The Numerical Stability of Hyperbolic Representation Learning

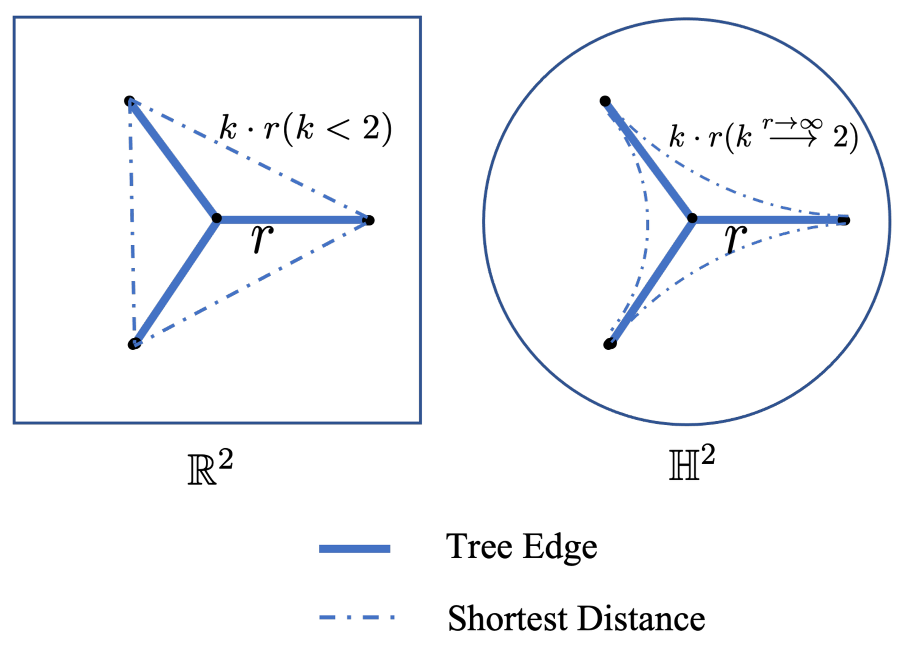
Gal Mishne[1], Zhengchao Wan[1], Yusu Wang[1], and Sheng Yang[2]

[1]University of California San Diego [2]Harvard University

Introduction

- Distances grow exponentially in hyperbolic spaces $\mathbb{H}^n$
- Finite trees can be embeded into $\mathbb{H}^n$ with arbitrarily low distortion[1]
- Motivates hyperbolic embeddings for hierarchical datasets such as words and images
- **Problem**: highly numerically unstable; mysteriously better performance of Lorentz than Poincaré despite being isometric.

Hyperbolic Distances are more Tree-like



Visualization of Poincaré and Lorentz model

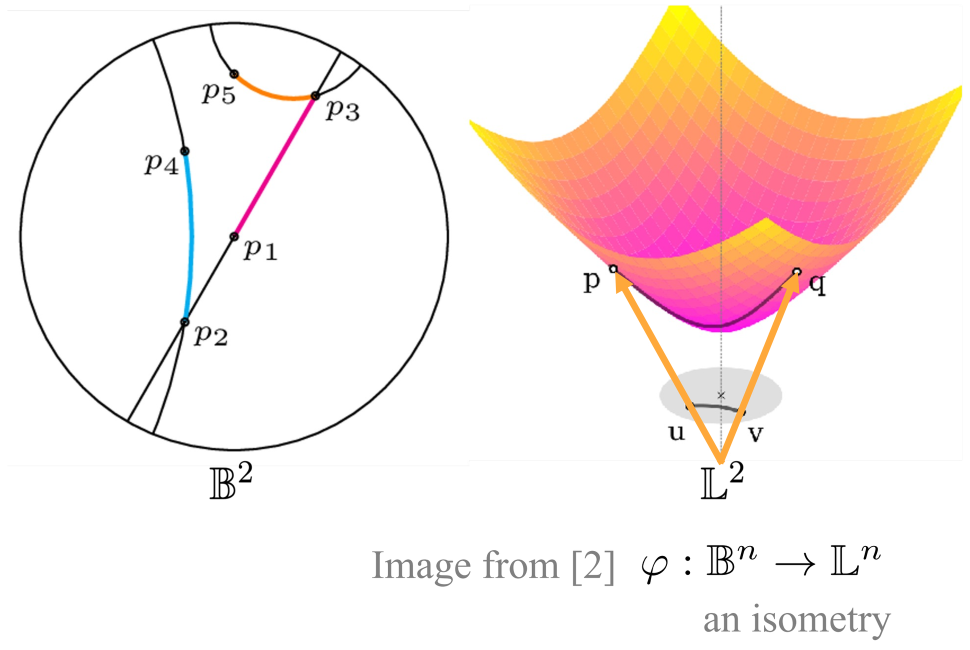
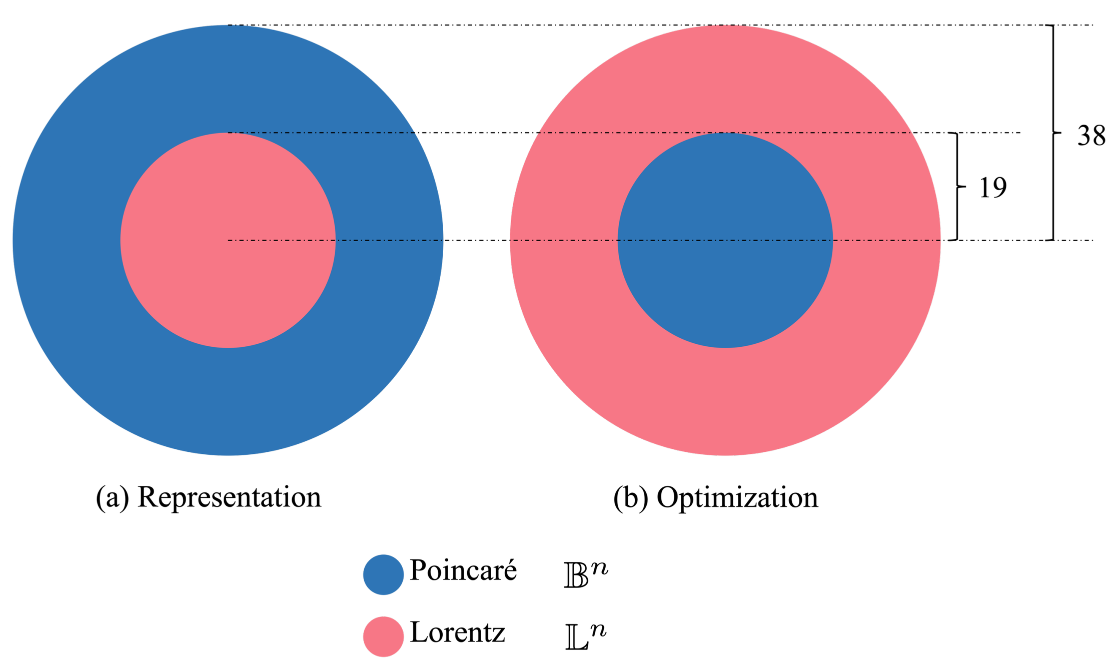


Image from [2] $\varphi : \mathbb{B}^n \to \mathbb{L}^n$ an isometry

Poincaré vs Lorentz

Regions of Accurate Representation/Optimization



- Identify the source of numerical instability in two popular isometric hyperbolic models, Lorentz $\mathbb{L}^n$ and Poincaré $\mathbb{B}^n$
- Clarify comparative advantage of $\mathbb{L}^n$ and $\mathbb{B}^n$
  - In **representation**, $\mathbb{B}^n$ has a larger representation diameter: Poincaré can represent points accuately in a ball with radius of 38 whereas Lorentz has a radius of 19.
  - In **optimization**, $\mathbb{L}^n$ suffers less severe gradient vanishing problem: *for any small* δ *and* $x \in \mathbb{B}^n : ||x|| = 1 - \delta$, *we then have*

$$\|\nabla_{\mathbb{B}^n} f(x)\| = \Omega\left(\delta^2 \|\nabla f(x)\|\right),$$
$$\|\nabla_{\mathbb{L}^n} g(\tilde{x})\| = O\left(\|\nabla f(x)\|\right).$$

$$\begin{tikzcd}
\mathbb{L}^n \arrow[r, shift left, "\psi"] \arrow[dr, "g"'] & \mathbb{B}^n \arrow[l, shift left, "\varphi"] \arrow[d, "f"] \\
& \mathbb{R}
\end{tikzcd}$$

Euclidean Parametrization

- The exponential map is a natural parameterization of hyperbolic model
  - No representation limitation
  - Optimization performance is similar to the Lorentz model

$$\|\nabla h(z)\| = \Omega(\delta \|\nabla f(x)\|)$$

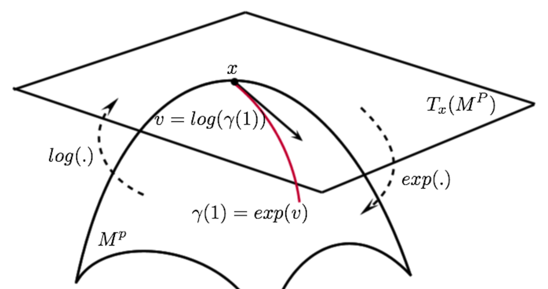


Image from [4]: exp/log map

- Propose a new hyperbolic SVM method
  - Lorentz SVM from [3] (LSVM)

$$\min_{w \in \mathbb{R}^{n+1}} \frac{1}{2}\|w\|^2_{\mathbb{L}^n} + C \sum_{i=1}^{n} l_{\mathbb{L}^n}(-y_i[w, x_i])$$
$$s.t.\ \ [w, w] > 0$$

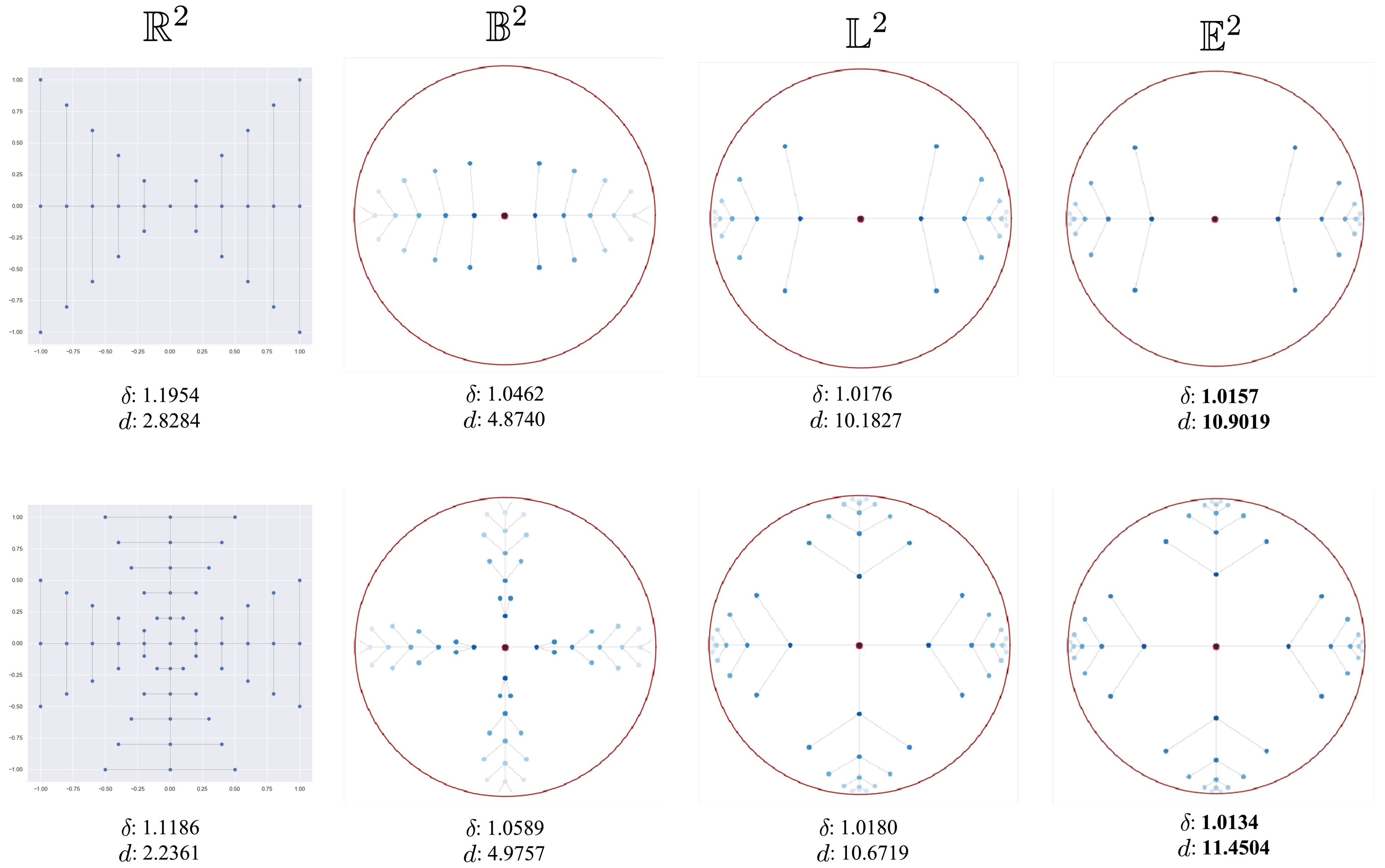
$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "F_D"] \arrow[dr, "F_L"'] \arrow[ddrr, bend right, "h"'] & \mathbb{B}^n \arrow[dr, "f"] & \\
& \mathbb{L}^n \arrow[u, "\psi"] \arrow[r, "g"] & \mathbb{R}
\end{tikzcd}$$

- Proposed hyperbolic SVM (LSVMPP)

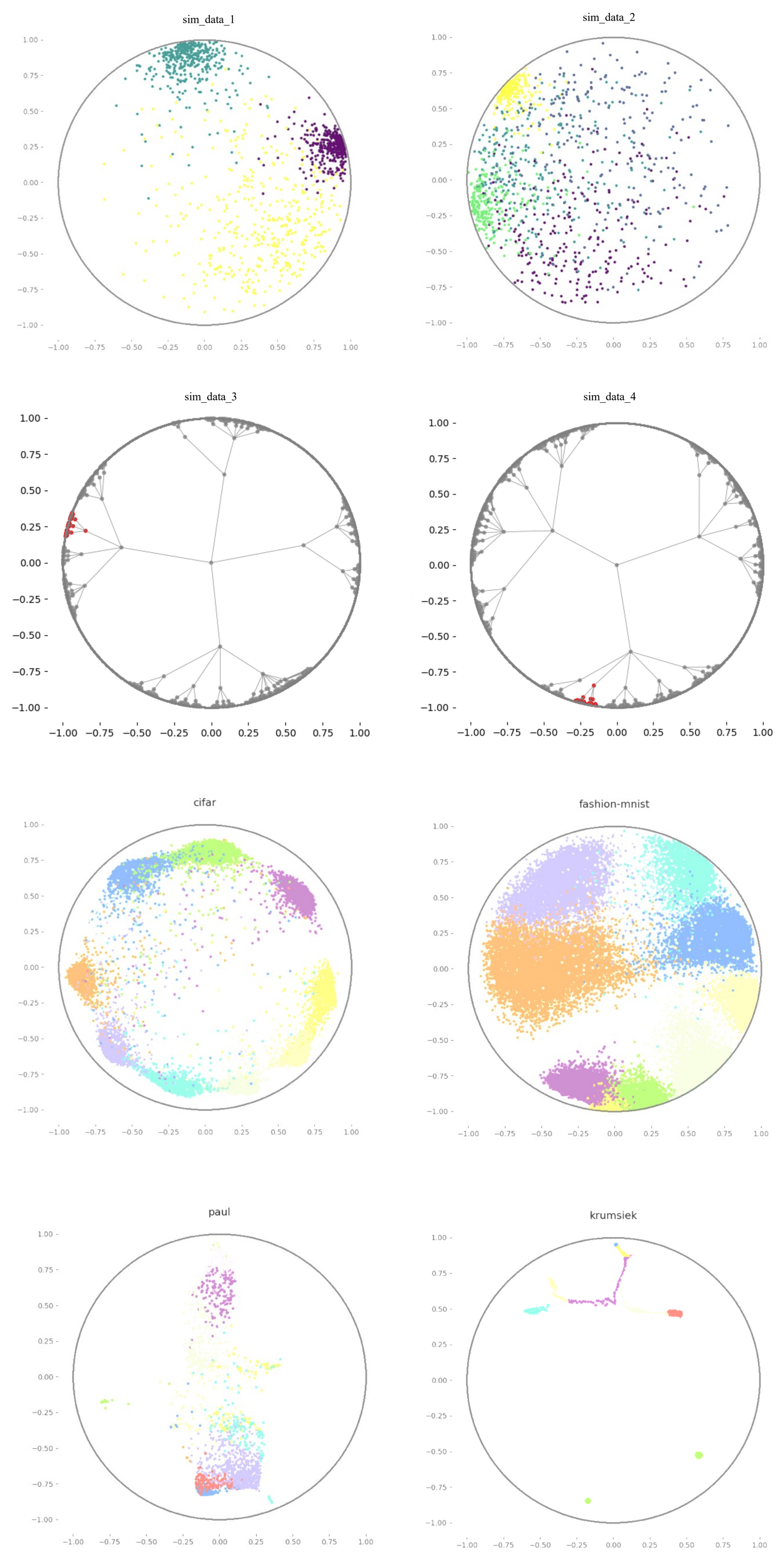
$$\min_{z \in \mathbb{R}^n, a \in \mathbb{R}} \frac{1}{2}\|z\|^2 + C \sum_{i=1}^{n} l_{\mathbb{L}^n}\left(y_i(\sinh(a)\,\|z\|\,x_0 - \cosh(a) z x_r^{\mathrm{T}})\right)$$

Results

- **Tree Embeddings**: reparametrized model has the lowest distortion and largest diameter.



- **SVM**: parametrized model boosts genearlization performance in classification.



	algorithm	accuracy (%)	F1
sim 1	ESVM	94.67 ± 0.00	0.9465 ± 0.00
	LSVM	94.67 ± 0.00	0.9465 ± 0.00
	PSVM	91.66 ± 0.00	0.9151 ± 0.00
	LSVMPP	94.67 ± 0.00	0.9464 ± 0.00
sim 2	ESVM	68.67 ± 0.00	0.6656 ± 0.00
	LSVM	66.66 ± 0.00	0.6336 ± 0.00
	PSVM	64.66 ± 0.00	0.6143 ± 0.00
	LSVMPP	$\mathbf{70.33} \pm \mathbf{0.00}$	$\mathbf{0.6920} \pm \mathbf{0.00}$
sim 3	ESVM	93.40 ± 0.00	0.0000 ± 0.00
	LSVM	96.04 ± 3.24	0.4000 ± 0.49
	PSVM	99.53 ± 0.00	0.9630 ± 0.00
	LSVMPP	$\mathbf{100.00} \pm \mathbf{0.00}$	$\mathbf{1.0000} \pm \mathbf{0.00}$
sim 4	ESVM	97.64 ± 0.00	0.7805 ± 0.00
	LSVM	99.63 ± 0.13	0.9711 ± 0.01
	PSVM	97.91 ± 0.00	0.8095 ± 0.00
	LSVMPP	$\mathbf{99.63} \pm \mathbf{0.13}$	$\mathbf{0.9714} \pm \mathbf{0.01}$

	algorithm	accuracy (%)	F1
CIFAR-10	ESVM	91.88 ± 0.00	0.9191 ± 0.00
	LSVM	91.88 ± 0.00	0.9189 ± 0.00
	PSVM	91.81 ± 0.00	0.9182 ± 0.00
	LSVMPP	$\mathbf{91.96} \pm \mathbf{0.00}$	$\mathbf{0.9197} \pm \mathbf{0.00}$
fashion-MNIST	ESVM	86.37 ± 0.00	0.8665 ± 0.00
	LSVM	71.59 ± 0.07	0.6588 ± 0.08
	PSVM	86.57 ± 0.00	0.8665 ± 0.00
	LSVMPP	$\mathbf{89.49} \pm \mathbf{0.00}$	$\mathbf{0.8955} \pm \mathbf{0.00}$
paul	ESVM	55.05 ± 0.00	0.4073 ± 0.00
	LSVM	58.36 ± 0.07	0.4517 ± 0.00
	PSVM	55.25 ± 0.00	0.3802 ± 0.00
	LSVMPP	$\mathbf{62.64} \pm \mathbf{0.05}$	$\mathbf{0.5024} \pm \mathbf{0.00}$
krumsiek	ESVM	82.19 ± 0.00	0.6770 ± 0.00
	LSVM	85.62 ± 0.39	0.6933 ± 0.92
	PSVM	84.06 ± 0.00	0.6908 ± 0.00
	LSVMPP	$\mathbf{86.25} \pm \mathbf{0.00}$	$\mathbf{0.7079} \pm \mathbf{0.00}$

Acknowledgement

This work is partially supported by NSF under grants CCF-2112665, CCF- 2217058, and by NIH under grant RF1MH125317.

References

1. Sarkar, Rik. "Low distortion delaunay embedding of trees in hyperbolic plane." *International Symposium on Graph Drawing*. Springer, Berlin, Heidelberg, 2011.
2. Nickel, Maximillian, and Douwe Kiela. "Learning continuous hierarchies in the lorentz model of hyperbolic geometry." *ICML*. PMLR, 2018.
3. Cho, H., DeMeo, B., Peng, J., and Berger, B. "Large-margin classification in hyperbolic space." *AISTATS* 2019.
4. Simeoni, Matthieu Martin Jean-Andre. "Statistics on manifolds applied to shape theory." *No. STUDENT*. 2013.