Lorentzian Distance Learning for Hyperbolic Representations

Marc Law, **Renjie Liao**, Jake Snell, Richard Zemel

June 13, 2019

University of Toronto, Vector Institute

Introduction: Hyperbolic Representations

Manifolds with constant curvature:

- zero curvature: Euclidean space
- positive curvature $1/r^2$: Hypersphere of radius r
- negative curvature $-1/\beta$: Hyperboloid model $\mathcal{H}^{d,\beta}$

Manifolds with constant curvature:

- zero curvature: Euclidean space
- positive curvature $1/r^2$: Hypersphere of radius r
- negative curvature $-1/\beta$: Hyperboloid model $\mathcal{H}^{d,\beta}$

$$\mathcal{H}^{d,\beta} := \{\mathbf{a} = (a_0, \cdots, a_d) \in \mathbb{R}^{d+1} : \langle \mathbf{a}, \mathbf{a} \rangle_{\mathcal{L}} = -\beta, a_0 > 0\} \quad (1)$$

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}} := -a_0 b_0 + \sum_{i=1}^{d} a_i b_i \quad (2)$$

Manifolds with constant curvature:

- zero curvature: Euclidean space
- positive curvature $1/r^2$: Hypersphere of radius r
- negative curvature $-1/\beta$: Hyperboloid model $\mathcal{H}^{d,\beta}$

$$\mathcal{H}^{d,\beta} := \{\mathbf{a} = (a_0, \cdots, a_d) \in \mathbb{R}^{d+1} : \langle \mathbf{a}, \mathbf{a} \rangle_{\mathcal{L}} = -\beta, a_0 > 0\} \quad (1)$$

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}} := -a_0 b_0 + \sum_{i=1}^{d} a_i b_i \quad (2)$$

• Any finite tree can be mapped into a finite hyperbolic space while approximately preserving distances between nodes (Gromov, 1987).

Introduction: Hyperbolic Distances

Poincaré distance: defined for $\beta = 1$

$$\forall \mathbf{a} \in \mathcal{H}^{d,1}, \mathbf{b} \in \mathcal{H}^{d,1} \quad \mathrm{d}_{\mathcal{P}}(\mathbf{a}, \mathbf{b}) = \cosh^{-1}\left(-\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}}\right) \tag{3}$$

Squared Lorentzian distance: defined and smooth for any $\beta > 0$

$$\forall \mathbf{a} \in \mathcal{H}^{d,\beta}, \mathbf{b} \in \mathcal{H}^{d,\beta} \quad \mathrm{d}^2_{\mathcal{L}}(\mathbf{a}, \mathbf{b}) = -2\beta - 2\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}} \tag{4}$$

Introduction: Hyperbolic Distances

Poincaré distance: defined for $\beta = 1$

$$\forall \mathbf{a} \in \mathcal{H}^{d,1}, \mathbf{b} \in \mathcal{H}^{d,1} \quad \mathrm{d}_{\mathcal{P}}(\mathbf{a}, \mathbf{b}) = \cosh^{-1}\left(-\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}}\right) \qquad (3)$$

Squared Lorentzian distance: defined and smooth for any $\beta > 0$

$$\forall \mathbf{a} \in \mathcal{H}^{d,\beta}, \mathbf{b} \in \mathcal{H}^{d,\beta} \quad \mathrm{d}^2_{\mathcal{L}}(\mathbf{a}, \mathbf{b}) = -2\beta - 2\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}} \qquad (4)$$

Advantages:

- Easy to optimize with standard gradient descent
- Closed-form expression for the center of mass
- Preserved order of Euclidean norms between Poincaré ball and hyperboloid
- The Euclidean norm of the centroid decreases as $\beta > 0$ decreases: ideal to represent hierarchies

Center of mass

Theorem (Centroid of the squared Lorentzian distance)

The point $\boldsymbol{\mu} \in \mathcal{H}^{d,\beta}$ *that minimizes the problem*

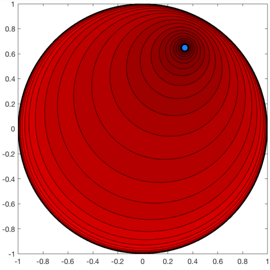
$$\min_{\boldsymbol{\mu} \in \mathcal{H}^{d,\beta}} \sum_{i=1}^{n} \nu_i \mathrm{d}_{\mathcal{L}}^2(\mathbf{x}_i, \boldsymbol{\mu}) \tag{5}$$

where $\forall i, \mathbf{x}_i \in \mathcal{H}^{d,\beta}$, $\nu_i \geq 0, \sum_i \nu_i > 0$ *is formulated as:*

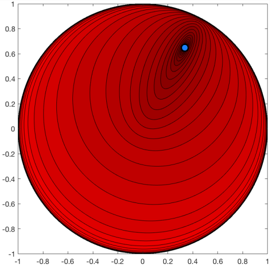
$$\boldsymbol{\mu} = \sqrt{\beta} \frac{\sum_{i=1}^{n} \nu_i \mathbf{x}_i}{|\|\sum_{i=1}^{n} \nu_i \mathbf{x}_i\|_{\mathcal{L}}|} \tag{6}$$

where $|\|\mathbf{a}\|_{\mathcal{L}}| = \sqrt{|\|\mathbf{a}\|_{\mathcal{L}}^2|}$ *is the modulus of the imaginary Lorentzian norm of the positive time-like vector* $\mathbf{a}$.

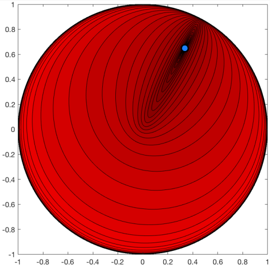
Distance as a function of the curvature $-1/\beta$



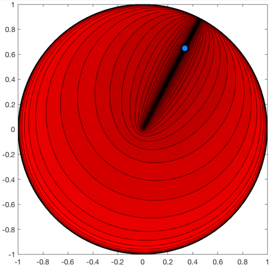
$\beta = 1$



$\beta = 10^{-1}$

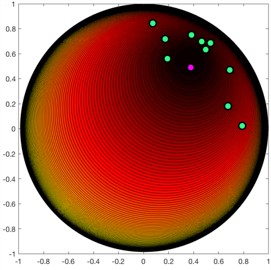


$\beta = 10^{-2}$

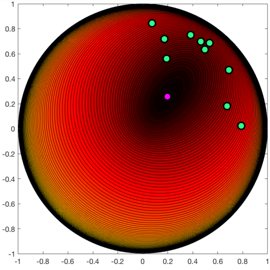


$\beta = 10^{-4}$

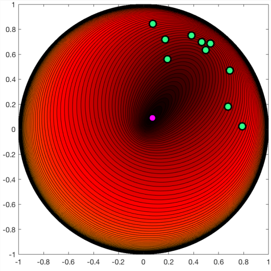
Centroid as a function of the curvature $-1/\beta$



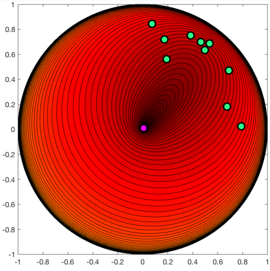
$\beta = 1$



$\beta = 10^{-1}$



$\beta = 10^{-2}$



$\beta = 10^{-4}$

Retrieval Evaluation performance

Method		$d_{\mathcal{P}}$ in $\mathcal{P}^d$	$d_{\mathcal{P}}$ in $\mathcal{H}^d$	Ours $\beta = 0.01$	Ours $\beta = 0.1$	Ours $\beta = 1$
WordNet Nouns	MR	4.02	2.95	**1.46**	1.59	1.72
	MAP	86.5	92.8	**94.0**	93.5	91.5
WordNet Verbs	MR	1.35	1.23	**1.11**	1.14	1.23
	MAP	91.2	93.5	**94.6**	93.7	91.9
EuroVoc	MR	1.23	1.17	**1.06**	**1.06**	1.09
	MAP	94.4	**96.5**	**96.5**	96.0	95.0
ACM	MR	1.71	1.63	**1.03**	1.06	1.16
	MAP	94.8	97.0	**98.8**	96.9	94.1
MeSH	MR	12.8	12.4	1.31	**1.30**	1.40
	MAP	79.4	79.9	90.1	**90.5**	85.5

MR = Mean Rank
MAP = Mean Average Precision

Smaller values of β improve recognition performance

Binary Classification Evaluation performance

Test F1 scores of the Wordnet Nouns subtree:

Dataset	animal.n.01	group.n.01	worker.n.01	mammal.n.01
(Ganea et al., 2018)	$99.26 \pm 0.59\%$	$91.91 \pm 3.07\%$	$66.83 \pm 11.83\%$	$91.37 \pm 6.09\%$
Euclidean dist	$99.36 \pm 0.18\%$	$91.38 \pm 1.19\%$	$47.29 \pm 3.93\%$	$77.76 \pm 5.08\%$
$\log_0$ + Eucl	$98.27 \pm 0.70\%$	$91.41 \pm 0.18\%$	$36.66 \pm 2.74\%$	$56.11 \pm 2.21\%$
Ours ($\beta = 0.01$)	$99.77 \pm 0.17\%$	$99.86 \pm 0.03\%$	$96.32 \pm 1.05\%$	$97.73 \pm 0.86\%$

Conclusion

- We show that the Euclidean norm of the center of mass decreases as the curvature decreases
- The performance of the learned model can be improved by decreasing the curvature of the hyperboloid model
- Decreasing the curvature implicitly enforces high-level nodes to have smaller Euclidean norm than their descendants

Thank You!

Welcome to our poster # 30