Geometric and Algebraic Aspects of Machine Learning

Giovanni Luca Marchetti

Royal Institute of Technology (KTH)

I. Overview

Bio

- B.Sc. & M.Sc.: pure mathematics at La Sapienza (Rome)
- Ph.D.: machine learning and robotics at KTH (Stockholm)
- PostDoc: theoretical machine learning at KTH (Stockholm)

Bio

- B.Sc. & M.Sc.: pure mathematics at La Sapienza (Rome)
- Ph.D.: machine learning and robotics at KTH (Stockholm)
- PostDoc: theoretical machine learning at KTH (Stockholm)

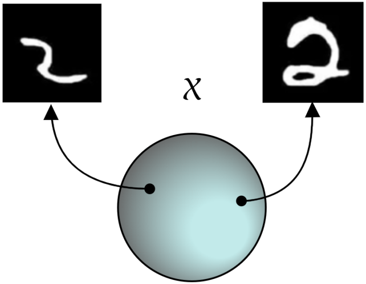
Research Interests

Geometric methods for machine learning and high-dimensional statistics:

- Geometric deep learning
- Manifold/representation learning
- Computational geometry & topology
- Statistical learning theory
- Robotics

The Geometry of Data

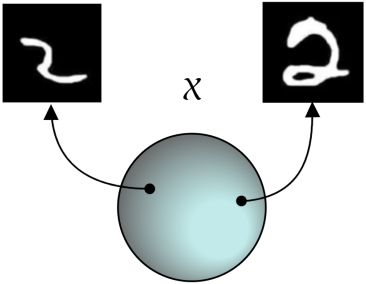
Data naturally carry geometric structure.



A data-set

The Geometry of Data

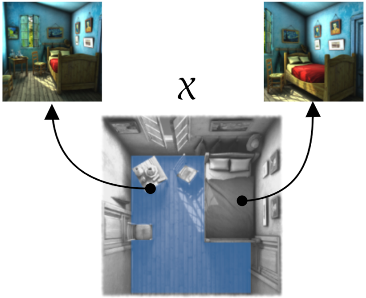
Data naturally carry geometric structure.



A data~~-set~~-manifold

The Geometry of Data

Data naturally carry geometric structure.



A data~~-set~~-manifold

The Geometry of Data

Data naturally carry geometric structure.

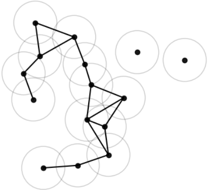
$\mathcal{X}$

Question: How can the geometric structure be extracted statistically from data and exploited for inference?

A data-~~set~~-manifold

Geometric Methods

Geometric structure can be extracted in several forms.



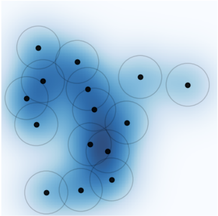
- Combinatorial: graphs and simplicial complexes [3]

[3] Edelsbrunner et al., *"Computational topology: an introduction"*, 2022
[21] Scott, *"Multivariate density estimation: theory, practice, and visualization"*, 2015
[1] Bengio et al., *"Representation learning: A review and new perspectives"*, 2013

Geometric Methods

Geometric structure can be extracted in several forms.

- Combinatorial: graphs and simplicial complexes [3]
- Fuzzy: density estimators [21]

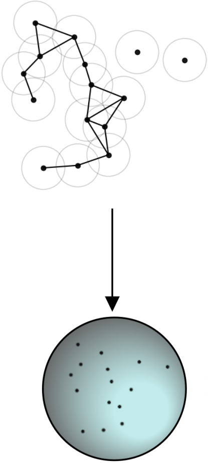


[3] Edelsbrunner et al., *"Computational topology: an introduction"*, 2022
[21] Scott, *"Multivariate density estimation: theory, practice, and visualization"*, 2015
[1] Bengio et al., *"Representation learning: A review and new perspectives"*, 2013

Geometric Methods

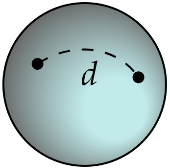
Geometric structure can be extracted in several forms.

- Combinatorial: graphs and simplicial complexes [3]
- Fuzzy: density estimators [21]
- Smooth: (neural) representations [1]

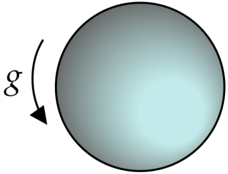


[3] Edelsbrunner et al., *"Computational topology: an introduction"*, 2022
[21] Scott, *"Multivariate density estimation: theory, practice, and visualization"*, 2015
[1] Bengio et al., *"Representation learning: A review and new perspectives"*, 2013

Symmetries and Metrics

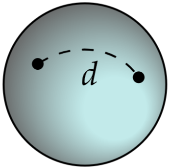


Metrics

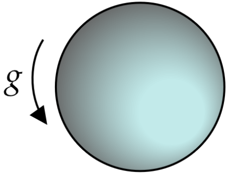


Symmetries

Symmetries and Metrics



Metrics



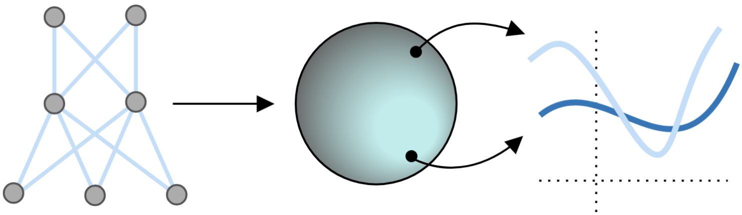
Symmetries

- Local in nature
- Can be inferred from data
- Suitable for non-parametric approaches

- Global in nature
- Unknown a priori
- Require powerful parametric approaches

The Geometry of Models

Machine learning models parametrize spaces of functions, whose (continuous) geometry is central in learning aspects.



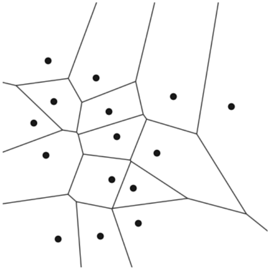
II. Geometric Density Estimation

Voronoi Tessellations

Definition

Let $\mathcal{X}$ be a metric space and $P \subseteq \mathcal{X}$ finite. The Voronoi cell of $p \in P$ is defined by the nearest-neighbor relation:

$$C(p) = \{x \in \mathcal{X} \mid \forall q \in P \; d(x,q) \geq d(x,p)\}.$$



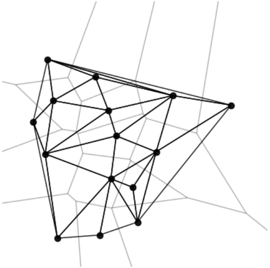
Voronoi Tessellations

Definition

Let $\mathcal{X}$ be a metric space and $P \subseteq \mathcal{X}$ finite. The Voronoi cell of $p \in P$ is defined by the nearest-neighbor relation:

$$C(p) = \{x \in \mathcal{X} \mid \forall q \in P \; d(x,q) \geq d(x,p)\}.$$

The Delaunay triangulation is dual to the Voronoi tessellation.



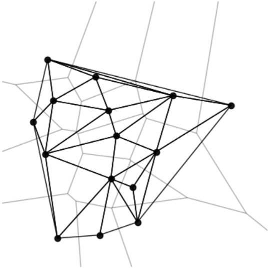
Voronoi Tessellations

Definition

Let $\mathcal{X}$ be a metric space and $P \subseteq \mathcal{X}$ finite. The Voronoi cell of $p \in P$ is defined by the nearest-neighbor relation:

$$C(p) = \{x \in \mathcal{X} \mid \forall q \in P \; d(x, q) \geq d(x, p)\}.$$

The Delaunay triangulation is dual to the Voronoi tessellation.



$C(p)$ is:

- ✓ An arbitrary convex polytope
- ✓ An adaptive neighbourhood of p
- ✗ Expensive to compute

Voronoi Density Estimator (VDE)

Non-Parametric Density Estimation

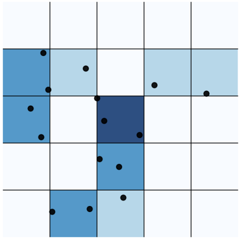
Estimate a closed-form density $\hat{\rho} \in \mathrm{L}^1(\mathcal{X})$ from finite samples $P \subseteq \mathcal{X}$, $P \sim \rho$.

[15] Ord, *"How many trees in a forest"*, 1978
[16] Pearson, *"Contributions to the mathematical theory of evolution"*, 1894
[20] Rosenblatt, *"Remarks on Some Nonparametric Estimates of a Density Function"*, 1956

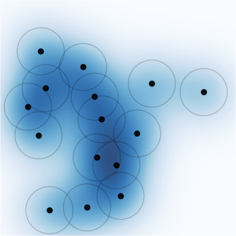
Voronoi Density Estimator (VDE)

Non-Parametric Density Estimation

Estimate a closed-form density $\hat{\rho} \in \mathrm{L}^1(\mathcal{X})$ from finite samples $P \subseteq \mathcal{X}$, $P \sim \rho$.



Histograms [16]



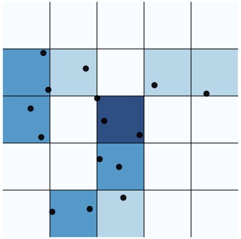
KDE [20]

[15] Ord, *"How many trees in a forest"*, 1978
[16] Pearson, *"Contributions to the mathematical theory of evolution"*, 1894
[20] Rosenblatt, *"Remarks on Some Nonparametric Estimates of a Density Function"*, 1956

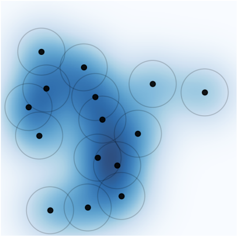
Voronoi Density Estimator (VDE)

Non-Parametric Density Estimation

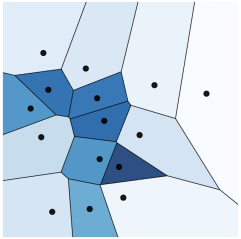
Estimate a closed-form density $\hat{\rho} \in \mathrm{L}^1(\mathcal{X})$ from finite samples $P \subseteq \mathcal{X}$, $P \sim \rho$.



Histograms [16]



KDE [20]



VDE [15]

In VDE, $\hat{\rho}(x)$ is inversely proportional to the volume of Voronoi cell $C(p) \ni x$.

[15] Ord, *"How many trees in a forest"*, 1978
[16] Pearson, *"Contributions to the mathematical theory of evolution"*, 1894
[20] Rosenblatt, *"Remarks on Some Nonparametric Estimates of a Density Function"*, 1956

Convergence

In [17], we prove convergence of Voronoi-based estimators to the ground-truth density ρ.

[17] Polianskii*, Marchetti* et al., *"Voronoi Density Estimator for High-Dimensional Data"*, UAI 2022.

Convergence

In [17], we prove convergence of Voronoi-based estimators to the ground-truth density ρ.

Theorem

Consider a density estimator such that

$$\int_{C(p)} \hat{\rho} = \frac{1}{|P|}$$

for every $p \in P$. *If* $P \sim \rho$, *then* $\hat{\rho} \to \rho$ *in an appropriate sense as* $|P| \to \infty$.

[17] Polianskii*, Marchetti* et al., *"Voronoi Density Estimator for High-Dimensional Data"*, UAI 2022.

Convergence

In [17], we prove convergence of Voronoi-based estimators to the ground-truth density ρ.

Theorem

Consider a density estimator such that

$$\int_{C(p)} \hat{\rho} = \frac{1}{|P|}$$

for every $p \in P$. If $P \sim \rho$, then $\hat{\rho} \to \rho$ in an appropriate sense as $|P| \to \infty$.

In contrast, KDE requires asymptotically-vanishing bandwidth for convergence.

[17] Polianskii*, Marchetti* et al., *"Voronoi Density Estimator for High-Dimensional Data"*, UAI 2022.

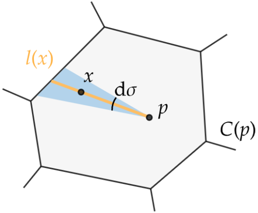
Improving the Voronoi Density Estimator

VDE is:

- ✓ Adaptive and convergent
- ✗ Expensive to compute
- ✗ Discontinuous

[11] Marchetti et al., *"An Efficient and Continuous Voronoi Density Estimator"*, AISTATS 2023 Notable Paper Award

Improving the Voronoi Density Estimator



VDE is:

- ✓ Adaptive and convergent
- ✗ Expensive to compute
- ✗ Discontinuous

In [11], we improve this by exploiting the radial geometry of Voronoi tessellations: $l(x)$ is continuous and computable in $O(|P|)$.

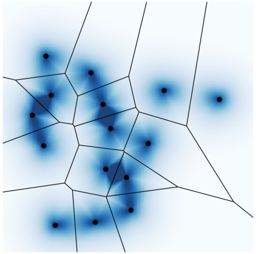
[11] Marchetti et al., *"An Efficient and Continuous Voronoi Density Estimator"*, AISTATS 2023 Notable Paper Award

Radial Voronoi Density Estimator

Given a kernel K, the estimated density is $K(\beta d(x, p))$, where $x \in C(p)$ and the radial bandwidth $\beta = \beta(l(x))$ is defined implicitly via an integral equation:

Radial Voronoi Density Estimator

Given a kernel K, the estimated density is $K(\beta d(x, p))$, where $x \in C(p)$ and the radial bandwidth $\beta = \beta(l(x))$ is defined implicitly via an integral equation:



RVDE

$$\underbrace{\int_0^{l(x)} t^{n-1} K(\beta t)\, \mathrm{d}t}_{\textit{Conical integral}} = \text{const}$$

In particular, $\int_{C(p)} \hat{\rho}$ is constant, implying convergence.

Modes and Performance

The map $I \mapsto \beta(I)$ generalizes Lambert's function. In particular:

Theorem

The modes *of RVDE are located among nodes and edges of the Delaunay triangulation.*

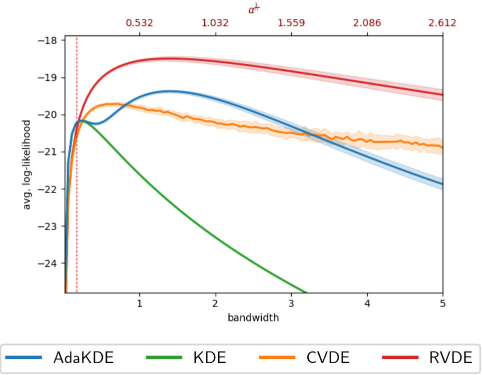
Modes and Performance

The map $l \mapsto \beta(l)$ generalizes Lambert's function. In particular:

Theorem

The modes of RVDE are located among nodes and edges of the Delaunay triangulation.

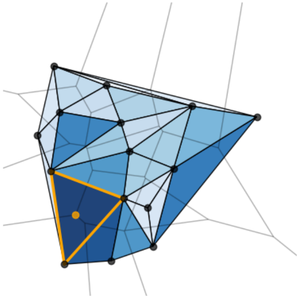
RVDE performs well:



Related Works

We have explored similar ideas for:

Non-parametric active learning via Delaunay triangulations [9].

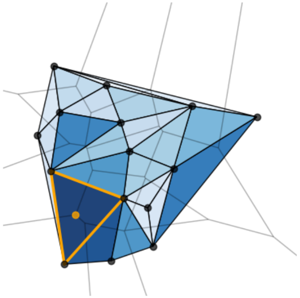


[9] Kravberg*, Marchetti* et al., *"Active Nearest Neighbor Regressor Through Delaunay Refinement"*, ICML 2022
[4] Garcia-Castellanos*, Medbouhi*, Marchetti et al., *"HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees"*, ALENEX 2025

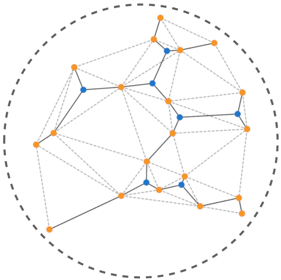
Related Works

We have explored similar ideas for:

Non-parametric active learning via Delaunay triangulations [9].



Statistics in non-Euclidean geometries, in particular hyperbolic spaces [4].

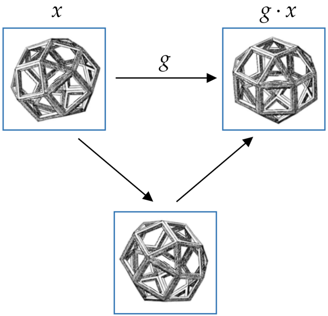


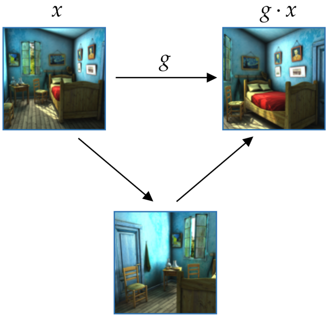
[9] Kravberg*, Marchetti* et al., *"Active Nearest Neighbor Regressor Through Delaunay Refinement"*, ICML 2022
[4] Garcia-Castellanos*, Medbouhi*, Marchetti et al., *"HyperSteiner: Computing Heuristic Hyperbolic Steiner Minimal Trees"*, ALENEX 2025

III. Symmetry-Based Representation Learning

Symmetries of Data

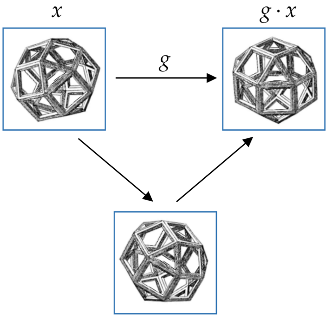
Data naturally exhibit symmetries.

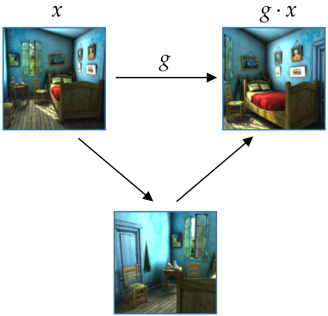




Symmetries of Data

Data naturally exhibit symmetries.

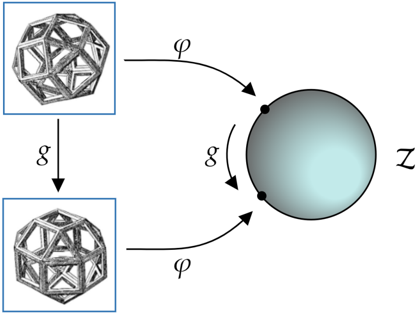




Symmetries are modelled as an unknown action $G \times \mathcal{X} \to \mathcal{X}$ by a group G on $\mathcal{X}$.

Equivariant Representations

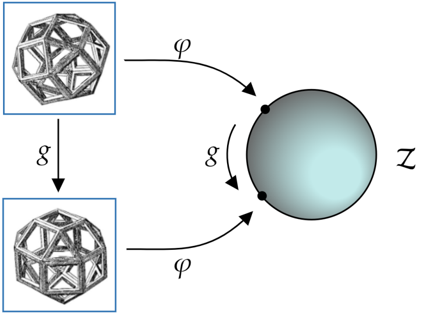
A representation respecting symmetries is deemed equivariant.



[2] Bronstein et al., *"Geometric deep learning: Grids, groups, graphs, geodesics, and gauges"*, 2021
[8] Kipf et al., *"Contrastive learning of structured world models"*, 2020

Equivariant Representations

A representation respecting symmetries is deemed equivariant.

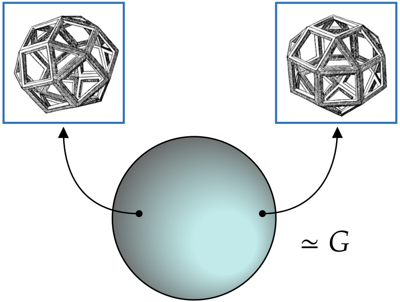


They arise in:

- Convolutional and graph neural networks [2]
- World models, incorporating interactions into representations [8]

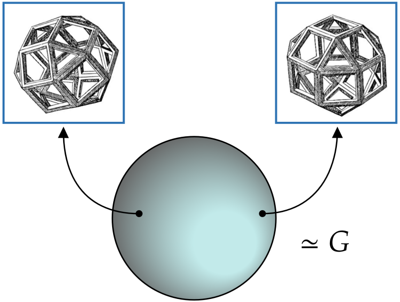
[2] Bronstein et al., *"Geometric deep learning: Grids, groups, graphs, geodesics, and gauges"*, 2021
[8] Kipf et al., *"Contrastive learning of structured world models"*, 2020

Geometry from Symmetries



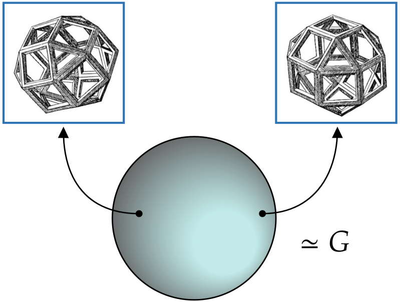
- Group actions determine classes deemed orbits $\mathcal{E} = \mathcal{X}/G$

Geometry from Symmetries



- Group actions determine classes deemed orbits $\mathcal{E} = \mathcal{X}/G$
- For free group actions, each orbit is isomorphic to G and thus $\mathcal{X} \simeq \mathcal{E} \times G$

Geometry from Symmetries



- Group actions determine classes deemed orbits $\mathcal{E} = \mathcal{X}/G$
- For free group actions, each orbit is isomorphic to G and thus $\mathcal{X} \simeq \mathcal{E} \times G$
- Every orbit-preserving equivariant map $\varphi \colon \mathcal{X} \to \mathcal{E} \times G$ is an isomorphism

Equivariant Isomorphic Networks (EquIN)

In [13], we propose to learn a representation $\varphi \colon \mathcal{X} \to \mathcal{E} \times G$. Given triplet data $(x, g, y = g \cdot x)$, EquIN optimizes:

[13] Marchetti et al., *"Equivariant Representation Learning via Class-Pose Decomposition"*, AISTATS 2023
[6] Higgins et al., *"Towards a definition of disentangled representations"*, 2018

Equivariant Isomorphic Networks (EquIN)

In [13], we propose to learn a representation $\varphi\colon \mathcal{X} \to \mathcal{E} \times G$. Given triplet data $(x, g, y = g \cdot x)$, EquIN optimizes:

- A (contrastive) loss forcing invariance in $\mathcal{E}$.
- A loss forcing equivariance in G.

[13] Marchetti et al., *"Equivariant Representation Learning via Class-Pose Decomposition"*, AISTATS 2023
[6] Higgins et al., *"Towards a definition of disentangled representations"*, 2018

Equivariant Isomorphic Networks (EquIN)

In [13], we propose to learn a representation $\varphi\colon \mathcal{X} \to \mathcal{E} \times G$. Given triplet data $(x, g, y = g \cdot x)$, EquIN optimizes:

- A (contrastive) loss forcing invariance in $\mathcal{E}$.
- A loss forcing equivariance in G.

An ideal learner φ will infer an isomorphism. In particular, the representation is:

✓ Lossless

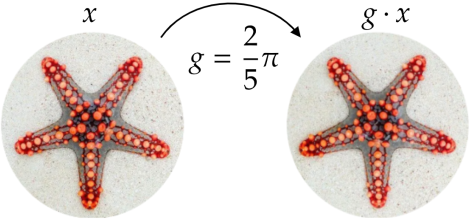
✓ Disentangled [6]

✗ Based on the assumption that G known a priori

[13] Marchetti et al., *"Equivariant Representation Learning via Class-Pose Decomposition"*, AISTATS 2023
[6] Higgins et al., *"Towards a definition of disentangled representations"*, 2018

Non-Free Group Actions

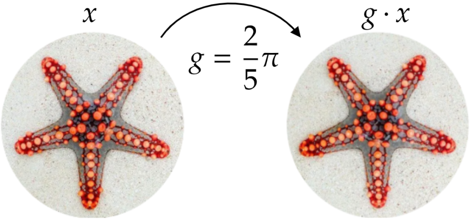
When data is stabilized, the action is not free.



[19] Rey*, Marchetti* et al., *"Equivariant Representation Learning in the Presence of Stabilizers"*, ECML 2023

Non-Free Group Actions

When data is stabilized, the action is not free.

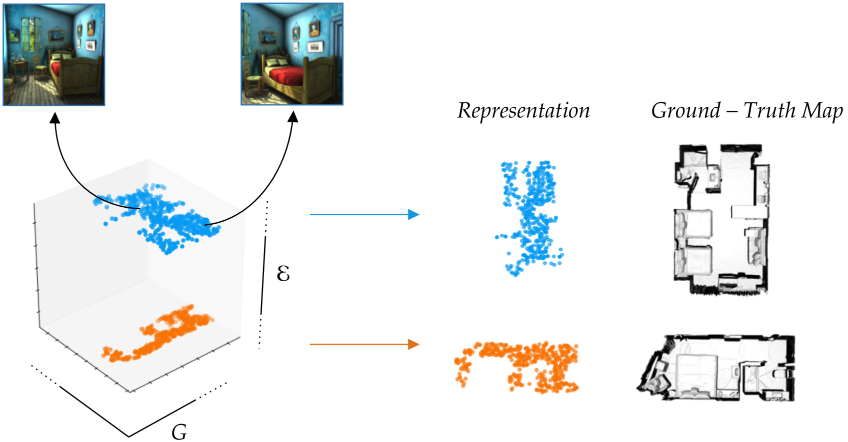


EquIN extends to this generality, but the theory is more subtle [19].

[19] Rey*, Marchetti* et al., *"Equivariant Representation Learning in the Presence of Stabilizers"*, ECML 2023

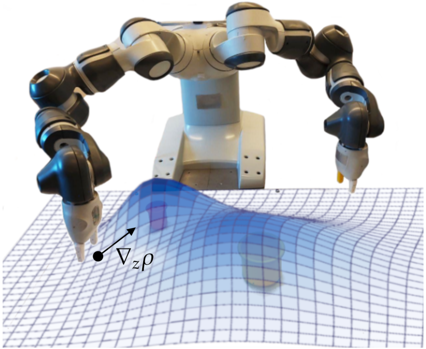
Extracting Geometry

EquIN extracts isometric maps of the world, which is crucial in robotic perception.



Applications to Robotics

We apply EquIN to vision-based manipulation in [18].



[18] Reichlin, Marchetti et al., "*Back to the Manifold: Recovery from Out-of-Distribution States*", IROS 2022

Abstract Harmonic Analysis

Suppose that G is unknown. Is it possible to discover symmetries from data?

Abstract Harmonic Analysis

Symmetries are intimately related to harmonic analysis:

Abstract Harmonic Analysis

Symmetries are intimately related to harmonic analysis: the Fourier transform can be defined over arbitrary groups G.

The Fourier transform decomposes functions $G \to \mathbb{C}$ into harmonics, i.e. (irreducible) unitary representations $G \to \mathrm{U}(\mathbb{C}^n)$.

Abstract Harmonic Analysis

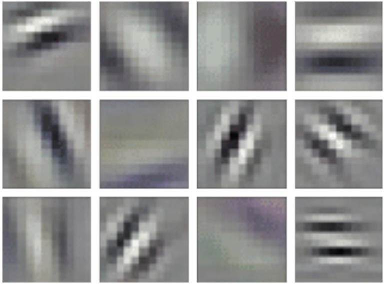
Symmetries are intimately related to harmonic analysis: the Fourier transform can be defined over arbitrary groups G.

The Fourier transform decomposes functions $G \to \mathbb{C}$ into harmonics, i.e. (irreducible) unitary representations $G \to \mathrm{U}(\mathbb{C}^n)$.

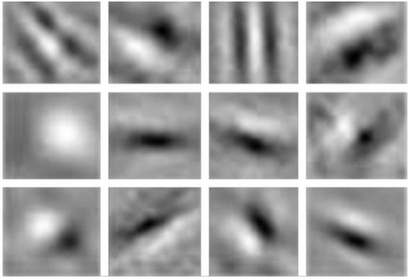
For commutative groups, $n = 1$, and $\mathrm{U}(\mathbb{C})$ is a circle.

Emergence of Harmonics

Harmonics are ubiquitous in both biological and artificial networks.



AlexNet [10]



Macaque [23]

[10] Krizhevsky et al., *"Imagenet classification with deep convolutional neural networks"*, 2012
[23] Zylberberg et al., *"A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields"*, 2011

Harmonics of Learning

In [14], we show that for a neural network $\varphi(W, x)$:

[14] Marchetti et al., "Harmonics of Learning: Universal Fourier Features Emerge in Invariant Networks", COLT 2024

Harmonics of Learning

In [14], we show that for a neural network $\varphi(W, x)$:

Theorem

If φ is invariant in x w.r.t. G, then each neuron learns a harmonic of G. In particular, if W is orthonormal, it coincides with the Fourier transform.

[14] Marchetti et al., "Harmonics of Learning: Universal Fourier Features Emerge in Invariant Networks", COLT 2024

Harmonics of Learning

In [14], we show that for a neural network $\varphi(W, x)$:

Theorem

If φ is invariant in x w.r.t. G, then each neuron learns a harmonic of G. In particular, if W is orthonormal, it coincides with the Fourier transform.

Theorem

If φ is 'almost invariant' and the W is 'almost orthonormal', then G can be recovered from W up to isomorphism.

[14] Marchetti et al., "Harmonics of Learning: Universal Fourier Features Emerge in Invariant Networks", COLT 2024

Harmonics of Learning

In [14], we show that for a neural network $\varphi(W, x)$:

Theorem

If φ is invariant in x w.r.t. G, then each neuron learns a harmonic of G. In particular, if W is orthonormal, it coincides with the Fourier transform.

Theorem

If φ is 'almost invariant' and the W is 'almost orthonormal', then G can be recovered from W up to isomorphism.

This explains the emergence of harmonics in learning systems via invariance, and can be exploited for symmetry discovery via contrastive learning.

[14] Marchetti et al., "Harmonics of Learning: Universal Fourier Features Emerge in Invariant Networks", COLT 2024

IV. Algebraic Geometry of Neural Networks

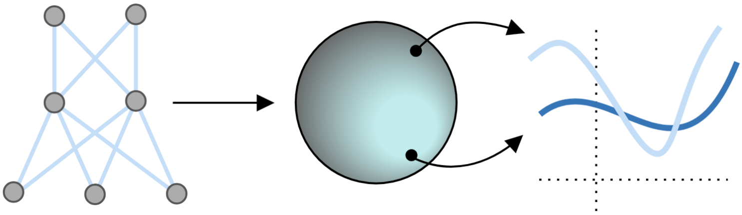
Neuromanifolds

Machine learning models parametrize spaces of functions,

Neuromanifolds

Machine learning models parametrize spaces of functions, referred to as:

Neuromanifolds	Hypothesis Spaces
Information geometry	Statistical learning theory



Algebraic Neuromanifolds

Consider a neural network with polynomial activation function.

Algebraic Neuromanifolds

Consider a neural network with polynomial activation function.

The neuromanifold can be defined by polynomial (in-) equalities, i.e. it is a (semi-) algebriac variety. These lie at the heart of algebraic geometry.

Algebraic Neuromanifolds

Consider a neural network with polynomial activation function.

The neuromanifold can be defined by polynomial (in-) equalities, i.e. it is a (semi-) algebriac variety. These lie at the heart of algebraic geometry.

By Weierstrass, any continuous activation function can be approximated by polynomial ones.

Neuroalgebraic Geometry

We refer to the study of neuromanifolds via algebraic geometry as neuroalgebraic geometry [12].

[12] Marchetti et al., "*An Invitation to Neuroalgebraic Geometry*", preprint 2025

Neuroalgebraic Geometry

We refer to the study of neuromanifolds via algebraic geometry as neuroalgebraic geometry [12].

Deep Learning	Algebraic Geometry
sample complexity, expressivity	dimension, degree
subnetworks	singularities
identifiability	fibers of the parametrization
learning dynamics	gradient flow, Morse theory

[12] Marchetti et al., "*An Invitation to Neuroalgebraic Geometry*", preprint 2025

Advances in Neuroalgebraic Geometry

We have recently studied the neuroalgebraic geometry of:

[22] Shahverdi* Marchetti* and Kohn*, "*On the Geometry and Optimization of Polynomial Convolutional Networks*", AISTATS 2025
[5] Henry* Marchetti* and Kohn*, "*Geometry of Lightning Self-Attention: Identifiability and Dimension*", ICLR 2025
[7] Kileel et al., "*On the expressive power of deep polynomial neural networks*", 2019

Advances in Neuroalgebraic Geometry

We have recently studied the neuroalgebraic geometry of:

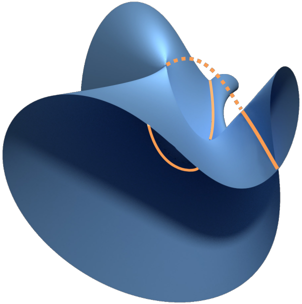


Figure: Neuromanifold of self-attention

- Deep polynomial convolutional networks [22]
- Deep lightning/linear self-attention networks (Transformers) [5]

[22] Shahverdi* Marchetti* and Kohn*, "*On the Geometry and Optimization of Polynomial Convolutional Networks*", AISTATS 2025
[5] Henry* Marchetti* and Kohn*, "*Geometry of Lightning Self-Attention: Identifiability and Dimension*", ICLR 2025
[7] Kileel et al., *"On the expressive power of deep polynomial neural networks"*, 2019

Advances in Neuroalgebraic Geometry

We have recently studied the neuroalgebraic geometry of:

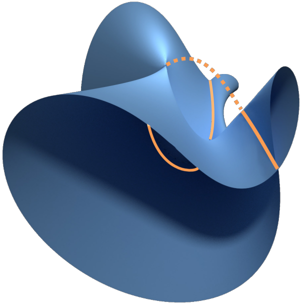


Figure: Neuromanifold of self-attention

- Deep polynomial convolutional networks [22]
- Deep lightning/linear self-attention networks (Transformers) [5]

For fully-connected networks, most problems are still open [7]!

[22] Shahverdi* Marchetti* and Kohn*, "*On the Geometry and Optimization of Polynomial Convolutional Networks*", AISTATS 2025
[5] Henry* Marchetti* and Kohn*, "*Geometry of Lightning Self-Attention: Identifiability and Dimension*", ICLR 2025
[7] Kileel et al., "*On the expressive power of deep polynomial neural networks*", 2019

Conclusions and Future Work

We have explored various geometric and algebraic aspects of deep learning and high-dimensional statistics, including:

Conclusions and Future Work

We have explored various geometric and algebraic aspects of deep learning and high-dimensional statistics, including:

- Non-parametric density estimation based on Voronoi cells.
- Group-equivariant representation learning, with applications to robotics.
- Theoretical deep learning via algebraic geometry.

Conclusions and Future Work

We have explored various geometric and algebraic aspects of deep learning and high-dimensional statistics, including:

- Non-parametric density estimation based on Voronoi cells.
- Group-equivariant representation learning, with applications to robotics.
- Theoretical deep learning via algebraic geometry.

Directions for future work:

- Voronoi cells for social networks/graphs.

Conclusions and Future Work

We have explored various geometric and algebraic aspects of deep learning and high-dimensional statistics, including:

- Non-parametric density estimation based on Voronoi cells.
- Group-equivariant representation learning, with applications to robotics.
- Theoretical deep learning via algebraic geometry.

Directions for future work:

- Voronoi cells for social networks/graphs.
- Emergence of features, especially in large vision/language models.

Conclusions and Future Work

We have explored various geometric and algebraic aspects of deep learning and high-dimensional statistics, including:

- Non-parametric density estimation based on Voronoi cells.
- Group-equivariant representation learning, with applications to robotics.
- Theoretical deep learning via algebraic geometry.

Directions for future work:

- Voronoi cells for social networks/graphs.
- Emergence of features, especially in large vision/language models.
- A prescriptive theory of neural architectures; what's next after Transformers?

Grazie!

References I

[1] Yoshua Bengio, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives". In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1798–1828.

[2] Michael M Bronstein et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: (2021).

[3] Herbert Edelsbrunner and John L Harer. "Computational topology: an introduction". In: (2022).

[4] Alejandro García-Castellanos et al. "Hypersteiner: Computing heuristic hyperbolic steiner minimal trees". In: *Symposium on Algorithm Engineering and Experiments (ALENEX)* (2025).

[5] Nathan W Henry, Giovanni Luca Marchetti, and Kathlén Kohn. "Geometry of lightning self-attention: Identifiability and dimension". In: *International Conference on Learning Representations (ICLR)*. 2025.

References II

[6] Irina Higgins et al. “Towards a definition of disentangled representations”. In: *arXiv preprint arXiv:1812.02230* (2018).

[7] Joe Kileel, Matthew Trager, and Joan Bruna. “On the expressive power of deep polynomial neural networks”. In: *Advances in neural information processing systems* 32 (2019).

[8] Thomas Kipf, Elise van der Pol, and Max Welling. “Contrastive learning of structured world models”. In: *ICLR* (2020).

[9] Alexander Kravberg et al. “Active Nearest Neighbor Regression Through Delaunay Refinement”. In: *International Conference on Machine Learning (ICML)*. 2022.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012).

References III

[11] Giovanni Luca Marchetti et al. "An efficient and continuous Voronoi density estimator". In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2023.

[12] Giovanni Luca Marchetti et al. "An Invitation to Neuroalgebraic Geometry". In: *arXiv preprint arXiv:2501.18915* (2025).

[13] Giovanni Luca Marchetti et al. "Equivariant representation learning via class-pose decomposition". In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2023.

[14] Giovanni Luca Marchetti et al. "Harmonics of Learning: Universal Fourier Features Emerge in Invariant Networks". In: *Conference on Learning Theory (COLT)* (2024).

[15] JK Ord. "How many trees in a forest". In: *Mathematical Scientist* 3 (1978), pp. 23–33.

References IV

[16] Karl Pearson. "Contributions to the mathematical theory of evolution". In: *Philosophical Transactions of the Royal Society of London. A* 185 (1894), pp. 71–110.

[17] Vladislav Polianskii et al. "Voronoi density estimator for high-dimensional data: Computation, compactification and convergence". In: *Uncertainty in Artificial Intelligence (UAI)*. 2022.

[18] Alfredo Reichlin et al. "Back to the Manifold: Recovering from Out-of-Distribution States". In: *International Conference on Intelligent Robots and Systems (IROS)*. 2022.

[19] Luis Perez Rey et al. "Equivariant Representation Learning in the Presence of Stabilizers". In: *European Conference on Machine Learning (ECML-PKDD)*. 2023.

References V

[20] Murray Rosenblatt. "Remarks on Some Nonparametric Estimates of a Density Function". In: *The Annals of Mathematical Statistics* 27.3 (1956), pp. 832–837. ISSN: 00034851. URL: http://www.jstor.org/stable/2237390.

[21] David W Scott. "Multivariate density estimation: theory, practice, and visualization". In: (2015).

[22] Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. "On the Geometry and Optimization of Polynomial Convolutional Networks". In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2025.

References VI

[23] Joel Zylberberg, Jason Timothy Murphy, and Michael Robert DeWeese. "A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields". In: *PLoS computational biology* 7.10 (2011), e1002250.