LAPLACIAN EIGENSPACES, HOROCYCLES AND NEURON MODELS ON HYPERBOLIC SPACES

Anonymous authors
Paper under double-blind review

ABSTRACT

We use hyperbolic Poisson kernel to construct the horocycle neuron model on hyperbolic spaces, which is a spectral generalization of the classical neuron model. We prove a universal approximation theorem for horocycle neurons. As a corollary, this theorem leads to a state-of-the-art result on the expressivity of neurons of the hyperbolic MLR. Our experiments get state-of-the-art results on the Poincare-embedding tree classification task and the classification accuracy of the two-dimensional visualization of images.

1 INTRODUCTION

Conventional deep network techniques attempt to use architecture based on compositions of simple functions to learn representations of Euclidean data (LeCun et al., 2015). They have achieved remarkable successes in a wide range of applications (Hinton et al., 2012; Silver et al., 2016; He et al., 2016). Geometric deep learning attempts to generalize conventional learning techniques to non-Euclidean spaces (Bronstein et al., 2017; Monti et al., 2017).

There has been growing interest in using hyperbolic spaces in machine learning tasks because they are well-suited for tree-like data representation (Alanis-Lobato et al., 2016; Nickel & Kiela, 2017; Chamberlain et al., 2018; Nickel & Kiela, 2018; Sala et al., 2018; Ganea et al., 2018b; Tifrea et al., 2019; Chami et al., 2019; Liu et al., 2019; Balazevic et al., 2019; Yu & Sa, 2019; Gulcehre et al., 2019; Law et al., 2019). Many authors have introduced hyperbolic analogs of classical learning tools (Ontrup & Ritter, 2005; Ganea et al., 2018a; Cho et al., 2019; Nagano et al., 2019; Grattarola et al., 2019; Mathieu et al., 2019; Ovinnikov, 2020; Khrulkov et al., 2020; Shimizu et al., 2020).

Spectral methods are successful in machine learning, from nonlinear dimensionality reduction (Belkin & Partha, 2002) to clustering (Shi & Malik, 2000; Ng et al., 2002) to hashing (Weiss et al., 2009) to graph CNNs (Bruna et al., 2014) to spherical CNNs (Cohen et al., 2018) and to inference networks (Pfau et al., 2019). This paper study a spectral hyperbolic generalization of the FC (affine) layer.

Let $(\cdot,\cdot)_E$ be the inner product, $|\cdot|$ the Euclidean norm, and ρ an activation function. The Poincaré ball model of the hyperbolic space $\mathbf{H}^n (n{\geq}2)$ is a manifold $\{x{\in}\mathbf{R}^n : |x|{<}1\}$ equipped with a Riemannian metric $ds^2_{\mathbf{H}^n}{=}\sum_{i=1}^n 4(1{-}|x|^2)^{-2}dx_i^2$. The boundary of $\mathbf{H}^n$ under its canonical embedding in $\mathbf{R}^n$ is the unit sphere S^{n-1}. The classical neuron $y{=}\rho((x,w)_E{+}b)$ is of input $x{\in}\mathbf{R}^n$, output $y{\in}\mathbf{R}$, with trainable parameters $w{\in}\mathbf{R}^n, b{\in}\mathbf{R}$. An affine layer $\mathbf{R}^n \to \mathbf{R}^m$ is a concatenation of m neurons. An alternative representation of the neuron $x{\mapsto}\rho((x,w)_E{+}b)$ is given by [1]

$$x{\in}\mathbf{R}^n \mapsto \rho(\lambda(x,\omega)_E{+}b), \quad \omega{\in}S^{n-1}, \lambda, b{\in}\mathbf{R}. \tag{1}$$

This neuron is constant over any hyperplane that is perpendicular to a fixed direction ω. In $\mathbf{H}^n$, a horocycle is a $n{-}1$ dimensional sphere (one point deleted) that is tangential to S^{n-1}. Horocycles are hyperbolic counterparts of hyperplanes (Bonola, 2012). Horocyclic waves $\langle x,\omega\rangle_H := \frac{1}{2}\log\frac{1-|x|^2}{|x-\omega|^2}$ are constant over any horocycle that is tangential to S^{n-1} at ω. Therefore,

$$x{\in}\mathbf{H}^n \mapsto \rho(\lambda\langle x,\omega\rangle_H{+}b), \quad \omega{\in}S^{n-1}, \lambda, b{\in}\mathbf{R} \tag{2}$$

generalizes the classical neuron model (1), and a concatenation of finitely many (2) generalizes the FC (affine) layer. We call (2) a horocycle neuron.

[1] if $w \neq (0,\ldots,0)$, one can take $\omega = w/|w|, \lambda = |w|$; else, one can take $\lambda = 0$ and any $\omega \in S^{n-1}$.

There exists another popular neuron model for functions defined on $\mathbf{H}^n$. Let $p \in \mathbf{H}^n, a \in T_p(\mathbf{H}^n)$, $\oplus$ the Möbius addition (Ungar, 2008). The following functions $f^1_{a,p}(x) = \frac{2|a|}{1-|p|^2} \sinh^{-1}\left(\frac{2(-p\oplus x,a)_E}{(1-|-p\oplus x|^2)|a|}\right)$ are building blocks of many hyperbolic learning tools (Ganea et al., 2018a; Mathieu et al., 2019; Shimizu et al., 2020). Figure 1 displays different neuron models on $\mathbf{H}^2$. The colorbar represents values of neurons, $\omega{=}(1,0), a{=}(1,0), p{=}(0.5,0)$ and ρ is tanh.

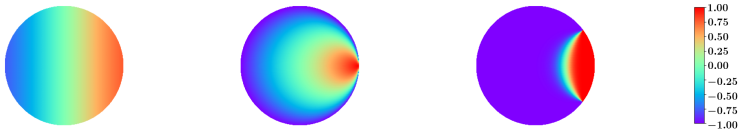

Figure 1: (Left) $\rho((\cdot,\omega)_E)$; (middle) $\rho(\langle\cdot,\omega\rangle_H)$; (right) $\rho(f^1_{a,p}(\cdot))$.

Neuron models (1, 2) are related to spectral theory because $(\cdot,\omega)_E$ (respectively $\langle\cdot,\omega\rangle_H$) are building blocks of the Euclidean (respectively hyperbolic) Laplacian eigenspace. Many L^2 spaces have a basis given by Laplacian eigenfunctions (Kulkarni, 1980; Einsiedler & Ward, 2017). On one side, all Euclidean (respectively hyperbolic) eigenfunctions are some kind of "superposition" of $(\cdot,\omega)_E$ (respectively $\langle\cdot,\omega\rangle_H$). On the other side, neural networks based on (1) (respectively (2)) express functions that are another kind of "superposition" of $(\cdot,\omega)_E$ (respectively $\langle\cdot,\omega\rangle_H$). They heuristically explain why networks constructed by (1, 2) are likely universal approximators. The above heuristic is not proof. By using the Hahn Banach theorem, an injectivity theorem of Helgason and an integral formula, we prove that finite sums of horocycle neurons (2) are universal approximators (Theorem 2).

This article mainly contributes theoretically to hyperbolic learning:

- we first apply spectral methods, horocycle, and Poisson kernel, to hyperbolic deep learning;
- we propose horocycle neuron, Poisson neuron, horocycle MLR, and end-based clustering;
- we prove expressivity results for horocycle neurons and $f^1_{a,p}$.

In experiments, using Tensorflow, Keras, and scikit-learn (Abadi et al., 2015; Chollet et al., 2015; Pedregosa et al., 2011), we obtain state-of-the-art results on the Poincaré-embedding tree classification task and the 2D visualization of images. We obtain 1D (S^1) reduction of image datasets (Figure 2).

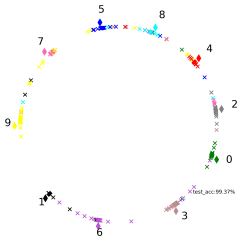

Figure 2: A single feature (angular coordinate) classifies MNIST with test error 0.63%

2 Related work

Universal approximation There is a vast literature on universal approximation (see Cybenko (1989); Hornik et al. (1989); Funahashi (1989); Leshno et al. (1993) and others). Cybenko (1989)'s existential approach uses the Hahn Banach theorem and Fourier transform of Radon measures. To prove Theorem 2, we also use the Hahn Banach theorem, and additionally an integral formula (6) and an injectivity Theorem 1 of Helgason. Generalizing integral formulas and injectivity theorems is easier than generalizing Fourier transform of Radon measures on most non-Euclidean spaces. (Carroll & Dickinson, 1989) used the inverse Radon transform to prove universal approximation theorems. This method is related to ours, as injectivity theorems and generalized inverse Radon transforms are related. However, using the injectivity theorem is an existential approach while using the inverse Radon transform is a constructive one.

Spectral methods Spectral methods in Bronstein et al. (2017); Bruna et al. (2014); Cohen et al. (2018) use a basis of $L^2(X)$ given by eigenfunctions, where X is a finite graph or the sphere. Because $L^2(\mathbf{H}^n)$ has no eigenfunctions as a basis, our work is an addition to the existing literature. This paper is motivated by a connection between the classical neuron model and Euclidean eigenfunctions, particularly by a heuristic connection between the universal approximation theorem and the expressivity of eigenfunctions. We are trying yet unable to locate a reference in the literature.

Hyperbolic deep learning One part of the hyperbolic learning concerns embedding data into the hyperbolic space (Nickel & Kiela, 2017; Sala et al., 2018), and another part concerns learning architectures with hyperbolic data as the input (Ganea et al. (2018a); Cho et al. (2019)). Ganea et al. (2018a) proposed two ways to generalize the affine layer on hyperbolic spaces: one by replacing the linear and bias part of an affine map with (25, 26) of Ganea et al. (2018a); another one by using a concatenation of $f^1_{a,p}$ in their hyperbolic MLR. The latter seems more related to ours. A level set of $f^1_{a,p}$ has the same distance to a chosen geodesic hypersurface, and a level set of a horocycle neuron has the same "spectral distance" to an ideal point at infinity. Lemma 1 shows an interesting relationship between them. Based on functions similar to $f^1_{a,p}$, Mathieu et al. (2019); Shimizu et al. (2020) build the gyroplane layer and Poincaré FC layer. Ganea et al. (2018a); Cho et al. (2019) use geodesics as decision hyperplanes, and we (initially) use horocycles. Geodesic decision hypersurfaces finally appear in the horocycle MLR (due to Apollonian circles). Geodesics decision hyperplanes and geodesic decision hypersurfaces arise from different methods. Khrulkov et al. (2020) investigated the hyperbolic image embedding, where prototypes of each class are center-based. We study a different one, where prototypes are end-based.

3 HYPERBOLIC SPACES

Hyperbolic metric We use the Poincaré model. The hyperbolic space $\mathbf{H}^n (n{\geq}2)$ is the manifold $\{x{\in}\mathbf{R}^n : |x|{<}1\}$ equipped with a Riemannian metric $ds^2 = \sum_{i=1}^n 4(1{-}|x|^2)^{-2}dx_i^2$. Let o be the origin of $\mathbf{H}^n$. The distance function $d_{\mathbf{H}^n}$ satisfies $d_{\mathbf{H}^n}(o,x){=}2\operatorname{arctanh}|x|$.

Geodesics, horocycles and corresponding points Geodesics in $\mathbf{H}^n$ are precisely circular arcs that are orthogonal to S^{n-1}. Horocycles in $\mathbf{H}^n$ are precisely $n{-}1$ dimensional spheres that are tangential to S^{n-1} (Lobachevsky called them horospheres if $n{\geq}3$, and we follow Helgason's notion Helgason (1970)). In Gauss's theory of parallelism Bonola (2012)[p.67], parallel straight lines in $\mathbf{H}^n$ are geodesics ending at the same point in S^{n-1}. Gauss introduced the idea of corresponding points Bonola (2012)[p.73]. A locus of corresponding points of a family of parallel straight lines is a hyperplane in $\mathbf{R}^n$ and a horocycle in $\mathbf{H}^n$. This fact explains that horocycles are hyperbolic analogs of Euclidean hyperplanes. This analogy is a cornerstone of Gauss, Bolyai, and Lobachevsky's non-Euclidean geometry Bonola (2012), and it suggests a useful idea on the construction of decision hypersurfaces.

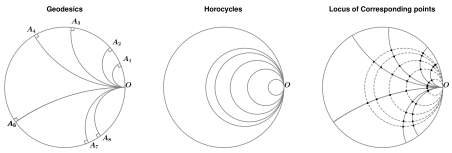

Figure 3: (Left) A_iO are pairwise parallel lines in the sense of Gauss; (Middle) A family of horocycles tangential to O; (Right) Each Horocycle is a locus of corresponding points of parallel lines.

Hyperbolic Poisson kernel The Poisson kernel for $\mathbf{H}^n$ is $P(x,\omega){=}\left(\frac{1-|x|^2}{|x-\omega|^2}\right)^{n-1}$, where $x{\in}\mathbf{H}^n, \omega{\in}S^{n-1}$ (Helgason (1970)[p.108]). The function $\langle\cdot,\omega\rangle_H = \frac{1}{2(n-1)}\log P(\cdot,\omega)$ or

$$\langle x,\omega\rangle_H = \frac{1}{2}\log\frac{1-|x|^2}{|x-\omega|^2} \tag{3}$$

is constant over any horocycle that is tangential to S^{n-1} at ω (Figure 1 (middle), (5)).

Riemannian volume The Riemannian volume induced by the metric ds^2 on $\mathbf{H}^n$ is

$$dm = 2^n(1-|x|^2)^{-n}dx_1\dots dx_n. \tag{4}$$

Horocycles Let Ξ be the set of horocycles of $\mathbf{H}^n$, and let Ξ_ω be the set of all horocycles that are tangential to S^{n-1} at ω. Given $\lambda \in \mathbf{R}$, we let $\xi_{\lambda,\omega}$ be the unique horocycle that connects ω and $\tanh(\lambda/2) \cdot \omega$. We have $\Xi_\omega = \cup_{\lambda \in \mathbf{R}} \{\xi_{\lambda,\omega}\}$ and $\Xi = \cup_{\omega \in S^{n-1}} \Xi_\omega$. The length of any geodesic (that ends at ω) line segment cut by $\xi_{\lambda_1,\omega}$ and $\xi_{\lambda_2,\omega}$ equals $|\lambda_1 - \lambda_2|$ (A.2). Therefore $|\lambda_1 - \lambda_2|$ is a natural distance function defined on Ξ_ω. The map $\lambda \to \xi_{\lambda,\omega}$ is an isometry between $\mathbf{R}$ and Ξ_ω. This isometry is closely related to $\langle \cdot, \omega \rangle_H$ (A.3): for any $x \in \xi_{\lambda,\omega}$,

$$\langle x, \omega \rangle_H = \lambda/2. \tag{5}$$

The annoying $/2$ in (5) is a tradeoff that the metirc here is different from that in Helgason (2000).

Integral formula For fixed $\omega \in S^{n-1}$, $\mathbf{H}^n = \cup_{\lambda \in \mathbf{R}} \xi_{\lambda,\omega}$. Let $dm_{\xi_{\lambda,\omega}}$ be the measure induced by ds^2 on $\xi_{\lambda,\omega}$. Let L be a family of geodesics that end at ω, $\delta > 0$, and $U = L \cap (\cup_{\lambda \le \alpha \le \lambda+\delta} \xi_{\alpha,\omega})$. For $l \in L$, $d_{\mathbf{H}}(l \cap \xi_{\lambda,\omega}, l \cap \xi_{\lambda+\delta,\omega}) = \delta$ (A.2), hence $dm(U) = \delta \cdot dm_{\xi_{\lambda,\omega}}(U \cap \xi_{\lambda,\omega})$ and therefore

$$\int_{\mathbf{H}^n} f(x) dm(x) = \int_{\mathbf{R}} \left(\int_{\xi_{\lambda,\omega}} f(z) dm_{\xi_{\lambda,\omega}}(z) \right) d\lambda. \tag{6}$$

The above proof (for $\mathbf{H}^n$) is essentially the same as that in (Helgason, 2000)[p.37] (for $\mathbf{H}^2$). To further convince the reader that (6) holds for all n, we give another simple proof in A.4.

Injectivity theorem With respect to the canonical measure on Ξ, Helgason (1970)[p.13] proved

Theorem 1 (Helgason)**.** *If $f \in L^1(\mathbf{H}^n)$ and $\int_\xi f(z) dm_\xi(z) = 0$ for a.e $\xi \in \Xi$, then $f = 0$ a.e..*

4 Learning architectures and eigenfunctions of the Laplacian

We first discuss a heuristic connection between the representation properties of eigenfunctions and of classical neurons. Next, we define some learning tools related to the horocycle neuron. Experiments will further demonstrate all of them.

4.1 Eigenspaces and neuron models

On a Riemannian manifold X, the Laplace-Beltrami operator L_X is the divergence of the gradient. The following representational power of Laplacian eigenfunctions is well-known (Kulkarni, 1980; Einsiedler & Ward, 2017): if X is a compact Riemannian manifold, finite graph (with a Laplacian matrix), or bounded domain in $\mathbf{R}^n$, then $L^2(X)$ has a basis given by eigenfunctions. This statement is false if X is $\mathbf{R}^n$ or $\mathbf{H}^n$ (see Borthwick (2007)[Theorem 4.2] for $\mathbf{H}^2$).

Eigenspaces of on $\mathbf{R}^n$ and $\mathbf{H}^n$ Given $\omega \in S^{n-1}$ and a complex number λ, $x \mapsto e^{\lambda (x,\omega)_E}$ is an eigenfunction of $L_{\mathbf{R}^n}$, and $x \mapsto e^{\lambda \langle x,\omega \rangle_H}$ is an eigenfunction of $L_{\mathbf{H}^n}$ (Helgason (1970)[p.140]). Euclidean (respectively hyperbolic) eigenfunctions are obtained from $(x,\omega)_E$ (respectively $\langle x,\omega \rangle_H$) by some kind of superposition. For example, all smooth eigenfunctions of $L_{\mathbf{R}^n}$ are precisely the functions (M. Hashizume & Okamoto, 1972)[p.543]

$$f(x) = \int_{S^{n-1}} e^{\lambda (x,\omega)_E} dT(\omega), \tag{7}$$

and eigenfunctions of $L_{\mathbf{H}^n}$ are precisely the functions (Helgason, 1970)[Theorem 1.7, p.139]

$$f(x) = \int_{S^{n-1}} e^{\lambda \langle x,\omega \rangle_H} dT(\omega), \tag{8}$$

where T in (7) and (8) are some technical linear forms of suitable functional spaces on S^{n-1}.

Neuron models It is clear from (7) and (1) that Euclidean eigenfunctions (respectively classical neurons) are superpositions, with homogeneity and additivity, of $(\cdot,\omega)_E$ and exp (respectively ρ). It is clear from (8) and (2) that hyperbolic eigenfunctions (respectively horocycle neurons) are superpositions of $\langle \cdot,\omega \rangle_H$ and exp (respectively ρ). We accept the representation properties of eigenfunctions on compact manifolds. Therefore, we may expect the same of classical and

horocycle neurons. However, this connection is heuristic. The spectral theory does not directly lead to universal approximation theorems because of the following: 1, superpositions in (1, 2) and (7, 8) are different (similarly, another kind of superposition in Hilbert's 13th problem (Hilbert, 1935; Arnold, 2009) was a driving force for universal approximation theorems (Nielsen, 1987), but only in rare cases, a result of Hilbert's 13th problem applies directly to networks (Girosi & Poggio, 1989)); 2, desired representation properties of hyperbolic eigenfunctions are unknown, partially because $\mathbf{H}^n$ is non-compact; 3, techniques (and theorems) in spectral theory favor Hilbert spaces, while universal approximation theorems cover more than L^2 space.

4.2 Horocycle based learning architectures

Horocycle feature Given a non-origin point $x \in \mathbf{H}^n$, for $y \in \mathbf{H}^n$ we define $h_x(y) = \langle y, x/|x| \rangle_H$. This h_x is the horocycle feature attached to x.

Horocycle decision hypersurface (HDH) Horocycles are hyperbolic analogs of Euclidean hyperplanes. Therefore horocycles could be a possible choice of decision hypersurfaces. A horocycle decision hypersurface (Figure 4[left]) may arise from a level set of a horocycle feature.

Horocycle layer, MLR We call a concatenation of (2) a horocycle layer. We call a horocycle layer followed by a softmax activation (with the cross-entropy loss) a horocycle MLR. In a horocycle MLR, the output of a horocycle layer is the score. In the classical MLR, the decision boundary determined by two score functions, $\{x : \lambda_1(x, \omega_1)_E + b_1 = \lambda_2(x, \omega_2)_E + b_2\}$, is a generic hyperplane. However, the decision boundary determined by $\{x : \lambda_1 \langle x, \omega_1 \rangle_H + b_1 = \lambda_2 \langle x, \omega_2 \rangle_H + b_2\}$ is more complicated. In experiments, we set all λ_i to be 1 (non-trainable). In this case, parameters ω_i are prototypes of each class (end-based), and decision boundaries (Figure 4[right]) are geodesics decision hypersurfaces (**GDH**). To verify the latter fact, we notice $\langle x, \omega_1 \rangle_H + b_1 = \langle x, \omega_2 \rangle_H + b_2$ is equivalent to $|x - \omega_1| = e^{\lambda_1 - \lambda_2} |x - \omega_2|$ and then we apply Apollonian circles. The fact that decision boundaries are geodesic hypersurfaces resembles classical linear regions (Montufar et al., 2014) and some classical case that clusters are separated linearly. As the Apollonius theorem often appears on $\mathbf{R}^2$, we present a proof for $\mathbf{R}^n$ in A.7. We use a standard algorithm (Absil et al., 2008; Bonnabel, 2013) to update ω in the horocycle neuron (A.8).

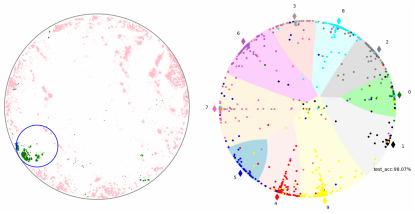

Figure 4: (Left) HDH; (Right) End-based clustering and GDH.

Poisson neuron, layer and MLR We call $x \in \mathbf{R}^n \to \rho\left(\lambda \frac{|w_1 - w_0|^2 - |x - w_0|^2}{|x - w_1|^2 + \epsilon} + b\right)$, with trainable parameters $w_0, w_1 \in \mathbf{R}^n, \lambda, b \in \mathbf{R}$, and ϵ a small constant for numerical stability, a Poisson neuron model. This model is well defined on $\mathbf{R}^n$. We call a concatenation of Poisson neurons a Poisson layer. We call a Poisson layer (Batch normalization may apply) followed by a softmax activation (with the cross-entropy loss) a Poisson MLR. Figure 5 displays a Poisson neuron, where ρ is $\tanh$, $w_0 = (0, 0)$, $w_1 = (1, 0), b = 0, \lambda = -1/3$. Suppose this Poisson neuron is non-trainable and is the score of the class 'bird' in a Poisson MLR, then the Poisson MLR is pulling feature vectors of birds to the deep red area and others to the deep blue. It seems unrealistic if there are more than two classes. Adding the Batch normalization (Ioffe & Szegedy, 2015) after the Poisson layer solves this problem in the experiment, after which bird objects still move to the deep red area, and others move to the blue but not necessarily the deep blue. We use the usual optimization to update parameters in the Poisson neuron.

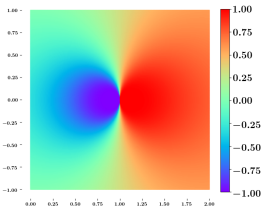

Figure 5: Poisson neuron

End-based clusters, end prototypes Natural clustering is a topic in representation learning (Bengio et al., 2013). The common prototype-based clusters are center-based (Tan et al., 2005). We want to embed objects into $\mathbf{H}^n$ such that each class i is close, in the $\mathbf{R}^n$ topology, to its prototype ω_i on S^{n-1} (Figure 4[right]). These clusters are not center-based (different from Khrulkov et al. (2020)). We call these clusters end-based, and we call ω_i end prototypes (in hyperbolic geometry, the end

is an equivalence class of parallel lines in Figure 3[left]). Different neurons may lead to different prototypes. In the feedforward network classifier, the first part provides a feature representation in $\mathbf{R}^n$ ($\mathbf{H}^n$ here) (Goodfellow et al., 2016)[p.525]. If RBFs calculate scores from this feature space, then the parameters of RBFs are roughly centroids of each class. (LeCun et al., 1998). If horocycle neurons $\langle \cdot, \omega_i \rangle_H + b_i$ do the job, then ω_i (thin diamonds in Figure 4[right]) are end prototypes. End prototypes are not points in $\mathbf{H}^n$ but at infinity. In the $\mathbf{R}^n$ topology, end prototypes are boundary points of each class. The Poisson neuron is useful for more general boundary-based clustering.

5 REPRESENTATIONAL POWER

In this section, ρ is sigmoidal(Cybenko, 1989), ReLU(Nair & Hinton, 2010), ELU(Clevert et al., 2016), or Softplus(Dugas et al., 2001).

Theorem 2. *Let K be a compact set in $\mathbf{H}^n$, and $1 \leq p < \infty$. Then finite sums of the form*

$$F(x) = \sum_{i=1}^{N} \alpha_i \rho(\lambda_i \langle x, \omega_i \rangle + b_i), \quad \omega_i \in S^{n-1}, \alpha_i, \lambda_i, b_i \in \mathbf{R} \tag{9}$$

are dense in $L^p(K, \mu)$, where μ is either dm (4) or the induced Euclidean volume.

Proof. We first treat the case ρ is sigmoidal and $\mu = dm$. Assume that these finite sums are not dense in $L^p(K, dm)$. By the Hahn-Banach theorem, there exists some nonzero $h \in L^q(K, dm)$, where $q = p/(p-1)$ if $p > 1$ and $q = \infty$ if $p = 1$, such that $\int_K F(x) h(x) dm(x) = 0$ for all finite sums of the form (9). As K is a compact set, by Hölder's inequality, $\int_K |h(x)| \, dm \leq (\int_K dm)^{1/p} ||h||_{L^q(K,dm)}$, which leads to $h \in L^1(K, dm)$. Extend h to be a function H that is defined on $\mathbf{H}^n$ by assigning $H(x) = h(x)$ if $x \in K$ and $H(x) = 0$ if $x \in \mathbf{H}^n \backslash K$. Then $H \in L^1(\mathbf{H}^n, dm) \cap L^q(\mathbf{H}^n, dm)$ and

$$\int_{\mathbf{H}^n} F(x) H(x) dm(x) = 0 \tag{10}$$

for all finite sums of the form (9). For any $\omega \in S^{n-1}$ and $\lambda, b \in \mathbf{R}$, we set $F_{\omega,\lambda,b}(x) = \rho(\lambda(\langle x, \omega \rangle_H - b))$. These functions are uniformly bounded, as $|F_{\omega,\lambda,b}(x)| \leq 1$. Moreover,

$$\lim_{\lambda \to \infty} F_{\omega,\lambda,b}(x) = \begin{cases} 1 & if \quad \langle x, \omega \rangle_H > b, \\ 0 & if \quad \langle x, \omega \rangle_H < b. \end{cases}$$

According to (10), for all ω, λ, b, we have $\int_{\mathbf{H}^n} F_{\omega,\lambda,b}(x) H(x) dm(x) = 0$. Functions $\{F_{\omega,\lambda,b}\}_{\lambda \in \mathbf{R}}$ converge pointwise as $\lambda \to \infty$, and they are uniformly bounded by $|H| \in L^1(\mathbf{H}^n, dm)$. By the bounded convergence theorem, for all $\omega \in S^{n-1}, b \in \mathbf{R}$, we have

$$\int_{\{x : \langle x, \omega \rangle_H > b\}} H(x) dm(x) = 0. \tag{11}$$

By the integral formula (6) (with notations defined there), (5) and (11), for all $b \in \mathbf{R}$,

$$\int_{2b}^{\infty} \left(\int_{\xi_t} H(z) dm_t(z) \right) dt = 0. \tag{12}$$

Taking the derivative of $\int_{2b}^{\infty} \left(\int_{\xi_t} H(z) dm_t(z) \right) dt$ with respect to b, we deduce from (12) that $\int_{\xi_{2b}} H(z) dm_{2b}(z) = 0$ for a.e. $b \in \mathbf{R}$. In other words, the integration of H on a.e. $\xi \in \Xi_\omega$ is zero. This fact is valid for all $\omega \in S^{n-1}$. Therefore, the integration of H on a.e. $\xi \in \Xi$ is zero. By the injectivity Theorem 1, $H = 0$ a.e., which contradicts our assumption. Therefore, finite sums of the form (9) are dense in $L^p(K, dm)$. The case ρ is ReLU, ELU or Softplus and $\mu = dm$ follows from the above case and the fact that $x \mapsto \rho(x+1) - \rho(x)$ is sigmoidal. The case μ is the Euclidean volume follows from previous cases and the fact that the Euclidean volume on compact K is bounded from above by λdm for some constant λ. □

In A.10, we prove Lemma 1 (see code for visualization).

Lemma 1. *Let K be a compact set in $\mathbf{H}^n$, $\omega \in S^{n-1}$ and $\epsilon > 0$. There exists $c, d \in \mathbf{R}$ $p \in \mathbf{H}^n$ and $a \in T_p(\mathbf{H}^n)$ such that the function $D(x) = cf^1_{a,p}(x) + d - \langle x, \omega \rangle_H$ satisfies $||D||_{L^p(K,dm)} < \epsilon$.*

By this lemma, Theorem 2 leads to a state-of-the-art result on the expressivity of $f^1_{a,p}$.

Corollary 1. *Let K be a compact set in $\mathbf{H}^n$ and $1 \leq p < \infty$. Finite sums of the form $F(x) = \sum_{i=1}^{N} \alpha_i \rho(c_i f^1_{a_i,p_i}(x) + d_i)$, $p_i \in H^n, a_i \in T_{p_i}(H^n), \alpha_i, c_i, d_i \in \mathbf{R}$, are dense in $L^p(K, \mu)$, where $\mu = dm$ or μ is the induced Euclidean volume.*

6 Experiments

6.1 Poincaré embedding subtree classification

The task (Ganea et al., 2018a) is to classify all other nodes as being part of the subtree rooted at a node in the Poincaré embedding (Nickel & Kiela, 2017). We use logistic regression with a horocycle feature as the predictor. Figure 4 (left) displays the decision hypersurface, where the red point is $x =$group.n.01, green points are from the subtree rooted at x, and pink points are the rest. In Table 1, [G], [S] and [H] refers to Ganea et al. (2018a), Shimizu et al. (2020) and this paper. Our model has $O(1)$ parameters, while MLR models in [G] and [S] have $O(\text{dim})$. The table demonstrates that our model obtains the state-of-the-art results on this task. A.11 and code contain details.

Table 1: Average test F1 classification scores (%) for five subtrees of WordNet noun tree

Wordnet subtree	**Reference**	$\mathbf{H}^2$	$\mathbf{H}^3$	$\mathbf{H}^5$	$\mathbf{H}^{10}$
animal.n.01	[G]/[S]	47.43/60.69	**91.92**/67.88	98.07/86.26	**99.26**/99.15
4016 nodes	[H]	**94.77**	87.00	**98.54**	98.84
group.n.01	[G]/[S]	81.72/74.27	89.87/63.90	87.89/84.36	91.91/85.60
8376 nodes	[H]	**90.32**	**91.87**	**97.64**	**97.63**
location.n.01	[S]	42.60	66.70	78.18	92.34
3362 nodes	[H]	**93.12**	**94.41**	**95.35**	**97.29**
mammal.n.01	[G]/[S]	32.01/63.48	87.54/94.98	87.73/**99.30**	91.37/**99.17**
1181 nodes	[H]	**98.42**	**96.33**	93.03	95.93
worker.n.01	[G]	12.68	24.09	55.46	66.83
1115 nodes	[H]	**94.17**	**91.98**	**95.59**	**94.73**

6.2 End-based clustering for 2D dimension reduction

Deep networks are useful for dimension reduction and visualization of high-dimensional data (Hinton & Salakhutdinov, 2006). We use a deep network with a horocycle MLR to find end-based clusters in $\mathbf{H}^2$. A.15 and code contain details. The first part of the network has four convolutional blocks

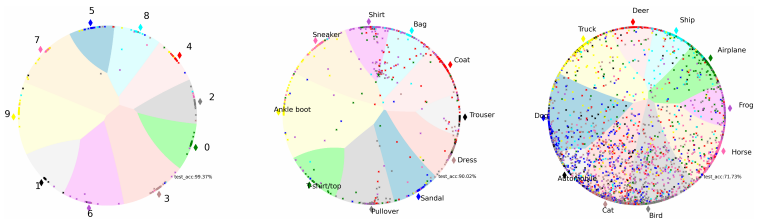

Figure 6: End-based clusters of MNIST, Fashion-MNIST and CIFAR-10

with 2D output, the second part is a logarithm layer that maps $\mathbf{R}^2$ to $\mathbf{H}^2$, and the third part is a horocycle MLR. In Figure 6, we obtain 2D clustering with test error $\mathbf{0.63}\%$ for MNIST, $\mathbf{9.98}\%$ for Fashion-MNIST(Xiao et al., 2017), and $\mathbf{28.27}\%$ for CIFAR-10(Krizhevsky, 2012). Shallow areas are decision regions. Parameters ω_i in the horocycle layer are end prototypes (thin diamonds in figures). We take numbers from Yang et al. (2018); Ghosh & Kirby (2020) as the benchmark for Table 2. Our accuracy for Fashion-MNIST is 8% higher than all numbers presented in McInnes et al. (2020).

Table 2: Test error rates on 2D embedded MNIST by dimensionality reduction techniques

GerDA	Centroidencoder	dG-MCML	dt-MCML	CPL	GCPL	This paper
3.2%	2.61%	2.13%	2.03%	0.72%	0.67%	**0.63**%

6.3 MNIST

MNIST (LeCun et al., 1998) task is popular for testing hyperbolic learning tools (Ontrup & Ritter, 2005; Nagano et al., 2019; Mathieu et al., 2019; Grattarola et al., 2019; Ovinnikov, 2020; Khrulkov et al., 2020). We train two different classifiers. A.12, A.14 and code contain details. The first one uses

Table 3: Test error rates of hyperbolic geometry related MNIST classifiers

Ontrup & Ritter (2005)	Grattarola et al. (2019)	Khrulkov et al. (2020)	This paper
5.4%	4.2%	1%	**0.35**%

a horocycle layer with outputs as features. The input layer is given by scaled 40-dimensional PCA features, the first layer is a 40 input/1000 output horocycle layer, and the second layer is the usual 1000 inputs/10 outputs MLR. The average test error rates after 600 epochs is $\mathbf{1.96}\%$. Theorem 2 is the rationale for this experiment (A.13). The second one uses a Poisson layer with outputs as scores. The first part of this network has four convolutional blocks. The second part is a 1000 inputs/10 outputs Poisson MLR. It is the best hyperbolic geometry related MNIST classifier (Table 3).

6.4 1D(S^1) DIMENSION REDUCTION

Figure 6 suggests that the horocycle MLR is pulling objects to their prototypes on S^{n-1} and therefore is adaptable for S^{n-1} embedding. We train the same model in Section 6.2 and replace the logarithm layer by a normalizing layer in the end. Figure 2, Figure 7, and code contain details.

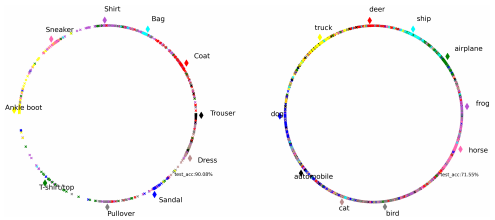

Figure 7: A single feature classifies Fashion-MNIST (CIFAR-10) with test error $9.92\%(28.45\%)$.

6.5 FASHION-MNIST, CIFAR-10 AND FLOWERS

We demonstrate the stability of Poisson MLR with several experiments.

Using four convolutional blocks and a Poisson MLR, we get classifiers with test error rates $\mathbf{5.88}\%$ on Fashion-MNIST and $\mathbf{17.31}\%$ on CIFAR-10.

The classification task of flowers (Tensorflow) is a typical example of overfitting. Replacing the MLR part of the Keras model (Tensorflow) by a Poisson MLR, the new Poisson model shows better generalization performance (Figure 8). A.16 and code contain details.

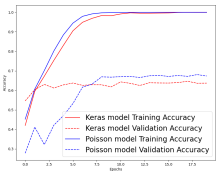

Figure 8: Flowers

7 CONCLUSION

We introduced spectral methods, such as the horocycle and Poisson kernel, for hyperbolic learning. With spectral methods, we obtain state-of-the-art results on the expressivity of $f^1_{a,p}$, the Poincaré embedding subtree classification task, and the classification accuracy of 2D visualization of images.

REFERENCES

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `http://tensorflow.org/`. Software available from tensorflow.org.

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

G. Alanis-Lobato, G. Mier, and M.A. Andrade-Navarro. Efficient embedding of complex networks to hyperbolic space via their Laplacian. *Scientific Reports*, 6, 2016.

V. I. Arnold. On functions of three variables. In *Collected Works. Vladimir I.Arnold-Collected Works*. Springer, 2009.

I. Balazevic, C. Allen, and T. Hospedales. Multi-relational Poincaré graph embeddings. In *Advances in Neural Information Processing Systems 32*, pp. 4463–4473. Curran Associates, Inc., 2019.

M. Belkin and N. Partha. Laplacian eigenmaps and spectral techniques for embedding and clustering. In T. G. Dietterich, S. Becker, and Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems 14*, pp. 585–591. MIT Press, 2002.

Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828, August 2013. ISSN 0162-8828.

S. Bonnabel. Stochastic gradient descent on Riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.

R. Bonola. *Non-Euclidean Geometry*. Dover Books on Mathematics. Dover Publications, 2012. ISBN 9780486155036.

D. Borthwick. *Spectral Theory of Infinite-Area Hyperbolic Surfaces*, volume 256. 01 2007. ISBN 978-0-8176-4524-3. doi: 10.1007/978-0-8176-4653-0.

M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.

S. Carroll and B. Dickinson. Construction of neural nets using the Radon transform. *International 1989 Joint Conference on Neural Networks*, pp. 607–611 vol.1, 1989.

B.P. Chamberlain, J.R. Clough, and M.P. Deisenroth. Hybed: Hyperbolic neural graph embedding, 2018.

I. Chami, Z. Ying, C. Ré, and J. Leskovec. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems 32*, pp. 4868–4879. Curran Associates, Inc., 2019.

H. Cho, B. DeMeo, J. Peng, and B. Berger. Large-margin classification in hyperbolic space. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1832–1840. PMLR, 16–18 Apr 2019.

F. Chollet et al. Keras. `https://keras.io`, 2015.

D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). In *Proceedings of the International Conference on Learning Representations(ICLR 2016)*, 2016.

T. S. Cohen, M. Geiger, J. Köhler, and M. Welling. Spherical CNNs. In *International Conference on Learning Representations*, 2018.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314, 1989.

C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating second-order functional knowledge for better option pricing. In T. K. Leen, T. G. Dietterich, and V. Tresp (eds.), *Advances in Neural Information Processing Systems 13*, pp. 472–478. MIT Press, 2001.

M. Einsiedler and T. Ward. *Functional Analysis, Spectral Theory, and Applications*, volume 276 of *Graduate Texts in Mathematics*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-58539-0. doi: 10.1007/978-3-319-58540-6.

K.-I. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

O.-E. Ganea, G. Becigneul, and T. Hofmann. Hyperbolic neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 31*, pp. 5345–5355. Curran Associates, Inc., 2018a.

O.-E. Ganea, G. Becigneul, and T. Hofmann. Hyperbolic entailment cones for learning hierarchical embeddings. volume 80 of *Proceedings of Machine Learning Research*, pp. 1646–1655. PMLR, 10–15 Jul 2018b.

T. Ghosh and M. Kirby. Supervised dimensionality reduction and visualization using centroid-encoder, 2020.

F. Girosi and T. Poggio. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Computation - NECO*, 1:465–469, 12 1989.

I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT Press, 2016.

D. Grattarola, L. Livi, and C. Alippi. Adversarial autoencoders with constant-curvature latent manifolds. *Appl. Soft Comput.*, 81, 2019.

C. Gulcehre, M. Denil, M. Malinowski, A. Razavi, R. Pascanu, K. M. Hermann, P. Battaglia, V. Bapst, D. Raposo, A. Santoro, and N. de Freitas. Hyperbolic attention networks. In *International Conference on Learning Representations*, 2019.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

S. Helgason. A duality for symmetric spaces with applications to group representations. *Adv. Math.*, 5:1–154, 1970.

S. Helgason. *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions*. Mathematical surveys and monographs. American Mathematical Society, 2000. ISBN 9780821826737.

D. Hilbert. Mathematische probleme. In *Dritter Band: Analysis · Grundlagen der Mathematik · Physik Verschiedenes*. Springer, 1935.

G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006.

G.E. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pp. 448–456. JMLR.org, 2015.

V. Khrulkov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempitsky. Hyperbolic image embeddings. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6417–6427, 2020.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

A. Krizhevsky. Learning multiple layers of features from tiny images. *University of Toronto*, 05 2012.

R.S. Kulkarni. Review: Sigurdur helgason, differential geometry, Lie groups and symmetric spaces. *Bull. Amer. Math. Soc. (N.S.)*, 2(3):468–476, 05 1980.

M. Law, R. Liao, J. Snell, and R. Zemel. Lorentzian distance learning for hyperbolic representations. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3672–3681, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015. doi: 10.1038/nature14539.

M. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

Q. Liu, M. Nickel, and D. Kiela. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems 32*, pp. 8230–8241. Curran Associates, Inc., 2019.

K. Minemura M. Hashizume, A. Kowata and K. Okamoto. An integral representation of an eigenfunction of the Laplacian on the Euclidean space. *Hiroshima Math. J.*, 2:535–545, 1972.

E. Mathieu, C. Le Lan, C. J. Maddison, R. Tomioka, and Y.W. Teh. Continuous hierarchical representations with Poincaré variational auto-encoders. In *Advances in Neural Information Processing Systems 32*, pp. 12565–12576. Curran Associates, Inc., 2019.

L. McInnes, J. Healy, and J. Melville. UMAP: Uniform manifold approximation and projection for dimension reduction, 2020.

F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5425–5434, 2017.

Guido F Montufar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio. On the number of linear regions of deep neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 27*, pp. 2924–2932. Curran Associates, Inc., 2014.

Y. Nagano, S. Yamaguchi, Y. Fujita, and M. Koyama. A wrapped normal distribution on hyperbolic space for gradient-based learning. In *ICML*, 2019.

V. Nair and G.E. Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pp. 807–814, 2010.

A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, and Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems 14*, pp. 849–856. MIT Press, 2002.

M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 30*, pp. 6338–6347. Curran Associates, Inc., 2017.

M. Nickel and D. Kiela. Learning continuous hierarchies in the Lorentz model of hyperbolic geometry. In *ICML*, 2018.

R.H. Nielsen. Kolmogorov's mapping neural network existence theorem. In *Proceedings of the IEEE First International Conference on Neural Networks* (San Diego, CA), volume III, pp. 11–13, 1987.

J. Ontrup and H. Ritter. A hierarchically growing hyperbolic self-organizing map for rapid structuring of large data sets. 2005.

I. Ovinnikov. Poincaré wasserstein autoencoder, 2020.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

D. Pfau, S. Petersen, A. Agarwal, D. G. T. Barrett, and K. L. Stachenfeld. Spectral inference networks: Unifying deep and spectral learning. In *International Conference on Learning Representations*, 2019.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

F. Sala, C. De Sa, A. Gu, and C. Re. Representation tradeoffs for hyperbolic embeddings. volume 80 of *Proceedings of Machine Learning Research*, pp. 4460–4469. PMLR, 10–15 Jul 2018.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

R. Shimizu, Y. Mukuta, and T. Harada. Hyperbolic neural networks++, 2020.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. P. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, May 2005.

Tensorflow. Image classification. `https://www.tensorflow.org/tutorials/images/classification`.

A. Tifrea, G. Becigneul, and O.-E. Ganea. Poincare glove: Hyperbolic word embeddings. In *International Conference on Learning Representations*, 2019.

A. A. Ungar. A gyrovector space approach to hyperbolic geometry. *Synthesis Lectures on Mathematics and Statistics*, 1(1):1–194, 2008.

Y. Weiss, A. Torralba, and R. Fergus. Spectral hashing. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (eds.), *Advances in Neural Information Processing Systems 21*, pp. 1753–1760. Curran Associates, Inc., 2009.

Jr. William H. Hayt and John A. Buck. *Engineering Electromagnetics*. McGraw-Hill Education, 6 edition, 2001.

H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms, 2017.

H. Yang, Xu-Yao Zhang, F. Yin, and C. Liu. Robust classification with convolutional prototype learning. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3474–3482, 2018.

T. Yu and C. De Sa. Numerically accurate hyperbolic embeddings using tiling-based models. In *Advances in Neural Information Processing Systems 32*, pp. 2023–2033. Curran Associates, Inc., 2019.