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

![Image](https://via.placeholder.com/150)

\[
\text{Exp}_0^c(v) = \tanh(\sqrt{c}\|v\|) \frac{v}{\sqrt{c}\|v\|}
\]

Better on hierarchical classification and few-shot learning

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.

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

Hyperbolic Distance in Poincaré Ball Increases Exponentially

Stereographic Projection

Hyperboloid Model

Poincaré ball Model
Hyperbolic Distance in Poincaré Ball Increases Exponentially

Stereographic Projection

Hyperboloid Model

Poincaré ball Model
Non-trivial Optimization Challenges of HNNs
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
Non-trivial Optimization Challenges of HNNs

Optimizing HNNs end-to-end with backpropagation

\[
\frac{\partial \ell}{\partial \mathbf{w}^E} = \left( \frac{\partial \mathbf{x}^H}{\partial \mathbf{w}^E} \right)^T \frac{(1 - \|\mathbf{x}^H\|_2^2)^2}{4} \nabla \ell(\mathbf{x}^H)
\]

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

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

Results on Standard Benchmarks

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Baseline HNN</th>
<th>Clipped HNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>CIFAR10</td>
<td>85</td>
<td>95</td>
</tr>
<tr>
<td>CIFAR100</td>
<td>70</td>
<td>85</td>
</tr>
<tr>
<td>ImageNet</td>
<td>55</td>
<td>70</td>
</tr>
</tbody>
</table>
Clipped HNNs $\rightarrow$ ENN on Flat Categorization
Super HNNs on Few-shot Learning

Super HNNs on Few-shot Learning

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

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

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

CLIP(x^E; r) = min\{1, \frac{r}{||x^E||}\} \cdot x^E

Hyperbolic Space

- Non-Euclidean space with constant negative curvature
- Can embed tree-like data continuously with low distortion

HNNs Underperform ENNs on Standard Benchmarks

Hyperbolic Feature Space

HNNs consist of ENN feature extractors and hyperbolic classifiers

Feature Clipping: Clip the Euclidean embedding before the exponential map

HNNs

Clipped HNNs

- Escaping the ill-conditioned region
- Preserving the hyperbolic property

HNNs show stronger OOD detection ability than ENNs

Hyperbolic Classification

HNNs show stronger adversarial robustness

Clipped HNNs show better results compared with baseline HNNs

Standard Benchmarks and Few-shot Learning

Clipped HNNs learn more balanced and discriminative feature space

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).