On the Uncertainty in Classification

Editors: List of editors' names

Abstract

We develop a theoretical framework for geometric deep learning that incorporates ambiguous data in learning tasks. This framework uncovers deep connections between noncommutative geometry and learning tasks. Namely, it turns out that learning tasks naturally arise from groupoids, and vice versa. We also find that learning tasks are closely linked to the geometry of its groupoid $*$-algebras. This point of view allows us to answer the question of what actually constitutes a classification problem and link unsupervised learning tasks to random walks on the second groupoid cohomology of its groupoid $*$-algebra.

Keywords: Geometric Deep Learning, Noncommutative Geometry, Unsupervised Learning, Learning Tasks, Classification Problem, Ambiguous Data

1. Introduction

In this paper we develop a theoretical framework for geometric deep learning that incorporates ambiguous data in learning tasks. Thinking of smoothly morphing an image to an another image let us come to the conclusion that the real world applications are full of ambiguous data. As far as we could survey an approach that incorporates ambiguous data does not exist yet. For this we start studying the approaches from different areas in mathematics that tackle classification problems. Naturally in these contexts the study of group actions arise as model for the classification problem. We start with studying approaches in the algebraic setting (Section 3). As an ex-curs into the application-driven deep learning approach (Section 4) suggests to us, we continue further to the harmonic analysis imprinted approaches (Section 5). We find out that in both the algebraic and the harmonic analysis setting, the representation theory of the regular representation of the group algebra plays a central role. By this investigations we discover a new class of functions, so-called *expected-to-be-invariant* functions (Definition 8) which generalizes the notion of *invariant* functions and seems better suited for applications. So far all classification problems were induced by a group action. In particular, we did not deal with ambiguous data yet. In Section 6 we find out that we can capture the distribution of the data from the sample space via so-called probability kernels. After that we enter into the groupoid setting. Groupoids allow us to handle ambiguous data. According to our previous observations in the algebraic and harmonic analysis setting we focus on the groupoid algebra and its regular representation. For locally compact Hausdorff groupoid with left Haar system we introduce its twisted groupoid $*$-algebra following Renault (2006) in Section 7. After these steps we are finally able to answer the question of what actually constitutes a classification problem/learning task:

Definition 14 A classification problem consists of a locally compact groupoid $\mathcal{G}$ with a left Haar system $\{\lambda^u\}_{u\in\mathcal{G}_0}$, and a normalized continuous $\mathbb{T}$-valued 2-cocycle σ on $\mathcal{G}$ such that $\sigma(\alpha,\beta)=\int_{\gamma\in\mathcal{G},s(\gamma)=r(\beta)}\sigma(\alpha,\gamma)\sigma(\gamma,\beta)d\lambda^{d(\alpha)}(\gamma)$ for all $\alpha,\beta\in\mathcal{G}$. The space $\mathcal{G}_0$ is called *sample space*, and σ is called *labeling*.

© 2022 .

This new viewpoint allows us to link unsupervised learning to random walks on the second groupoid cohomology with values in $\mathbb{T}$, the multiplicative group of complex numbers of modulus 1.

2. Definitions and Notation

At first, we need to introduce a generic definition of a classification problem:

Definition 1 *A classification problem* C *is a tuple* $(\Omega, (\Omega_l)_{l\in L})$ *consisting of a set* Ω*, so-called sample space, and family* $(\Omega_l)_{l\in L}$ *of subsets of* Ω *such that*

a) $\Omega = \bigcup_{l\in L} \Omega_l$,

b) $\bigcup_{l\in L'} \Omega_l \subsetneq \Omega$ *for all* $L' \subsetneq L$.

Thereby, a subset Ω_l $(l \in L)$ *is called cluster and the index set* L *is called the set of classes of* C*. We call the classification problem*

In this note, we will distinguish between two types of classification problems:

Definition 2 *We call a classification problem* C *unambiguous if its clusters are pairwise disjoint. Otherwise, we call it ambiguous.*

Definition 3 *A classification problem* C *is induced by a group action if there is a group* G *acting on its sample-space* Ω *such that the cluster are given by the orbits of the* G *action.*

Remark 4 *For a classification problem* $\mathsf{C} = (\Omega, (\Omega_l)_{l\in L})$ *induced by a group action* $G : \Omega$*, there is an one-to-one relation between the index set* L *and* Ω/G *(in the set-theoretical sense).*

Lemma 5 *A classification problem induced by a group action is an unambiguous classification problem, and vice versa.*

3. Algebraic classifications

For this section, we fix an algebraically closed field $\mathbb{K}$, and for $n \in \mathbb{N}$ let $\mathbb{A}^n$ denote the affine variety $\mathbb{K}^n$. All algebraic varieties in this section will be over $\mathbb{K}$.

Let $\mathsf{C} = (\Omega, (\Omega_l)_{l\in L})$ be a classification problem with an affine variety Ω as sample space. Further assume that C is induced by an algebraic group G acting regular on Ω. Then, the cluster are locally closed subsets of the sample space and we have an associated unambiguous classification problem $\bar{\mathsf{C}} = (\Omega, (\bar{\Omega}_l)_{l\in \bar{L}})$ where the cluster are induced by the closure of the orbits. Regarding the classification problem in this setting, it is natural to consider first the *associated regular* G*-representation* $\mathbb{K}[\Omega] = \{f : \Omega \to \mathbb{K} \,|\, f \text{ is regular}\}$ given by $(g.f)(\omega) := f(g^{-1}\omega)$, $g \in G$, $f \in \mathbb{K}[\Omega]$, $\omega \in \Omega$ (Kraft and Wiedemann (1985)), then the *invariant ring* $\mathbb{K}[\Omega]^G := \{f \in \mathbb{K}[\Omega] \ : \ g.f = f \text{ for all } g \in G\} \subseteq \mathbb{K}[\Omega]$ and focusing on the challenge to find $(f_i)_{i\in I} \subseteq \mathbb{K}[\Omega]$ such that $\mathbb{K}[f_i : \ i \in I] = \mathbb{K}[\Omega]^G$ holds. To understand the intuition why such a set of generators of the invariant ring is desirable assume that $I = \{1, \ldots, m\}$ for some $m \in \mathbb{N}$. In such a case, one can consider the regular function

$p : \Omega \to \mathbb{A}^m, \omega \mapsto (f_1(\omega), \ldots, f_m(\omega))$. This function detects the cluster of $\bar{\mathsf{C}}$, that is, for all $v, \omega \in \Omega$ we have: $\overline{G\omega} = \overline{Gv}$ iff $p(\omega) = p(v)$. More than detecting the orbit closures could not have been expected in order not to violate the continuity requirements!

If we summarize, desirable properties would be, first, that I is finite and, second, that these functions could be calculated effectively using a 'simple' formula or algorithm.

The first desirable property on finite generation does not hold for arbitrary groups. There is a famous counterexample by Nagata (1959). Adding further assumptions on G, namely to be linear reductive, releases more methods from representation theory of groups to apply. In that way, the invariant ring occurs naturally as part of the isotypic decomposition of the regular G-representation $\mathbb{K}[\Omega]$ and one can conclude that the invariant ring is finitely generated:

Theorem 6 (Hilbert (1890), Noether (1915), Hilbert (1933)) *The invariant ring respective an linear reductive group is finitely generated.*

Regarding the second desirable property, we would like to give a sketch of a constructive proof for finite groups G and $\mathrm{char}(\mathbb{K}) = 0$ by following Noether (1915).

Proof We extend the *regular* G-representation to a *representation of the group algebra of* G, denoted by $\mathbb{K}[G]$, in $\mathrm{End}_{\mathbb{K}}(\mathbb{K}[\Omega])$ induced by the $\mathbb{K}$-algebra morphism

$$\mathrm{H} : \mathbb{K}[G] \to \mathrm{End}_{\mathbb{K}}(\mathbb{K}[\Omega]), \alpha \mapsto \left(f \mapsto \frac{1}{|G|} \sum_{g \in G} \alpha(g) \cdot (g.f) \right).$$

We have $\mathrm{im}(\mathrm{H}(\alpha)) = \mathbb{K}[\Omega]^G$ iff $\alpha \in \mathbb{K}[G]^G$. The constant function $\underline{1} : G \to \mathbb{K}, h \mapsto 1$ is idempotent and spans linear $\mathbb{K}[G]^G = \mathbb{K}.\underline{1}$. Since Ω is an affine variety, there is a projection of $\mathbb{K}$-algebras $\pi : \mathbb{K}[X_1, \ldots, X_n] \to \mathbb{K}[\Omega]$. For $\mu \in \mathbb{N}^n$ define $X^\mu := X_1 \cdot \ldots \cdot X_n$, and $|\mu| := \sum_i \mu_i$. Then $\mathbb{K}[\Omega]^G$ is generated by the $J_\mu := \mathrm{H}(\underline{1})\pi(X^\mu)$ for all $\mu \in \mathbb{N}^n$ with $|\mu| \leq |G|$. ■

4. Deep learning classifications

Definition 7 *Let $n \in \mathbb{N}$. For a subset $X \subseteq \mathbb{R}^n$, we define $\bar{\mathrm{co}}(X) := \{\sum_{i=1}^m \alpha_i x_i : x_i \in X, m \in \mathbb{N}, \sum_{i=1}^m \alpha_i \leq 1, \alpha_i \geq 0\}$.*

After mathematical modeling, let $\mathsf{C} = (\Omega, (\Omega_l)_{l \in L})$ denote the related classification with $\Omega \subseteq \mathbb{R}^n$ compact, and $|L| = p < \infty$. Deep learning provides a computational framework allowing us to find a map $\nu = (\nu_1, \ldots, \nu_p)$ given by composition of N preactivation and activation functions, a so-called N-layer feedforward neural network that 'solves C accurately'. Assume the soft-max score function $\mathbb{R}^p \to \bar{\mathrm{co}}(e_1, \ldots, e_p) \subseteq \mathbb{R}^p$ in the $(N-1)$-th layer, that is $\nu : \Omega \to \bar{\mathrm{co}}(e_1, \ldots, e_p)$, where $e_1, \ldots, e_p$ denote the canonical linear basis of $\mathbb{R}^p$. In order to formulate precisely, what 'solving C accurately' means, in fact Ω must be a probability space, the cluster measurable, and ν a measurable map!

A resulting ν is accepted by the majority as an accurate solution if $\mathrm{E}[\nu_m \mid \Omega_l] = \delta_{ml} e_l$ holds for all $m, l = 1, \ldots, p$, where $\mathrm{E}[. \mid .]$ denotes the conditional expectation, and $\delta_{.,.}$ the

Kronecker delta. Heuristically, when ν solves C accurately we have $\nu^{-1}\big(\nu(\omega)\big) \approx \Omega_l$ for almost all $\omega \in \Omega_l$, for each $l \in L$. In particular, ν_i is almost constant almost everywhere on Ω_l, for each $l \in L$, for all $i = 1, \ldots, p$.

This consideration shows that when talking about classification, the distribution of data cannot be ignored! In Section 3 the distribution of data was assumed to be uniformly distributed implicitly which can be seen particularly well in the formula of the operator $\mathrm{H}(\underline{1})$ used in the proof of Theorem 6.

5. Classifications considering data distribution

For the sake of simplicity let G be a compact group. Let $L^1(G)$ denote the space of measurable functions on G which are integrable with respect to normalized Haar measure, where, as usual, two such functions are identified if they are equal almost everywhere. The Banach space $L^1(G)$ becomes a Banach *algebra* when equipped with the *convolution product*. We shall require some concepts from the theory of Banach algebras, and for these refer to Bonsall and Duncan (2012). In particular, Banach modules are defined in Definition 9.12. If $(\Omega, \mathcal{F})$ is a measurable space (where $\mathcal{F}$ is a σ-algebra of subsets of Ω) and Ω is a G-space, then $(\Omega, \mathcal{F})$ is a measurable G-space if the map $(x, \omega) \mapsto x \cdot \omega$ is measurable. To keep it technically as simple as possible, you can think of $\mathcal{F}$ as a Borel σ-algebra induced by a Fréchet topology.

Let Ω be a measurable G-space and $\mathbb{P}$ be a quasi-invariant probability measure on Ω. For the measurable quotient of $(\Omega, \mathcal{F})$ by G, we write $\pi : (\Omega, \mathcal{F}) \to (\Omega/G, \mathcal{F}/G)$.

To understand this subsection as a continuation of the previous subsections, here, we assume that the decomposition of the sample space Ω into cluster is induced by an action of a compact group G that respects the distribution $\mathbb{P}$ of the data.

As pointed out in Section 4, it is natural to consider neural networks as random variables. Then the 'almost constant almost everywhere on Ω_l for each $l \in L$' condition translates to the requirement of minimal variance on the cluster.

When starting with an arbitrary random variable $f \in L^1(\Omega, \mathbb{P})$ the canonical way to minimize the variance of f on the orbits is given by the conditional expectation with respect to π. The conditional exception with respect to the σ-algebra generated by π gives rise to an almost sure unique, idempotent linear operator $\mathbf{E}[? \mid \Omega/G] : L^1(\Omega, \mathbb{P}) \to L^1(\Omega, \mathbb{P})$. By definition, for $\mathbf{E}[f \mid \Omega/G]$ the variance on the orbits are minimal, in fact 0. So we find a linear operator that transforms arbitrary (L^1-)random variables to a class of *universal approximators* (Rosenblatt (1958)) as the source for functions of our interest as pointed out in Section 3 and Section 4. Unfortunately, the class we end up this way is given by the simplest feedforward neural networks with just one hidden layer and Binary step as activation function (Figure 1).

If μ is a measure on the measurable G-space Ω and $g \in G$, then the translate of μ by g will be denoted by μ_g and is defined by $\mu_g(E) = \mu(g^{-1} \cdot E)$ for each measurable set E contained in Ω. Since $\mathbb{P}$ is assumed to be a quasi-invariant probability measure, $L^1(\Omega, \mathbb{P})$ is a Banach left $L^1(G)$-module under the action of $L^1(G)$ defined via the Banach algebra morphism

$$\mathrm{H} : L^1(G) \to \mathcal{B}(L^1(\Omega, \mathbb{P})), \alpha \mapsto \left(\xi \mapsto \int_G \alpha(g) P(g, ?) \xi(g^{-1} \cdot ?) dg \right),$$

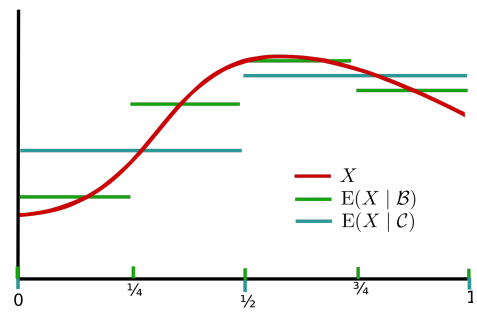

Figure 1: Conditional expectation with respect to σ-algebras

.

where $\mathcal{B}(L^1(\Omega, \mathbb{P}))$ denotes the bounded linear operators on $L^1(\Omega, \mathbb{P})$ and $P(g, \omega) = (d\mathbb{P}_g/d\mathbb{P})(\omega)$ is the Radon-Nikodym derivative.

As you can see, we have for compact groups as for finite groups (in the proof of Theorem 6) the *regular* representation of the group algebra of the compact group G. We can derive this representation even for locally compact groups!

For compact G the idempotent function $\underline{1} : G \to \mathbb{C}, h \mapsto 1$ belongs to $L^1(G)$. Analog, we end up with the projection $\mathrm{H}(\underline{1})$. In fact, this is an explicit description of the conditional expectation $\mathbf{E}[? \mid \Omega/G]$!

If we compare the projection in the proof of Theorem 6 with this one we notice that the formula basically differs in P. This discussion leads to the conclusion that the cocycle P given by the Radon-Nikodym derivative captures the distribution of the data.

In the non-compact case, say locally compact case, the constant function $\underline{1}$ is no more available in the domain of H. But we can evaluate (L^1)-random variables u of G with expected value 1. They give rise to a new class of functions that are closely related to the original classification problem:

Definition 8 *Let G be a locally compact group, and $E := \{u \in L^1(G) \ : \ \mathbf{E}[u] = 1\}$. The set of expected-to-be-invariant functions is defined as $\bigcup_{u \in E} \mathrm{im}\,(\mathrm{H}(u)) \subseteq L^1(\Omega, \mathbb{P})$.*

6. Motivation

Our considerations have shown that classification problems induced by group actions are very closely related to their associated regular representations of group algebras! Even the invariant ring naturally arises of the representation theory of the regular representation.

Unfortunately, the image of $\mathrm{H}(\underline{1})$ when defined is not well suited for applications since the resulting functions are badly realizable in the application as demonstrated in the discussion.

As is all too often the case, the solution lies in the problem itself. Let us recall these famous examples Figure 2. Or think of smoothly morphing an image to an another image. Thus one comes to the conclusion that in applications the requirement of a disjoint decomposition of the sample space into cluster is unrealistic. And if one thinks about it carefully it is also the reason that the resulting functions $\mathbf{E}[f \mid \Omega/G]$ of the conditional expecta-

tion operator are badly realizable in applications. Namely when thinking heuristically, the disjoint decomposition causes the 'jumps' in the graph of $\mathbf{E}[f \mid \Omega/G]$.

(a) Witch or young lady

(b) Vase or faces

Figure 2: Ambiguous data.

6.1. Probability kernel

Lets have a closer look at the idempotent integral operator $\mathrm{H}(\underline{1})$. For $f \in L^1(\Omega, \mathbb{P})$, $\omega \in \Omega$ we can rewrite:

$$\begin{aligned} \mathrm{H}(\underline{1})(f)(\omega) &= \int_G P(g,\omega) f(g^{-1} \cdot \omega) dg \\ &= \int_\Omega k(\omega, v) f(v) d\mathbb{P}(v), \end{aligned}$$

where $k : \Omega \times \Omega \to \mathbb{R}$ is defined by

$$k(\omega, v) = \begin{cases} P(g,\omega) & \text{if } \exists\, g \in G \,:\, v = g^{-1}\omega, \\ 0 & \text{else.} \end{cases}$$

The cocycle over the G-action P can be interpreted as kind of probability property, that is, $k(\omega, v)$ can be seen as the probability to decide w is related to a given v respectively the occurrence of v. Further, $\mathrm{H}(\underline{1})$ is an idempotent operator, that is $\mathrm{H}(\underline{1})^2 = \mathrm{H}(\underline{1})$ and P satisfies the cocycle identities

$$\begin{aligned} P(gh,\omega) &= P(h, g^{-1} \cdot \omega) P(g,\omega), \qquad \text{('local')} \\ P(e,\omega) &= 1 \qquad (g, h \in G, \text{ a.e. } \omega \in \Omega). \end{aligned}$$

These properties translates to

$$\begin{aligned} k(\omega,\omega) &= 1, \\ k(\omega, v) &= \int_\Omega k(\omega, z) k(z, v) d\mathbb{P}(z) \qquad \text{('global')} \end{aligned}$$

for almost all $\omega, v \in \Omega$. In this way, we end up with an integral operator with respect to a probability kernel that have densities with respect to some reference measure.

Working out the defining properties of P was straightforward. We just have to keep in mind that the local property could not be transferred yet. Now, lets tackle the problem of disjoint cluster. As pointed out in Section 6 disjoint cluster make not so much sense in application oriented problems. In particular, considering a group action as a starting point of a classification problem makes no sense. We have to switch to the more general notion of a groupoid!

7. Groupoids

7.1. Definitions and Notation

A group can be understood as a symmetry transformation that relates an object isomorphically to itself (the symmetries of an object, for example, the isometries of a polyhedron), while a groupoid is a collection of symmetry transformations acting between possibly more than one object (nLab authors (2022)).

Definition 9 *A groupoid $\mathcal{G}$ is a small category in which every morphism is invertible. We denote by $\mathcal{G}$ its set of morphisms, $\mathcal{G}_0$ its set of objects, $s : \mathcal{G} \to \mathcal{G}_0, \gamma \mapsto \gamma^{-1}\gamma$ its source map, and by $r : \mathcal{G} \to \mathcal{G}_0, \gamma \mapsto \gamma\gamma^{-1}$ its range map.*

Example 1 *Let Q be an arbitrary quiver (see Definition in Appendix B.1). We double the quiver Q by adding to each arrow $\alpha : x \to y$ in Q an arrow $\alpha^* : y \to x$ and obtain the doubled quiver $\bar{Q}$. Requiring that the arrows α, α^* are inverses of each other in the category of paths $\mathcal{P}\bar{Q}$ for all $\alpha \in Q_a$ gives rise to a groupoid $\mathcal{G}_Q$. In fact, every groupoid $\mathcal{G}$ can be obtained in that way.*

Remark 10 *In the following we will be interested in groupoids $\mathcal{G}_Q$ where the map $Q_1 \times_{s,r} Q_1 \to Q_0 \times Q_0, (\beta, \alpha) \mapsto (r(\beta), d(\alpha))$ is one-to-one. Thereby, $Q_1 \times_{s,r} Q_1 \subseteq Q_1 \times Q_1$ denotes the subset of composable arrows.*

Remark 11 *Let $\mathcal{G}$ be a groupoid. It induces the subset of composable morphisms $\mathcal{G} \times_{s,r} \mathcal{G} \subseteq \mathcal{G} \times \mathcal{G}$ and functions $\mathcal{G} \times_{s,r} \mathcal{G} \to \mathcal{G}, (\gamma, \omega) \mapsto \gamma\omega$, and $\mathcal{G} \to \mathcal{G}, \gamma \mapsto \gamma^{-1}$. These data give rise to a groupoid as Hahn (1978) defined (definition 1.1). And every groupoid in the sense of Hahn (1978) (definition 1.1) gives rise to a groupoid as we defined.*

Definition 12 *A topological groupoid $\mathcal{G}$ is a groupoid where the set of morphisms $\mathcal{G}$ and $\mathcal{G}_0$ are topological spaces, and all structure maps (source, range, composition, inverse) are continuous maps. Thereby, $\mathcal{G} \times_{s,r} \mathcal{G}$ has the induced topology from $\mathcal{G} \times \mathcal{G}$. It is called an etalé groupoid if in addition the source map s is a local homeomorphism.*

8. Twisted groupoid $*$-algebras

From now on, let $\mathcal{G}$ denote a topological groupoid whose topology is *Hausdorff* and locally compact. We also assume that the topology of the groupoid is second countable. We now recall the relevant notion from the cohomology theory for groupoids. For a detailed overview see Renault (2006) Section I.1.

Definition 13 *Let A be a topological abelian group with identity e_A. A continuous A-valued 2-cocycle on $\mathcal{G}$ is a continuous map $\sigma : \mathcal{G} \times_{r,s} \mathcal{G} \to A$ that satisfies the 2-cocycle identity: $\sigma(\alpha, \beta)\sigma(\alpha\beta, \gamma) = \sigma(\alpha, \beta\gamma)\sigma(\beta, \gamma)$ for all $(\alpha, \beta), (\beta, \gamma) \in \mathcal{G} \times_{r,s} \mathcal{G}$, and is called normalized if $\sigma(r(\gamma), \gamma) = \sigma(\gamma, s(\gamma)) = e_A$ for all $\sigma \in \mathcal{G}$. We write $Z^2(\mathcal{G}, A)$ for the group of continuous A-valued 2-cocycles on $\mathcal{G}$.*

We write $\mathbb{T}$ for the multiplicative group of complex numbers of modulus 1. As you can suggest a $\sigma \in Z^2(\mathcal{G}, \mathbb{T})$ will be the counterpart of the observed probability kernel in Section 6. We rediscover the 'local' property in the definition of the continuous $\mathbb{T}$-valued 2-cocycles. What is still missing is the 'global' property. To formulate this property we need to be able to integrate. We need compatible measures. Therefore, we refer to Renault (2006) for the definition of a left Haar system and quasi-invariant measure in this setting. Bourbaki's theory of integration on locally compact spaces is used (Bourbaki (2013), Bourbaki).

We add to the prerequisites that $\mathcal{G}$ is a locally compact groupoid with left Haar system $\{\lambda^u\}_{u \in \mathcal{G}_0}$. We revise our generic definition given at the beginning. The following definition is one of the main achievements of this paper:

Definition 14 *A classification problem consists of a locally compact groupoid $\mathcal{G}$ with a left Haar system $\{\lambda^u\}_{u \in \mathcal{G}_0}$, and a normalized continuous $\mathbb{T}$-valued 2-cocycle σ on $\mathcal{G}$ such that $\sigma(\alpha, \beta) = \int_{\gamma \in \mathcal{G}, s(\gamma) = r(\beta)} \sigma(\alpha, \gamma)\sigma(\gamma, \beta) d\lambda^{d(\alpha)}(\gamma)$ for all $\alpha, \beta \in \mathcal{G}$. The space $\mathcal{G}_0$ is called sample space, and σ is called labeling.*

Remark 15 *A labeling σ can be thought as continuation of a method for summing up all possible random walks as suggested in Figure 3 (physicstravelguide.com). Indeed, σ incorporates the uncertainty in classification for ambiguous data.*

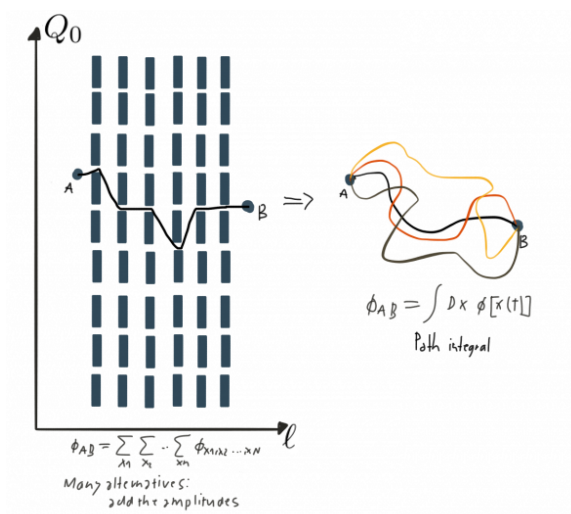

Figure 3: Labeling as a 'path integral'.

Let $C_c(\mathcal{G})$ denote the locally convex space of complex-valued continuous functions with compact support, endowed with the inductive limit topology (Jarchow (2012), Köthe (1983)).

For $f, g \in C_c(\mathcal{G})$ define $f*g(\alpha) \int f(\alpha\beta)g(\beta^{-1})\sigma(\alpha\beta, \beta^{-1})d\lambda^{d(\alpha)}(\beta)$, $f^*(\alpha) = \overline{f(\alpha^{-1})}\,\overline{\sigma(\alpha, \alpha^{-1})}$. Under these operations, $C_c(\mathcal{G})$ becomes a topological $*$-algebra (Renault (2006)).

Definition 16 *This topological $*$-algebra is denoted by $C_c(\mathcal{G}, \sigma)$ and called the σ-twisted groupoid $*$-algebra of $\mathcal{G}$.*

The $*$-algebras induced by cohomologous 2-cocycles induce isomorphic $*$-algebras.

These σ-twisted groupoid $*$-algebras are the equivalent counterparts of the $*$-algebras raised in the previous settings. As pointed out before, its regular representation play an essential role regarding the original classification problem.

8.1. Regular representation - Intuitively

The σ-twisted groupoid $*$-algebra $C_c(\mathcal{G}, \sigma)$ operates on functions on the base $\mathcal{G}_0$. Let φ be a function on $\mathcal{G}$, and u a function on $\mathcal{G}_0$. Define

$$(\mathrm{Op}\,\varphi)u(x) := \int_{\alpha: y \to x} \varphi(\alpha)\sigma(\alpha, \alpha^{-1})u(y) \ \big(...\big).$$

Intuitively, if we think of the elements of $\mathcal{G}$ as 'arrows' on the base space $\mathcal{G}_0$, then this integral tells us to look at all the arrows α going into a given point $x \in \mathcal{G}_0$, evaluate the function u at the tail of each of those arrows, then move back to x and integrate over all arrows α with 'weight' given by φ and labeling given by σ (Da Silva and Weinstein (1999)).

9. Outlook

The insight of Definition 14 allows us to link unsupervised learning to random walks on the second groupoid cohomology with values in the multiplicative group of complex numbers of modulus 1.

Acknowledgments

The author is deeply indebted to his postdoctoral advisers Prof. Dr. A and Prof. Dr. B for their help, support, kindness and patience. The author was funded by X.

References

Ibrahim Assem, Daniel Simson, Andrzej Skowronski, et al. *Elements of the Representation Theory of Associative Algebras: Volume 1: Techniques of Representation Theory.* Cambridge University Press, 2006.

Frank F Bonsall and John Duncan. *Complete normed algebras*, volume 80. Springer Science & Business Media, 2012.

N Bourbaki. *Integration I: Chapters 1-6.* Springer Science & Business Media, 2013.

Nicolas Bourbaki. *Integration II: Chapters 7-9.* Springer.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Ana Cannas Da Silva and Alan Weinstein. *Geometric models for noncommutative algebras*, volume 10. American Mathematical Soc., 1999.

Peter Gabriel. Unzerlegbare darstellungen i. *Manuscripta mathematica*, 6(1):71–103, 1972.

Peter Gabriel and Andrei V Roiter. *Representations of finite-dimensional algebras*, volume 73. Springer Science & Business Media, 1992.

Peter Hahn. Haar measure for measure groupoids. *Transactions of the American Mathematical Society*, 242:1–33, 1978.

David Hilbert. Über die theorie der algebraischen formen. *Mathematische annalen*, 36(4): 473–534, 1890.

David Hilbert. Über die vollen invariantensysteme. In *Algebra· Invariantentheorie· Geometrie*, pages 287–344. Springer, 1933.

Hans Jarchow. *Locally convex spaces*. Springer Science & Business Media, 2012.

Gottfried Köthe. Topological vector spaces. In *Topological Vector Spaces I*, pages 123–201. Springer, 1983.

Hanspeter Kraft and A Wiedemann. *Geometrische methoden in der invariantentheorie*. Springer, 1985.

Saunders MacLane. *Homology*. Springer Science & Business Media, 2012.

Masayoshi Nagata. On the 14-th problem of hilbert. *American Journal of Mathematics*, 81 (3):766–772, 1959.

nLab authors. groupoid. https://ncatlab.org/nlab/show/groupoid, September 2022. Revision 86.

Emmy Noether. Der endlichkeitssatz der invarianten endlicher gruppen. *Mathematische Annalen*, 77(1):89–92, 1915.

physicstravelguide.com. Path integral quantum mechanics. URL https://physicstravelguide.com/theories/quantum_mechanics/path_integral#tab__intuitive.

Jean Renault. *A groupoid approach to C*-algebras*, volume 793. Springer, 2006.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

Appendix A. Equivariant functions

This section is an appendix for Section 3. As pointed out in Bronstein et al. (2021), beside invariant functions also equivariant functions get into the focus of interest. Let $\chi : G \to \mathbb{G}_m := \mathrm{Gl}_1(\mathbb{K})$ be a character of G, that is a group homomorphism. Then, we can define

$$\mathbb{K}[\Omega]^{G,\chi} := \{f \in \mathbb{K}[\Omega] \ : \ f(g.\omega) = \chi(g)f(\omega) \ \ \forall g \in G, \, \omega \in \Omega\},$$

and $\mathbb{K}[\Omega]^G_\chi := \bigoplus_{n\geq 0} \mathbb{K}[\Omega]^{G,\chi^n}$. By adjusting the sample-space to $\Omega \times \mathbb{A}^1$ and the action of G to $g.(\omega, t) := (g.\omega, \chi(g^{-1})t)$, $g \in G, \omega \in \Omega, t \in \mathbb{A}^1$, we can realize $\mathbb{K}[\Omega]^G_\chi$ as the invariant ring of this adjusted action, that is, we have $\mathbb{K}[\Omega \times \mathbb{A}^1]^G \ \simeq \ \mathbb{K}[\Omega]^G_\chi$.

Appendix B. Regular representation of twisted groupoid $*$-algebras - technically

At first we need to recall basic notions for algebras, modules and categories.

B.1. Algebras, modules, categories

In this section, we will follow closely chapter 2 of Gabriel and Roiter (1992). For the definition of a category and functor we refer to MacLane (2012). Throughout this section $\mathbb{K}$ denotes a fixed commutative ring.

Definition 17

i) A $\mathbb{K}$-category is category $\mathcal{A}$ whose morphism sets, here denoted by $\mathrm{Hom}(X, Y)$ or $\mathcal{A}(X, Y)$, are endowed with $\mathbb{K}$-module structures such that the composition maps are $\mathbb{K}$-linear. A $\mathbb{K}$-functor between $\mathbb{K}$-categories $\mathcal{A}$ and $\mathcal{B}$ is a functor $F : \mathcal{A} \to \mathcal{B}$ whose defining maps $F(X, Y) : \mathcal{A}(X, Y) \to \mathcal{B}(FX, FY)$ are $\mathbb{K}$-linear for all $X, Y \in \mathcal{A}$.

ii) A $\mathbb{K}$-category $\mathcal{A}$ is called svelte whose isomorphism classes will form a set.

iii) Let $\mathcal{C}$ be an arbitrary category. By $\mathbb{K}\mathcal{C}$ we denote its linearization, that is the $\mathbb{K}$-category with same objects as $\mathcal{C}$, and its morphism-spaces $\mathbb{K}\mathcal{C}(X, Y)$ are the free $\mathbb{K}$-modules with bases $\mathcal{C}(X, Y)$.

Remark 18

i) Each $\mathbb{K}$-algebra A gives rise to a svelte $\mathbb{K}$-category which has one object Ω, and satisfies $\mathrm{Hom}(\Omega, \Omega) = A$. Each svelte $\mathbb{K}$-category $\mathcal{A}$ gives rise to a $\mathbb{K}$-algebra $|\mathcal{A}|$ in the obvious way (unique up to isomorphism of $\mathbb{K}$-algebras). In the sequel, we shall identify svelte $\mathbb{K}$-categories with associated $\mathbb{K}$-algebras, and vice versa.

ii) A functor $E : \mathcal{C} \to \mathcal{A}$ of categories whose range is a $\mathbb{K}$-category uniquely extends to a $\mathbb{K}$-functor $F : \mathbb{K}\mathcal{C} \to \mathcal{A}$.

Definition 19 (Gabriel (1972))

i) Let Q be a quiver, that is, a set of vertices connected by arrows as illustrated in Figure 4. We denote by Q_v the set of vertices of Q, by Q_a the set of arrows. Thus we have two maps, $d, r : Q_a \to Q_v$ which map an arrow α onto its domain $d(\alpha)$ and its range $r(\alpha)$ ($d(\alpha) = x$, $r(\alpha) = z$ in Figure 4).

ii) A quiver Q gives rise to the category of paths $\mathcal{P}Q$ whose set of objects is Q_v, whereas each morphism set $\mathcal{P}Q(x, y)$ consists of the paths with origin x and terminus y. The composition is the juxtaposition of paths. We call the linearization $\mathbb{K}\mathcal{P}Q$ of $\mathcal{P}Q$ the $\mathbb{K}$-category of paths of Q and simply denote by $\mathbb{K}Q$.

The $\mathbb{K}$-category of paths of Q will play an important role in the sequel.

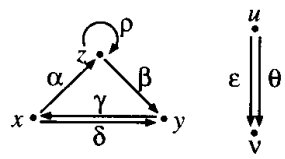

Figure 4: Quiver

Definition 20 *A (left) module M over a $\mathbb{K}$-category $\mathcal{A}$ is a covariant $\mathbb{K}$-functor $M : \mathcal{A} \to \mathbb{K}$-Mod, where $\mathbb{K}$-Mod denotes the $\mathbb{K}$-category of $\mathbb{K}$-modules.*

Remark 21 *If the $\mathbb{K}$-category $\mathcal{A}$ is svelte, the modules over $\mathcal{A}$ and their morphism again form a $\mathbb{K}$-category which we denote by $\mathcal{A}$-Mod.*

Example 2 (Representation of a quiver) *A representation of a quiver Q over $\mathbb{K}$ consists of $\mathbb{K}$-modules $V(x)$, one for each $x \in Q_v$, and for $\mathbb{K}$-linear maps $V(\alpha) : V(x) \to V(y)$, one for each arrow $(\alpha : x \to y) \in Q_a$. The maps $x \mapsto V(x)$ and $\alpha \mapsto V(\alpha)$ uniquely extend to a $\mathbb{K}$-functor $\mathbb{K}Q \to \mathbb{K}$-Mod, which we still denote by V. Thus we identify representation of Q with left modules over $\mathbb{K}Q$. This identification also allows us to transfer the notion of morphism from left modules to representations.*

Lemma 22 *There is an equivalence of $\mathbb{K}$-categories, induced by the $\mathbb{K}$-functor*

$$| \, . \, | : \mathbb{K}Q\text{-Mod} \to |\mathbb{K}Q|\text{-Mod}, V \mapsto \bigoplus_{x \in Q_v} V(x),$$

where $|\mathbb{K}Q|$-Mod denotes the $\mathbb{K}$-category of (left) modules over the path algebra of Q.

Proof A proof can be found for example in Assem et al. (2006). ■

B.2. Representation of σ-twisted groupoid algebras

The approach from Appendix B.1 will be basically transferred: The role of the $\mathbb{K}$-categories will be taken over by $*$-algebras of groupoids and the base category $\mathbb{K}$-Mod will be substituted by the category of separable Hilbert spaces with unitary operators.