ICML'19, Jun 12th, 2019

A Wrapped Normal Distribution on Hyperbolic Space for Gradient Based Learning

Yoshihiro Nagano[1], Shoichiro Yamaguchi[2], Yasuhiro Fujita[2], Masanori Koyama[2]

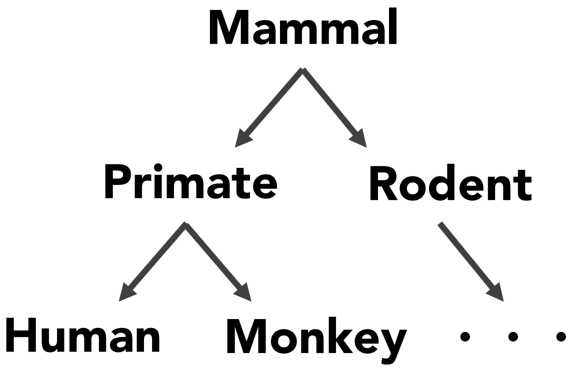
[1] Department of Complexity Science, The University of Tokyo, Japan

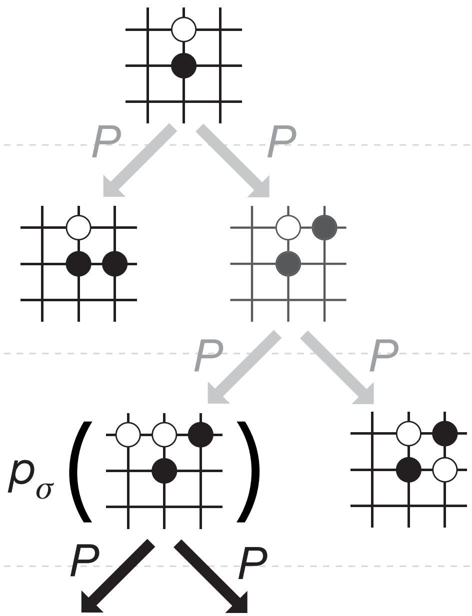
[2] Preferred Networks, Inc., Japan

Code: github.com/pfnet-research/hyperbolic_wrapped_distribution

Poster: 6:30-9:00 PM @Pacific Ballroom #7

Motivation

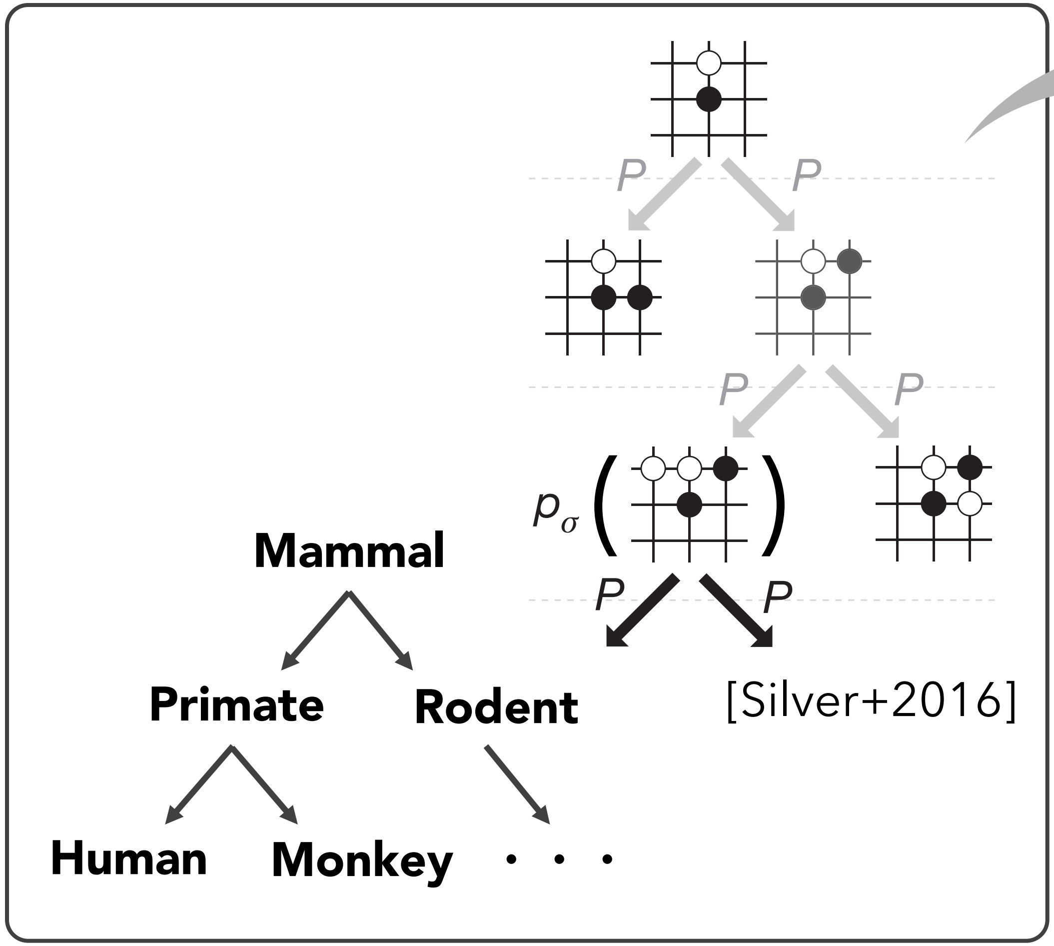




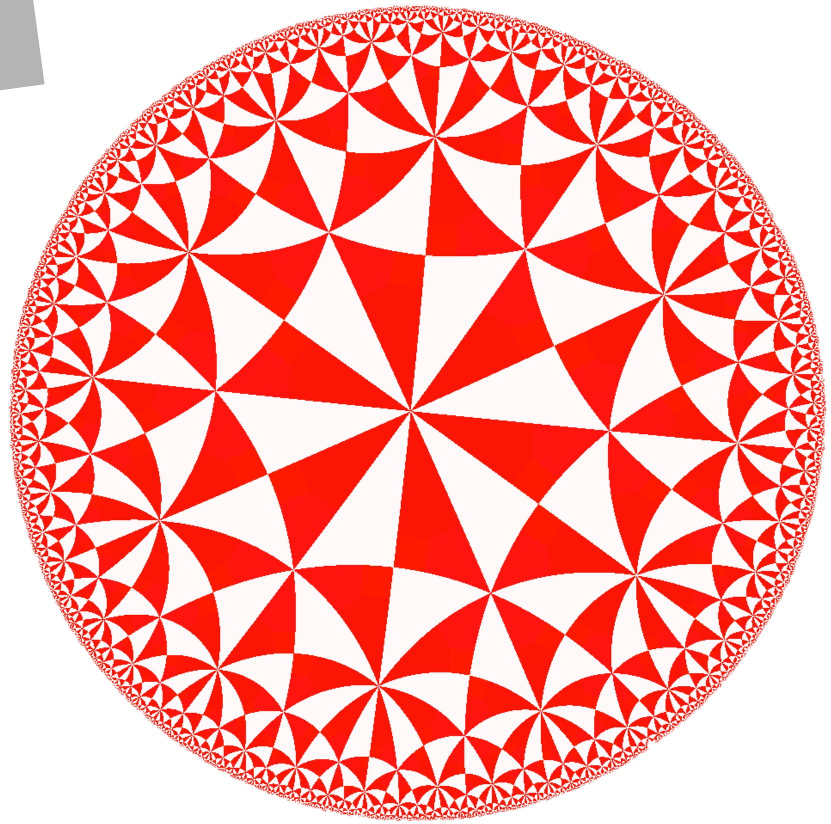
[Silver+2016]

Motivation

Hierarchical Datasets



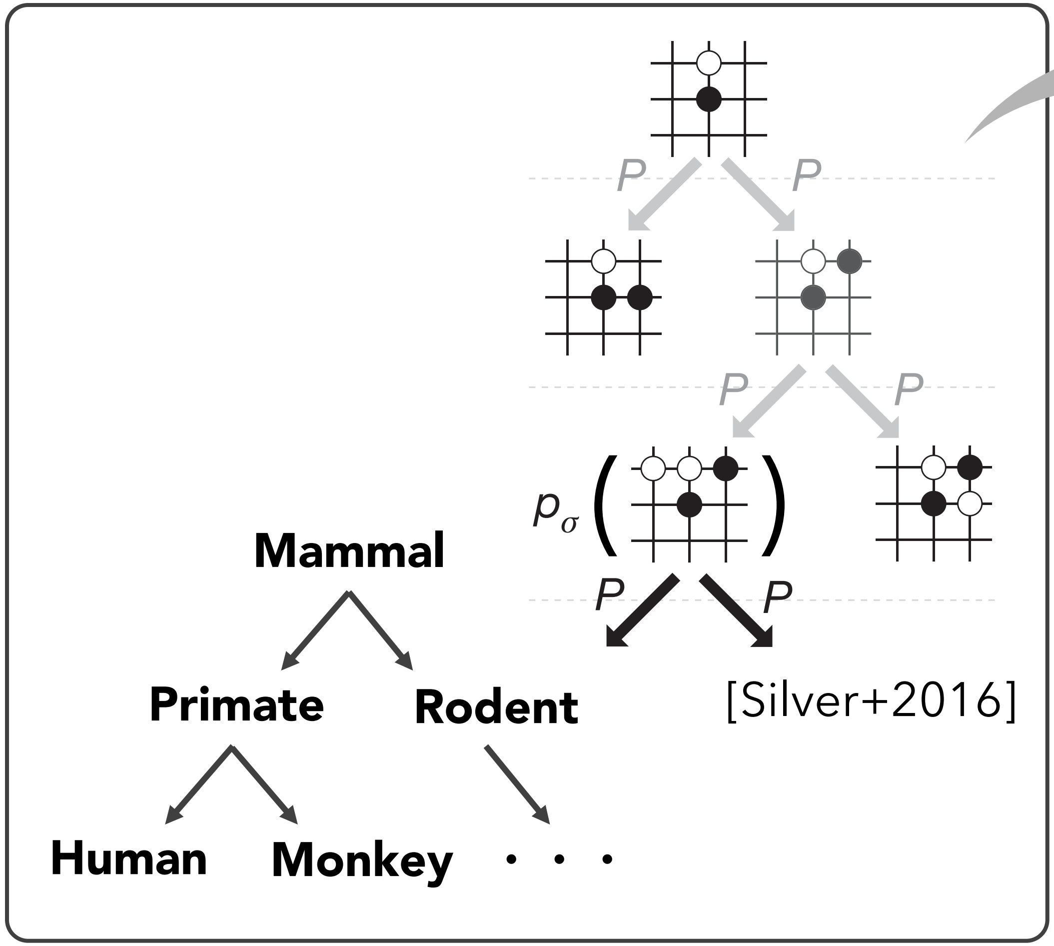
Hyperbolic Space



[Image: wikipedia.org]

Motivation

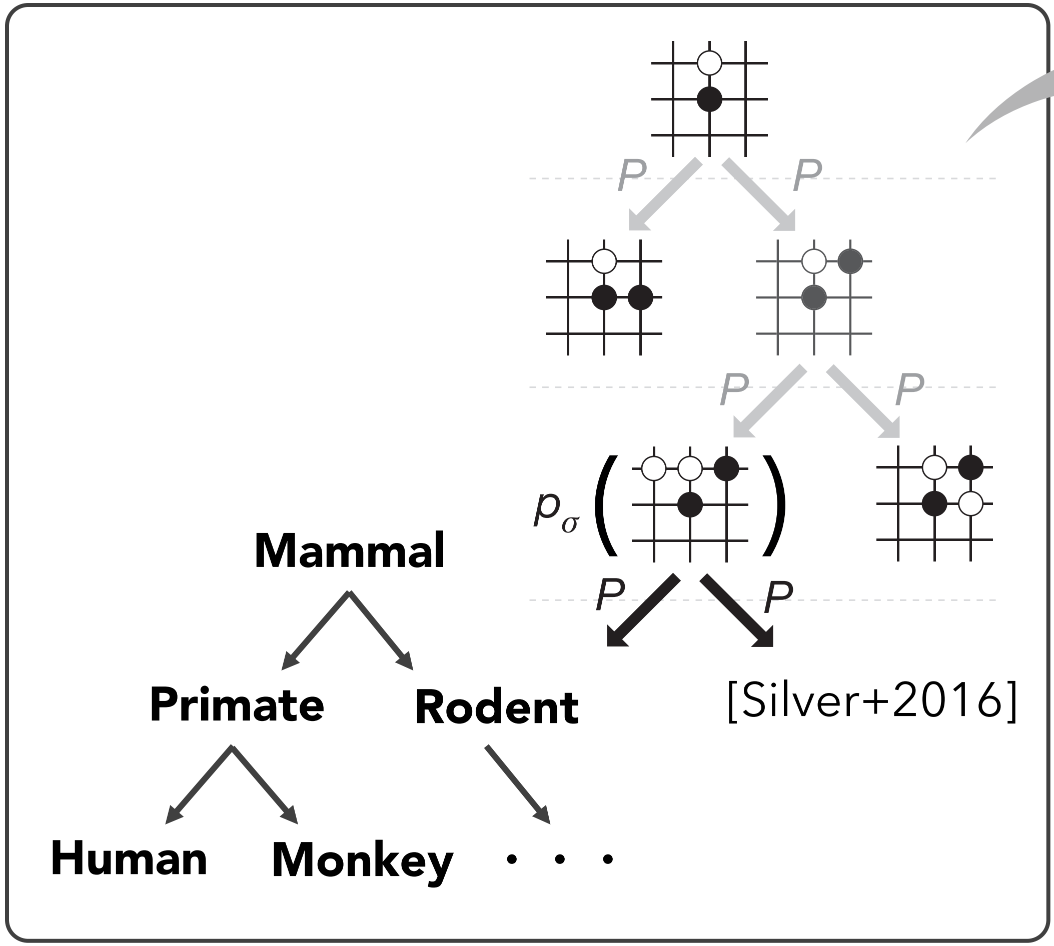
Hierarchical Datasets



Hyperbolic Space

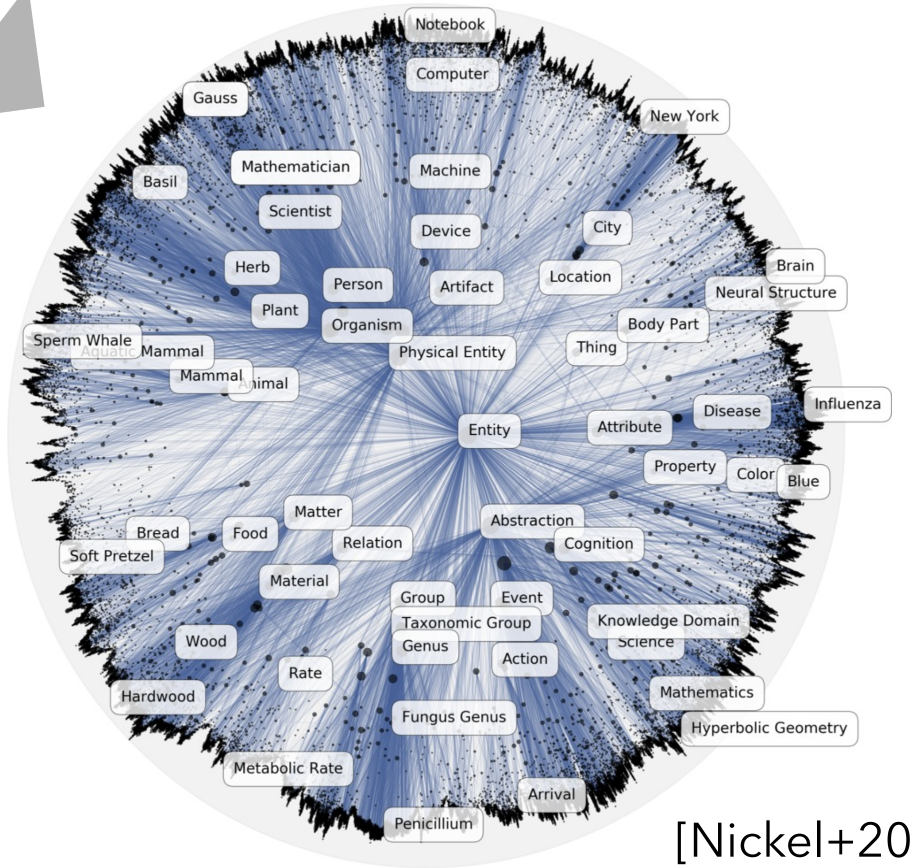
Motivation

Hierarchical Datasets



[Silver+2016]

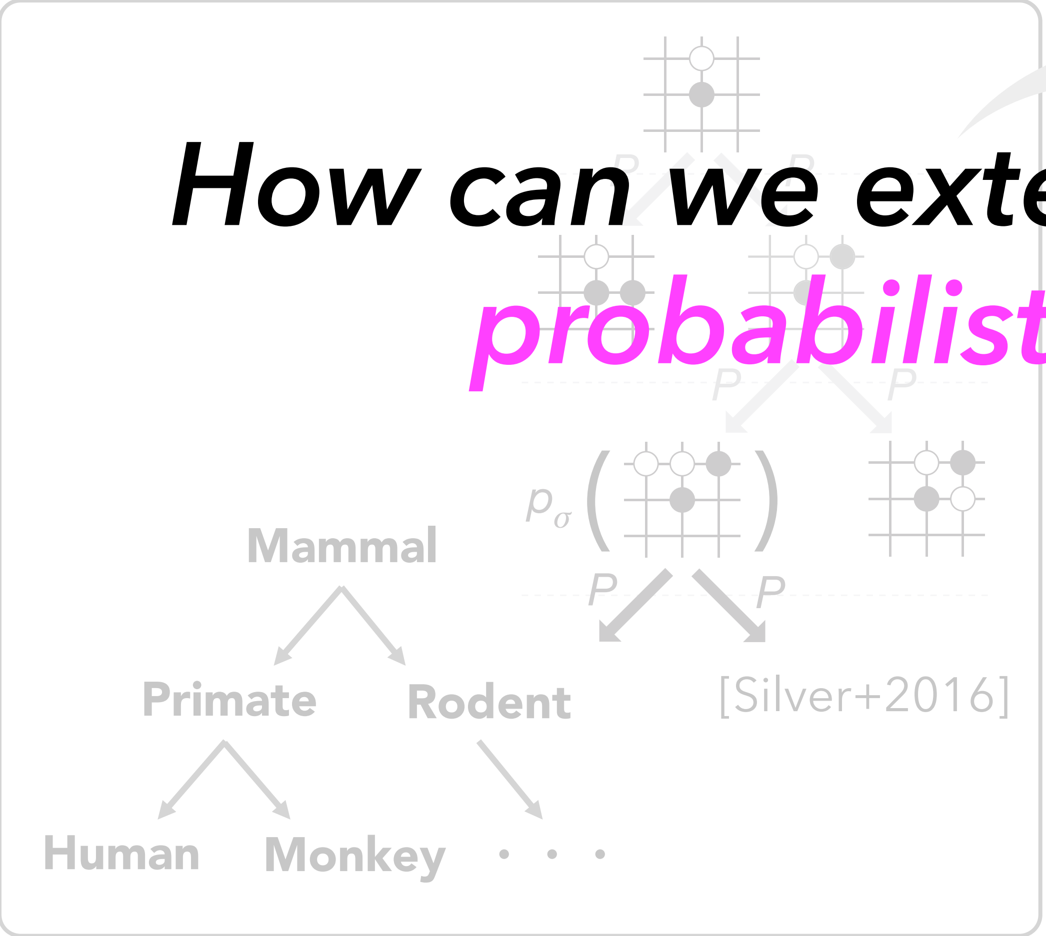
Hyperbolic Space



[Nickel+2017]

Motivation

Hierarchical Datasets



[Silver+2016]

Hyperbolic Space

[Nickel+2017]

How can we extend these works to probabilistic inference?

Difficulty: Probabilistic Distribution on Curved Space

VAEs w/ Riemannian distribution [Ovinnikov2019; Mathieu+2019]

$$\mathcal{N}_{\mathbb{B}_c^d}\left(\boldsymbol{x}|\boldsymbol{\mu},\sigma^2\right) = \frac{1}{Z(\sigma)} \exp\left(-\frac{d_p^c(\boldsymbol{\mu},\boldsymbol{x})^2}{2\sigma^2}\right)$$

$$Z_r(\sigma) = \sqrt{\frac{\pi}{2}}\sigma \frac{1}{(2\sqrt{c})^{d-1}} \sum_{k=0}^{d-1} (-1)^k \binom{d-1}{k} e^{\frac{(d-1-2k)^2}{2} c\sigma^2} \left[1 + \mathrm{erf}\left(\frac{(d-1-2k)\sqrt{c}\sigma}{\sqrt{2}}\right)\right]$$

- Only limited to the Gaussian w/ scalar variance
- Needs rejection sampling

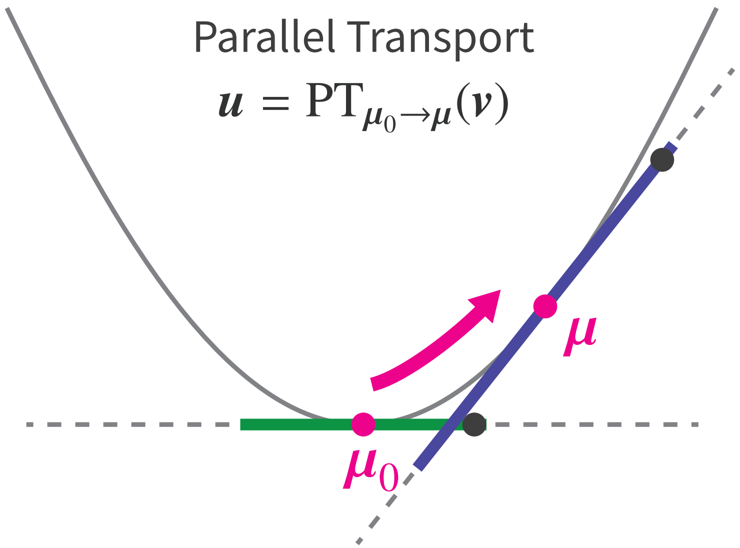
⇒ Construct distribution by sampling for flexible density and sampling

Construction of Hyperbolic Wrapped Distribution

Lorentz model: $\mathbb{H}^n = \left\{ z \in \mathbb{R}^{n+1} : \langle z, z \rangle_{\mathcal{L}} = -1 \right\}$

Defining probabilistic distribution on locally flat **tangent space** and projecting its random variable with the parallel transport and exponential map.

We can analytically get the log-density by calculating volumetric change.

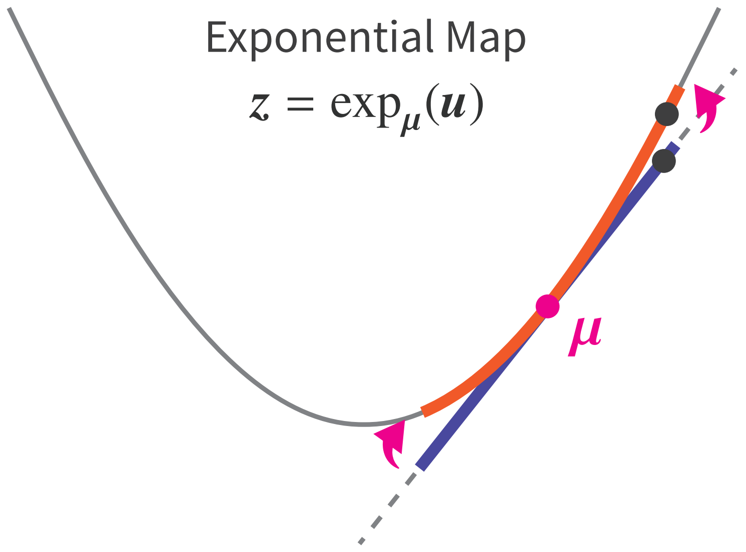


Construction of Hyperbolic Wrapped Distribution

Lorentz model: $\mathbb{H}^n = \left\{ z \in \mathbb{R}^{n+1} : \langle z, z \rangle_{\mathcal{L}} = -1 \right\}$

Defining probabilistic distribution on locally flat **tangent space** and projecting its random variable with the parallel transport and exponential map.

We can analytically get the log-density by calculating volumetric change.

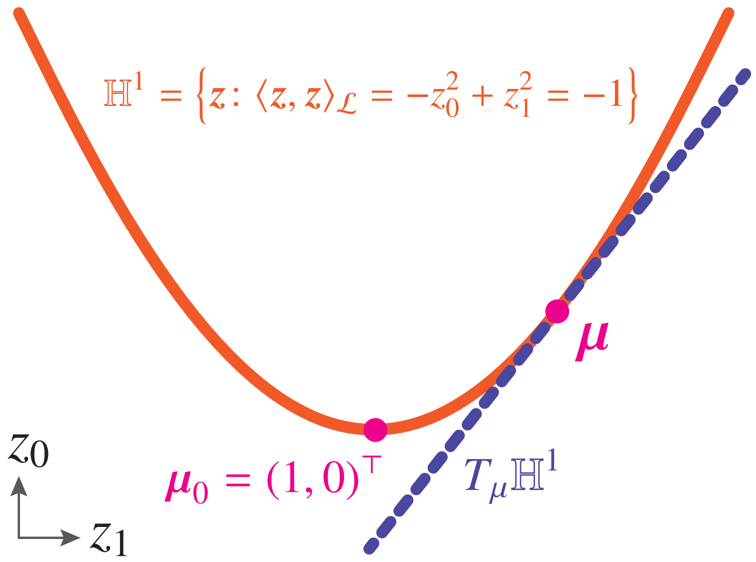


Construction of Hyperbolic Wrapped Distribution

Lorentz model: $\mathbb{H}^n = \left\{ z \in \mathbb{R}^{n+1} : \langle z, z \rangle_{\mathcal{L}} = -1 \right\}$

Defining probabilistic distribution on locally flat **tangent space** and projecting its random variable with the parallel transport and exponential map.

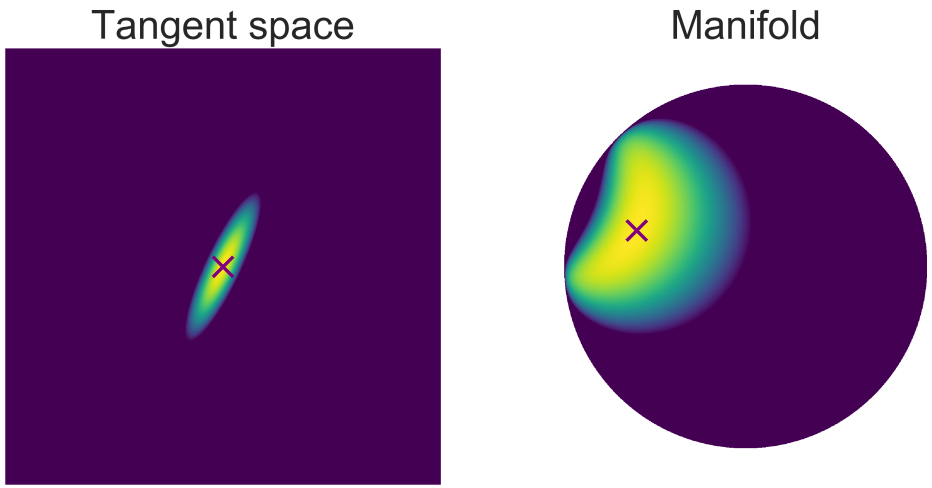
We can analytically get the log-density by calculating volumetric change.



Properties of Hyperbolic Wrapped Distribution

Density: $$\mathcal{G}(z;\mu,\Sigma) = \left(\frac{r}{\sinh r}\right)^{n-1} \mathcal{N}(v;\mathbf{0},\Sigma)$$

Projection: $$z = \exp_{\mu} \circ \mathrm{PT}_{\mu_0 \to \mu}(v)$$

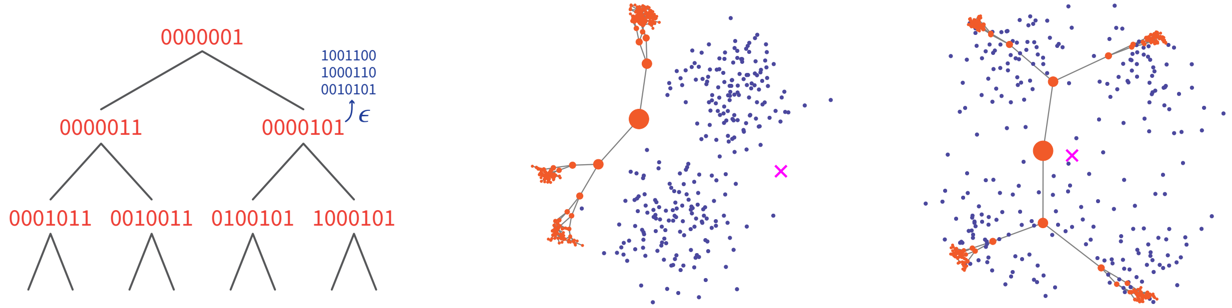


	Riemannian distribution	**Wrapped distribution (Ours)**
Construction	Maximum entropy on the manifold	Projecting r.v. defined on the tangent space
Sampling	Rejection sampling	Conventional sampling w/ deterministic transformation
Expressivity	Gaussian distribution with scalar variance	Any distribution on the tangent space ($\simeq \mathbb{R}^n$)

Numerical Evaluations

Variational Autoencoder

Hyperbolic VAE could learn not only the true hierarchical structure but also noisy unseen data without any explicit knowledge for tree.



(a) A tree representation of the training dataset

(b) Normal VAE ($\beta = 1.0$)

(c) Hyperbolic VAE

Word embedding

Our model outperformed Euclidean counterpart for WordNet nouns dataset.

	Euclid		Hyperbolic	
n	MAP	Rank	MAP	Rank
5	$0.296_{\pm.006}$	$25.09_{\pm.80}$	$\mathbf{0.506}_{\pm.017}$	$\mathbf{20.55}_{\pm1.34}$
10	$0.778_{\pm.007}$	$\mathbf{4.70}_{\pm.05}$	$\mathbf{0.795}_{\pm.007}$	$5.07_{\pm.12}$
20	$0.894_{\pm.002}$	$\mathbf{2.23}_{\pm.03}$	$\mathbf{0.897}_{\pm.005}$	$2.54_{\pm.20}$
50	$0.942_{\pm.003}$	$1.51_{\pm.04}$	$\mathbf{0.975}_{\pm.001}$	$\mathbf{1.19}_{\pm.01}$
100	$0.953_{\pm.002}$	$1.34_{\pm.02}$	$\mathbf{0.978}_{\pm.002}$	$\mathbf{1.15}_{\pm.01}$

Conclusion

Proposed a projection-based probabilistic distribution on hyperbolic space which is easy to use with gradient-based learning.

Constructed the wrapped normal distribution on Lorentz model by projecting the random variable on locally flat tangent space.

Numerically evaluated the performance of our model on various datasets including MNIST, Atari 2600 Breakout, and WordNet.

Poster: 6:30-9:00 PM @Pacific Ballroom #7