal{C}}$ of copresheaves is also enriched over $[0, 1]$. The $[0, 1]$-object between any pair of copresheaves $f, g : \mathcal{C} \to [0, 1]$ is given by*

$$\hat{\mathcal{C}}(f, g) = \inf_{c \in \mathcal{C}} \left\{ 1, \frac{gc}{fc} \right\}.$$

Lemma 1.9.3 says that enriched copresheaves form an enriched category.

Lemma 1.9.4. *Let $\mathcal{C}$ be a $[0, 1]$-category. For every object X, the function*

$$h^X := \mathcal{C}(X, -)$$

is a $[0, 1]$-functor.

Lemma 1.9.4 says that an object defines a representable copresheaf.

Definition 1.9.5. *For any object X in a $[0, 1]$-category, $\mathcal{C}$, the functor $h^X := \mathcal{C}(X, -)$ is the copresheaf represented by X. We say a copresheaf $f : \mathcal{C} \to [0, 1]$ is representable if $f = h^X$ for some X in $\mathcal{C}$.*

Theorem 1.9.6 (The Enriched Yoneda Lemma)**.** *For any object X in a $[0,1]$-category, $\mathcal{C}$, and any $[0,1]$-copresheaf $f : \mathcal{C} \to [0,1]$, we have $\hat{\mathcal{C}}\left(h^X, f\right) = f(X)$.*

Corollary 1.9.7. *For objects X and Y in a $[0,1]$-category, $\mathcal{C}$, we have $\mathcal{C}(y,x) = \hat{\mathcal{C}}(h^x, h^y)$.*

Therefore, $X \mapsto h^X$ embeds the opposite of a $[0,1]$-category within its category of copresheaves. That is, for any $[0,1]$-category $\mathcal{C}$, the assignment $X \mapsto h^X$ defines an enriched functor $\mathcal{C}^{\mathsf{op}} \to \hat{\mathcal{C}}$, embedding $\mathcal{C}^{\mathsf{op}}$ as an enriched subcategory of $\hat{\mathcal{C}}$.

2 Application

2.1 Large Language Models as Enriched Categories

We now embed the ideas of enriched categories and enriched copresheaves within the context of language models. A large language model aims to construct sequences of words in grammatically correct and meaningful ways. To formalise this process we are going to define a syntax category, which models the probability distribution on text continuations. The probability distribution will relate to how likely a particular word is to follow or be contained within another set of words. Therefore, the syntax category only captures the distributional structure of language. To encode meaning and introduce context, we consider the category of enriched copresheaves of the syntax category, namely the semantic category.

2.1.1 The Syntax Category

Definition 2.1.1. *The syntax category $\mathcal{L}$ is a category enriched over $[0,1]$ whose objects are expressions in language and where $[0,1]$-hom objects are defined by*

$$\mathcal{L}(x,y) := \pi(y|x)$$

with $\pi(y|x)$ denoting the probability that expression y extends expression x. If x is not a subtext of y then $\pi(y|x) = 0$.

Remark 2.1.2.

1. *Definition 2.1.1 is well-defined as*
$$\pi(x|x) = 1 \geq 1$$
and
$$\pi(z|y)\pi(y|x) = \pi(z|x)$$
for all expressions x, y, z.
2. *Note how a syntax category is inherently syntactical and neglects any semantics.*

2.1.2 The Semantics Category

We can extract information about the structure of a set by investigating how it maps into a different set that has a regular structure. For example, consider a set X and $\mathbb{R}$. If X is a finite set, say $|X| = n$, then the collection of maps $\{f : X \to \mathbb{R}\}$ can be identified with $\mathbb{R}^n$. Note how we can investigate properties of X using the language of linear algebra. For categories $\mathcal{C}$ we can perform a similar shift in perspective by studying the collection of functors $\mathcal{C} \to \mathsf{Set}$. Specifically for our case, we will use the unit interval to extract information about an $[0,1]$-category. If the category of consideration is $\mathcal{C}$ we will use $\hat{\mathcal{C}} = [0,1]^{\mathcal{C}}$ to denote precisely this set of functors. Note that we have shown that $\hat{\mathcal{C}}$ is also a category.

Definition 2.1.3. *Let $\mathcal{L}$ be the syntax category. The semantic category $\hat{\mathcal{L}} := [0,1]^{\mathcal{L}}$ is the $[0,1]$-category of $[0,1]$-enriched presheaves on the $[0,1]$-category $\mathcal{L}$. For each object x in $\mathcal{L}$, $h^x := \mathcal{L}(x,-)$ is given by*

$$c \mapsto h^X(c) := \begin{cases} \pi(c|x) & \text{if } x \leq c \\ 0 & \text{otherwise} \end{cases}$$

where $x \leq c$ if and only if x is contained as a subtext of c.

Remark 2.1.4. *The idea in Definition 2.1.3 is that h^x encodes the meaning of the expression x. Note that*

h^x has a support that only contains texts including x. The varying intensities of h^x concerning the texts it contains encodes the meaning of x. Intuitively, if you know that a piece of text x is more greatly associated with words of a particular context, demonstrated by h^x having a high intensity for texts within this context, then you can extract some meaning regarding the piece of text x.

Suppose y extends the text x and $\mathcal{L}(x, y) = \pi(y|x) = a \neq 0$. Then

$$x \xrightarrow{a} y$$

and

$$h^y \xrightarrow{a} h^x.$$

Our intuition is that h^x is the meaning of x which represents the varying potential context in which x might appear, and h^y represents the varying potential context in which y appears. Thus the above relation shows that the context x appears in extends the contexts y appears in. Consequently, continuing an expression means that we restrict the potential context in which the expression can be used.