A Wrapped Normal Distribution on Hyperbolic Space for Gradient Based Learning

Yoshihiro Nagano¹, Shoichiro Yamaguchi², Yasuhiro Fujita², Masanori Koyama²

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

Mammal

Primate  Rodent

Human  Monkey  ⋮

[Silver+2016]
Motivation

Hierarchical Datasets

Hyperbolic Space

[Image: wikipedia.org]
Motivation

Hierarchical Datasets

Hyperbolic Space

Volume increases exponentially with its radius

Mammal

Primate

Rodent

Human

Monkey

...
Motivation

Hierarchical Datasets

Hyperbolic Space

[Silver+2016]

[Nickel+2017]
Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?

Motivation

Hierarchical Datasets

Hyperbolic Space

How can we extend these works to probabilistic inference?
Difficulty: Probabilistic Distribution on Curved Space

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

\[ N_{B^c}^d (x | \mu, \sigma^2) = \frac{1}{Z(\sigma)} \exp \left( - \frac{d_p^c(\mu, x)^2}{2\sigma^2} \right) \]

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

- Only limited to the Gaussian w/ scalar variance
- Needs rejection sampling
  \[ \Rightarrow \text{Construct distribution by sampling for flexible density and sampling} \]
Construction of Hyperbolic Wrapped Distribution

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

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: \( \mathbb{H}^n = \{ z \in \mathbb{R}^{n+1} : \langle z, z \rangle_L = -1 \} \)

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:** \( \mathbb{H}^n = \left\{ z \in \mathbb{R}^{n+1} : \langle z, z \rangle_L = -1 \right\} \)

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

**Projection:** \( z = \exp_\mu \circ \text{PT}_{\mu_0 \rightarrow \mu}(v) \)

<table>
<thead>
<tr>
<th>Riemannian distribution</th>
<th>Wrapped distribution (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction</strong></td>
<td>Maximum entropy on the manifold</td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
<td>Rejection sampling</td>
</tr>
<tr>
<td><strong>Expressivity</strong></td>
<td>Gaussian distribution with scalar variance</td>
</tr>
</tbody>
</table>
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.

Word embedding

Our model outperformed Euclidean counterpart for WordNet nouns dataset.
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