Hyperbolic Neural Networks
Octavian-Eugen Ganea*, Gary Béciteauln*, Thomas Hofmann
Department of Computer Science, ETH Zürich, Switzerland

Hyperbolic Geometry
- Hyperbolic space - constant negative curvature
- Non-Euclidean embeddings
- Exponential volume growth (unlike polynomially in Euclidean space) ⇒ Exponential capacity increase
- Mathematically, can isometrically (preserve distances) embed:
  - approximate tree-like structures, or w/ heterogeneous topology
  - scale-free networks - node degree distributions follow a power-law

Recently, hyperbolic embeddings in ML - e.g. Nickel & Kiela,2017

How to use Hyperbolic Embeddings in Neural Nets
- Scale-free than in Euclidean space due to the negative curvature
  - from J. Lamping et al. “A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies.” SIGCHI 1995.
- Polynomially in Euclidean space

Downstream (preserve distances) embed:
- Heterogeneous topology

Our contributions
Use Gyro-vector spaces to generalize basic operations and neural networks from Euclidean to hyperbolic spaces:
- Gyro- vs. - analogue of Euclidean vector spaces
  - used in relativity theory (speeds of particles are hyperbolic)
- Vector addition: \( x + y \leftarrow x \oplus y \)
- Scalar multiplication: \( r x \leftarrow r \otimes x \)
- Closed form distance \( d_h(x, y) = (2/\sqrt{c}) \tanh^{-1} (\sqrt{c} \Vert x - y \Vert ) \)
- Closed form geodesics: \( \gamma_{x \rightarrow y}(t) := x \oplus t (x \oplus y - x) \otimes t \)

1) We connect Gyro-vs and Riemannian hyperbolic geometry \( T_h M \)
- Closed form \( \exp(x), \log(x) \)
- Closed parallel transport (move across tangent spaces)

2) Hyperbolic Feed-forward Neural Networks
- Möbius version of \( f: \mathbb{R}^n \rightarrow \mathbb{R}^n \) (e.g. pointwise non-linearity):
  \( f^M: \mathbb{D}^n \rightarrow \mathbb{D}^n, \quad f^M(x) := \exp(f(\log(x))) \)
- Matrix - vector multiplication:
  \( M^c(x) = (1/\sqrt{c}) \tanh \left( \frac{||Mx||}{||x||} \tanh^{-1}(\sqrt{c}||x||) \right) \frac{Mx}{||Mx||} \)
  - Properties: matrix associativity, scalar-matrix associativity, preserved rotations

3) Hyperbolic Softmax layer - Multiclass Logistic Regression
- Hyperbolic hyperplane:
  \( H^c_{a,p} = \{ x \in \mathbb{D}^n_+ : (p \otimes x, a) = 0 \} \)
- Theorem: closed form of \( d_h(x, H^c_{a,p}) \)
- Final MLR formula (based on Lebanon and Lafferty,2004):
  \( p(y = k|x) \propto \exp \left( \frac{\lambda^c_k ||a_k|| \sinh^{-1} \left( \frac{2\sqrt{c}(p_k \otimes x, a_k)}{(1 - c) - p_k \otimes x, a_k} \right) }{\sqrt{c}} \right) \)
  - Property: All our models recover their Euclidean variants when curvature \( c \rightarrow 0 \).

4) Hyperbolic Recurrent Networks, e.g. hGRU
  \[ \begin{align*}
  r_t &= \sigma \log \left( W_r x_t \oplus h_{t-1} \oplus U_r \otimes h_{t-1} \oplus b \right) \\
  \bar{h}_t &= \varphi^c \left( \left( W \operatorname{diag}(r_t) \right) \otimes \left( h_{t-1} \oplus U \otimes x_t \oplus b \right) \right) \\
  h_t &= h_{t-1} \oplus \operatorname{diag}(z_t) \otimes \left( -h_{t-1} \oplus h_t \right)
  \end{align*} \]
- Hyperbolic hidden states
  - Theorem: update-gate mechanism derived from time-warping invariance principle (via gyro-derivative and gyro-chain-rule)

Experiments
1) Textual Entailment tasks (semantic + syntactic).

<table>
<thead>
<tr>
<th>Test Set</th>
<th>SNLI</th>
<th>SLUI</th>
<th>PPREF</th>
<th>SFREF</th>
<th>Prefix-50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Euclidean RNN</td>
<td>79.34%</td>
<td>89.62%</td>
<td>81.71%</td>
<td>72.10%</td>
<td>76.50%</td>
</tr>
<tr>
<td>Hyp RNN/FFNN, Eucl MLR</td>
<td>98.18%</td>
<td>99.36%</td>
<td>98.74%</td>
<td>98.47%</td>
<td>98.77%</td>
</tr>
<tr>
<td>Fully Hyperbolic GRU</td>
<td>91.52%</td>
<td>95.96%</td>
<td>88.47%</td>
<td>95.04%</td>
<td>96.77%</td>
</tr>
<tr>
<td>Hyp RNN/FFNN, Eucl MLR</td>
<td>98.19%</td>
<td>99.74%</td>
<td>98.26%</td>
<td>98.44%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

2) MLR experiments.
Test F1 classification scores (%) for 4 subtypes of WordNet tree.

<table>
<thead>
<tr>
<th>WordNet subtype</th>
<th>Model</th>
<th>d = 2</th>
<th>d = 3</th>
<th>d = 5</th>
<th>d = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal</td>
<td>Hyp</td>
<td>47.43 ± 1.07</td>
<td>91.92 ± 0.64</td>
<td>98.07 ± 0.55</td>
<td>99.9 ± 0.59</td>
</tr>
<tr>
<td>Eccl</td>
<td>41.69 ± 0.39</td>
<td>64.61 ± 1.96</td>
<td>93.59 ± 1.18</td>
<td>99.36 ± 0.18</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Hyp</td>
<td>81.72 ± 0.17</td>
<td>89.57 ± 2.07</td>
<td>87.89 ± 0.80</td>
<td>91.22 ± 1.29</td>
</tr>
<tr>
<td>PNN-1</td>
<td>Eccl</td>
<td>61.13 ± 0.12</td>
<td>61.31 ± 0.21</td>
<td>63.72 ± 0.04</td>
<td>91.38 ± 1.29</td>
</tr>
<tr>
<td>Worker</td>
<td>Hyp</td>
<td>12.68 ± 0.82</td>
<td>24.09 ± 1.45</td>
<td>55.46 ± 5.19</td>
<td>68.64 ± 11.85</td>
</tr>
<tr>
<td>Eccl</td>
<td>10.96 ± 0.94</td>
<td>22.49 ± 0.94</td>
<td>52.25 ± 3.16</td>
<td>47.25 ± 4.91</td>
<td></td>
</tr>
<tr>
<td>Mammal</td>
<td>Hyp</td>
<td>15.58 ± 0.04</td>
<td>41.06 ± 1.87</td>
<td>35.35 ± 1.31</td>
<td>77.75 ± 5.08</td>
</tr>
<tr>
<td>Eccl</td>
<td>15.10 ± 0.31</td>
<td>44.99 ± 1.18</td>
<td>52.51 ± 0.85</td>
<td>56.1 ± 2.21</td>
<td></td>
</tr>
</tbody>
</table>

Hyperbolic (left) vs Direct Euclidean (right) binary MLR used to classify nodes as being part in the GROUP.P.01 subtype of the WordNet noun hierarchy solely based on their Poincaré embeddings.

This work to the rescue :)