Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers

Yunhui Guo  Xudong Wang  Yubei Chen  Stella X. Yu
Why Hyperbolic (Non-Euclidean) Space?

Hyperbolic space can embed tree metric with low distortion, unlike Euclidean space.
Hyperbolic Neural Network (HNN)

They are made of Euclidean feature extractor + hyperbolic classifier

Exponential Map

\[ \exp_0^c(v) = \tanh(\sqrt{c}||v||) \frac{v}{\sqrt{c}||v||} \]

Better on hierarchical classification and few-shot learning

Ganea, Octavian, et al. "Hyperbolic neural networks." NeurIPS 2018.
HNN Better than ENN on Hierarchical Tasks Only

Known to outperform standard Euclidean neural net (ENN) only for hierarchical categorization tasks, greatly limiting its applicability.

\[ \text{HNN} > \text{ENN} \]

Ganea, Octavian, et al. "Hyperbolic neural networks." NeurIPS 2018.
HNN Better than ENN on Hierarchical Tasks Only

Known to outperform standard Euclidean neural net (ENN) only for hierarchical categorization tasks, greatly limiting its applicability.

Ganea, Octavian, et al. "Hyperbolic neural networks." NeurIPS 2018.
Hyperbolic Distance in Poincaré Ball Increases Exponentially

Stereographic Projection

Hyperboloid Model

Poincaré ball Model
Hyperbolic Distance in Poincaré Ball Increases Exponentially
Non-trivial Optimization Challenges of HNNs
Non-trivial Optimization Challenges of HNNs

Optimizing HNNs end-to-end with backpropagation

\[ \frac{\partial \ell}{\partial \mathbf{w}^E} = \left( \frac{\partial x^H}{\partial \mathbf{w}^E} \right)^T \frac{(1 - \|x^H\|^2)^2}{4} \nabla \ell(x^H) \]

Gradients w.r.t hyperbolic embedding
Non-trivial Optimization Challenges of HNNs

Optimizing HNNs end-to-end with backpropagation

\[
\frac{\partial \ell}{\partial w^E} = \left( \frac{\partial x^H}{\partial w^E} \right)^T \frac{(1 - \|x^H\|^2)^2}{4} \nabla \ell(x^H)
\]

Gradients w.r.t hyperbolic embedding

Hyperbolic embedding close to the boundary -> Gradients vanish
Consequence: Gradients Vanish for Larger Features

The trajectories of six sampled points during training

Training loss increases at the end of training
Our Solution: Feature Clipping

Avoid the ill-conditioned region while well preserving the hyperbolic property

\[
\text{Clip}(x^E; \text{Clip Value}) = \begin{cases} 
  x^E & \text{if } ||x^E|| \leq \text{Clip Value} \\
  \frac{\text{Clip Value}}{||x^E||} \cdot x^E & \text{if } ||x^E|| > \text{Clip Value}
\end{cases}
\]
Clipped HNNs Are Super HNNs!

Results on Standard Benchmarks

Accuracy

MNIST  CIFAR10  CIFAR100  ImageNet

Baseline HNN  Clipped HNN
Clipped HNNs $\rightarrow$ ENN on Flat Categorization
Super HNNs on Few-shot Learning

Results on Few-shot Learning

Accuracy

- CUB 1-shot 5-way
- CUB 5-shot 5-way
- ImageNet 1-shot 5-way
- ImageNet 5-shot 5-way

Baseline HNN
Clipped HNN

Khrulkov, Valentin, et al. "Hyperbolic image embeddings." CVPR 2020.
Super HNNs on Few-shot Learning

Results on Few-shot Learning

- Baseline HNN
- Clipped HNN
- ENN

Khrulkov, Valentin, et al. "Hyperbolic image embeddings." CVPR 2020.
Generalize to Out-of-Distribution Detection

In-distribution data: CIFAR10 and CIFAR100
Out-of-distribution data: ISUN, Place365, Texture, SVHN, LSUN-Crop and LSUN-Resize
Generalize to Out-of-Distribution Detection

In-distribution data: CIFAR10 and CIFAR100
Out-of-distribution data: ISUN, Place365, Texture, SVHN, LSUN-Crop and LSUN-Resize
Impact on the Learned Hyperbolic Feature

HNNs

Clipped HNNs

Test Accuracy

MNIST Class Index

Clipped HNN
Baseline HNN

Poincaré Hyperplane
A Clipped Hyperbolic Space Is Still Hyperbolic
Learning Hyperbolic Word Embeddings

Nickel, Maximillian, et al. "Poincaré embeddings for learning hierarchical representations." NeurIPS 2017.
Clipped Hyperbolic Classifiers Are Super-Hyperbolic Classifiers

Yunhui Guo    Xudong Wang    Yubei Chen    Stella X. Yu

Contributions

- OOD Detection
- Training Dynamics of HNNs
- Feature Clipping: Clip the Euclidean embedding before the exponential map

Clipped HNNs learn more balanced and discriminative feature space

Hyperbolic Feature Space

- Hyperbolic Space
- Feature Clipping: $\text{CLIP}(x^E; r) = \min\{1, \frac{r}{\|x^E\|}\} \cdot x^E$

Training Dynamics of HNNs

HNNs and Clipped HNNs

Hyperbolic Feature Space

- HNNs underperform ENNs on standard benchmarks
- Non-Euclidean space with constant negative curvature
- Can embed tree-like data continuously with low distortion

OOD Detection

CIFAR10

HNNs show stronger OOD detection ability than ENNs

CIFAR100

Impact of Dimensionality

- HNNs outperform ENNs when the feature dimensionality is low

Impact of Clip Value

- The clipping value should not be too large (causing vanishing gradient problem) or too small (no enough capacity).

Hyperbolic Space

- HNNs consist of ENN feature extractors and hyperbolic classifiers
- Gradients vanish when the embedding is close to the boundary

Adversarial Robustness

HNNs show stronger adversarial robustness

Standard Benchmarks and Few-shot Learning

- HNNs perform better than baseline HNNs

Feature Clipping: Clip the Euclidean embedding before the exponential map

HNNs consist of ENN feature extractors and hyperbolic classifiers

- Gradients vanish when the embedding is close to the boundary

HNNs show stronger adversarial robustness

Impact of Dimensionality

- HNNs outperform ENNs when the feature dimensionality is low

Impact of Clip Value

- The clipping value should not be too large (causing vanishing gradient problem) or too small (no enough capacity).