HyperSPNs: Compact and Expressive Probabilistic Circuits

Andy Shih
Dorsa Sadigh
Stefano Ermon
Density Estimation
Density Estimation

Data → Distribution
Density Estimation

Data ➔ Distribution
Density Estimation
Density Estimation

Distribution

Sample
Density Estimation

Distribution

Sample
Density Estimation

\[ p(x) = 0.13 \]
Density Estimation

Distribution

Sample  Density  Inference?
Density Estimation

\[ p(\ ) = ? \]

Distribution

Sample  Density  Inference?
Inference - Marginals & Conditionals

What's the probability that:
- road 1 is under construction?
- road 7 is busy given road 9 is under construction?

\[ p(r_1 = c, r_2, \ldots, r_{100}) \]

\[ p(r_7 = b, r_9 = c) / p(r_9 = c) \]
Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]
Inference - Marginals & Conditionals

\( X = \{r_1, r_2, \ldots, r_{100}\} \)

What’s the probability that:
- road 1 is under construction?
Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]

What’s the probability that:
- road 1 is under construction?

\[ \sum_{r_2, \ldots, r_{100}} p(r_1 = c, r_2, \ldots, r_{100}) \]

\[ \frac{p(r_7 = b, r_9 = c)}{p(r_9 = c)} \]
Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]

What’s the probability that:
- road 1 is under construction?
  \[ \sum_{r_2, \ldots, r_{100}} p(r_1 = c, r_2, \ldots, r_{100}) \]
- road 7 is busy given road 9 is under construction?
Inference - Marginals & Conditionals

\[ X = \{r_1, r_2, \ldots, r_{100}\} \]

What’s the probability that:
- road 1 is under construction?

\[ \sum_{r_2, \ldots, r_{100}} p(r_1 = c, r_2, \ldots, r_{100}) \]

- road 7 is busy given road 9 is under construction?

\[ \frac{p(r_7 = b, r_9 = c)}{p(r_9 = c)} \]
Modeling Families

Mixture Models

tractable

expressive
Modeling Families

Mixture Models

tractable

expressive
Modeling Families

- Mixture Models
- tractable
- expressive efficient
Modeling Families

Mixture Models

tractable

Score-based
VAEs
Flows
GANs

expressive efficient
Modeling Families

tractable

expressive efficient

Mixture Models

DPPs

SPNs

ACs

Score-based

VAEs

Flows

GANs
Modeling Families

- Mixture Models
  - tractable

- Score-based
  - DPPs
  - VAEs
  - Flows
  - GANs

- Expressive
  - efficient

Stanford University
Modeling Families

- Mixture Models
  - tractable

- Determinantal Point Processes
- SPNs
- ACs
- Score-based
- VAEs
- Flows
- GANs

expressive efficient
Modeling Families

Mixture Models

tractable

DPPs

SPNs

ACs

Score-based

VAEs

Flows

GANs

expressive efficient
Modeling Families

Mixture Models

tractable

DPPs

SPNs

ACs

Score-based

VAEs

Flows

GANs

effective

Stanford University
Modeling Families

Mixture Models

tractable

DPPs

SPNs

ACs

Score-based

VAEs

Flows

GANs

expressive efficient

Stanford University
Modeling Families

Mixture Models

- tractable

- DPPs
- SPNs
- ACs

GANs
- Score-based
- VAEs
- Flows
- GANs

expressive efficient
Modeling Families

- Mixture Models
  - tractable

- SPNs
  - DPPs
  - SPNs
  - ACs

- Score-based
  - VAEs
  - Flows
  - GANs

Stanford University
Modeling Families

- **Mixture Models**
  - tractable
- **DPPs**
- **SPNs**
- **ACs**
- **Score-based**
- **VAEs**
- **Flows**
- **GANs**

**expressive**

**efficient**

Stanford University
Modeling Families

Mixture Models

- tractable

SPNs
- DPPs
- ACs

GANs

Score-based
- VAEs
- Flows
- GANs

expressive efficient
Modeling Families

- Mixture Models
  - tractable

- SPNs
- DPPs
- ACs

- Score-based
- VAEs
- Flows
- GANs

- expressive
  - efficient
Modeling Families

Mixture Models

tractable

DPPs

SPNs

Arithmetic Circuits

Score-based

VAEs

Flows

GANs

expressive efficient
Modeling Families

Mixture Models
- tractable

DPPs
SPNs
ACs

Score-based
VAEs
Flows
GANs

expressive
efficient
Modeling Families

Mixture Models

tractable

DPPs
SPNs
ACs

GANs
Flows
VAEs
Score-based

expressive efficient
Modeling Families

- Mixture Models
- ACs
  - SPNs
  - DPPs

- Score-based
  - VAEs
  - Flows
  - GANs

tractable

expressive efficient
Sum Product Networks

base distributions
Sum Product Networks

sum nodes

mixture

base distributions

0.3 0.7
Sum Product Networks

Sum nodes

Product nodes

Base distributions

Mixture

Factorization

0.3

0.7
Sum Product Networks
Sum Product Networks

Feed-forward network

Defines a computation graph
Sum Product Networks

Feed-forward network

Defines a computation graph
Sum Product Networks

Feed-forward network

Defines a computation graph

Train via gradient descent
Sum Product Networks

- Feed-forward network
- Defines a computation graph
- Train via gradient descent
Sum Product Networks

- Feed-forward network
- Defines a computation graph
- Train via gradient descent
- Multilinear polynomial over base distributions
Sum Product Networks

✅ Marginals in one forward pass
Sum Product Networks

✔ Marginals in one forward pass

Recall...

Inference - Marginals & Conditionals

\[ X = \{r_1, r_2, \ldots, r_{100}\} \]

What's the probability that:
- road 1 is under construction?
\[ \sum_{r_2 \ldots r_{100}} p(r_1 = c, r_2, \ldots, r_{100}) \]
Sum Product Networks

Marginals in one forward pass

Recall…

Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]

What's the probability that:
- road 1 is under construction?

\[ \sum_{r_1 = c_1, r_2, \ldots, r_{100}} p(r_1 = c_1, r_2, \ldots, r_{100}) \]
Sum Product Networks

Marginals in one forward pass

Recall...

Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]

What's the probability that:
- road 1 is under construction?

\[ \sum_{r_1, r_2, \ldots, r_{100}} \rho(r_1 = c, r_2, \ldots, r_{100}) \]
Sum Product Networks

✅ Marginals in one forward pass

Recall…

Inference - Marginals & Conditionals

\[ X = \{ r_1, r_2, \ldots, r_{100} \} \]

What's the probability that:
- road 1 is under construction?

\[ \sum_{r_1, r_2, \ldots, r_{100}} \rho(r_1 = c, r_2, \ldots, r_{100}) \]

Stanford University
Sum Product Networks

Marginals in one forward pass

Recall...

Inference - Marginals & Conditionals

\[ X = \{r_1, r_2, \ldots, r_{100}\} \]

What's the probability that:
- road 1 is under construction?

\[ \sum_{r_1, r_2, \ldots, r_{100}} p(r_1 = c, r_2, \ldots, r_{100}) \]
SPN Architecture

Structure Learning — expensive
SPN Architecture

Structure Learning — expensive

Prescribed Structure
  RAT-SPNs (UAI’19)

  EiNETs (ICML’20)
    - 9.4M parameters
SPN Architecture

Structure Learning — expensive

Prescribed Structure
- RAT-SPNs (UAI’19)
- EiNETs (ICML’20)
  - 9.4M parameters
Modeling Families

- Mixture Models
  - DPPs
  - SPNs
  - ACs

- Score-based
  - VAEs
  - Flows
  - GANs

tractable

expressive efficient
Modeling Families

The bigger the network, the more expressive the model
Large SPNs
Large SPNs

Regularization Choices
Large SPNs

Regularization Choices

<table>
<thead>
<tr>
<th>Dropout</th>
<th>Discriminative only</th>
</tr>
</thead>
</table>
Large SPNs

Regularization Choices

<table>
<thead>
<tr>
<th>Dropout</th>
<th>Discriminative only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight Decay</td>
<td>Many parameters</td>
</tr>
</tbody>
</table>
Large SPNs

Regularization Choices

<table>
<thead>
<tr>
<th>Dropout</th>
<th>Discriminative only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight Decay</td>
<td>Many parameters</td>
</tr>
<tr>
<td>HyperSPN</td>
<td>Few parameters</td>
</tr>
<tr>
<td></td>
<td>Memory efficient</td>
</tr>
<tr>
<td></td>
<td>Better generalization</td>
</tr>
<tr>
<td></td>
<td>our proposal</td>
</tr>
</tbody>
</table>

Stanford University
HyperSPN
HyperSPN

1 \times k

k \times k

k \times 2
HyperSPN

embedding
dimension $h$

underlying SPN

$1 \times k$

$k \times k$

$k \times 2$
HyperSPN

embedding

dimension $h$

$h$ inputs

$k^2$ outputs

underlying SPN

$1 \times k$

$k \times k$

$k \times 2$
HyperSPN

embeddings → neural network → SPN

$\chi$

forward pass

$p(x)$
HyperSPN

embeddings → neural network

backward pass → forward pass

$X$ → SPN → $P(X)$
HyperSPN
HyperSPN

![Diagram showing train and test sets]

- train
- test
HyperSPN

- train
- test
HyperSPN

Many clusters
Not constrained

SPN-Large

train  test  cluster
HyperSPN

SPN-Large
Many clusters
Not constrained

SPN-Small
Few clusters

train  test  cluster
HyperSPN

SPN-Large
- Many clusters
- Not constrained

SPN-Small
- Few clusters

HyperSPN
- Many clusters
- Constrained

- train
- test
- cluster
Better Generalization
Better Generalization
## Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLTCS</td>
<td>16</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
</tr>
<tr>
<td>Netflix</td>
<td>100</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
</tr>
<tr>
<td>20Newsgrpr</td>
<td>910</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay</th>
<th>Adam HyperSPN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-LH</td>
<td># Params</td>
<td>Log-LH</td>
</tr>
<tr>
<td>NLTCS</td>
<td>16</td>
<td>-6.02 40050</td>
<td>-6.01 9115</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05 42550</td>
<td>-6.05 9615</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14 160050</td>
<td>-2.13 33115</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36 172550</td>
<td>-13.26 35615</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18 25050</td>
<td>-39.83 51115</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98 25050</td>
<td>-52.75 51115</td>
</tr>
<tr>
<td>Netflix</td>
<td>100</td>
<td>-57.15 25050</td>
<td>-56.74 51115</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09 277550</td>
<td>-35.36 56615</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91 337550</td>
<td>-10.89 68615</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76 407550</td>
<td>-30.79 82615</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td>-98.41 45050</td>
<td>-98.49 91115</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93 47550</td>
<td>-10.89 96115</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40 735050</td>
<td>-9.90 148115</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01 125050</td>
<td>-34.90 251115</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99 125050</td>
<td>-51.32 251115</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91 2097550</td>
<td>-158.60 420615</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14 2222550</td>
<td>-85.65 445615</td>
</tr>
<tr>
<td>20Newsggrp</td>
<td>910</td>
<td>-154.37 2275050</td>
<td>-152.49 456115</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01 2645050</td>
<td>-254.44 530115</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23 3890050</td>
<td>-28.25 779115</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay</th>
<th>Adam HyperSPN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log-LH  # Params</td>
<td>Log-LH  # Params</td>
</tr>
<tr>
<td>NLTCS</td>
<td>16</td>
<td>-6.02 40050</td>
<td>-6.01 9115</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05 42550</td>
<td>-6.05 9615</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14 160050</td>
<td>-2.13 33115</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36 172550</td>
<td>-13.26 35615</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18 250050</td>
<td>-39.83 51115</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98 250050</td>
<td>-52.75 51115</td>
</tr>
<tr>
<td>Netflix</td>
<td>100</td>
<td>-57.15 250050</td>
<td>-56.74 51115</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09 277550</td>
<td>-35.36 56615</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91 337550</td>
<td>-10.89 68615</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76 407550</td>
<td>-30.79 82615</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td><strong>-98.41</strong> 450050</td>
<td>-98.49 91115</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93 475050</td>
<td>-10.89 96115</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40 735050</td>
<td>-9.90 148115</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01 1250050</td>
<td>-34.90 251115</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99 1250050</td>
<td>-51.32 251115</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91 2097550</td>
<td>-158.60 420615</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14 2222550</td>
<td>-85.65 445615</td>
</tr>
<tr>
<td>20Newsgrp</td>
<td>910</td>
<td>-154.37 2275050</td>
<td><strong>-152.49</strong> 456115</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01 2645050</td>
<td>-254.44 530115</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23 3890050</td>
<td><strong>-28.25</strong> 779115</td>
</tr>
</tbody>
</table>
### Results

- Better log-LH
- Fewer # params

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay</th>
<th>Adam HyperSPN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log-LH</td>
<td># Params</td>
</tr>
<tr>
<td>NLTCS</td>
<td>16</td>
<td>-6.02</td>
<td>40050</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05</td>
<td>42550</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14</td>
<td>160050</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36</td>
<td>172550</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18</td>
<td>250050</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98</td>
<td>250050</td>
</tr>
<tr>
<td>Netflix</td>
<td>100</td>
<td>-57.15</td>
<td>250050</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09</td>
<td>277550</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91</td>
<td>337550</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76</td>
<td>407550</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td>-98.41</td>
<td>450050</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93</td>
<td>475050</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40</td>
<td>735050</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01</td>
<td>1250050</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99</td>
<td>1250050</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91</td>
<td>2097550</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14</td>
<td>2222550</td>
</tr>
<tr>
<td>20Newsgrp</td>
<td>910</td>
<td>-154.37</td>
<td>2275050</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01</td>
<td>2645050</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23</td>
<td>3890050</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay</th>
<th>Adam HyperSPN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Log-LH # Params</td>
<td>Log-LH # Params</td>
</tr>
<tr>
<td>NLTCs</td>
<td>16</td>
<td>-6.02 40050</td>
<td><strong>-6.01</strong> 9115</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05 42550</td>
<td><strong>-6.05</strong> 9615</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14 160050</td>
<td><strong>-2.13</strong> 33115</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36 172550</td>
<td><strong>-13.26</strong> 35615</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18 250050</td>
<td><strong>-39.83</strong> 51115</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98 250050</td>
<td><strong>-52.75</strong> 51115</td>
</tr>
<tr>
<td>Netflix</td>
<td>100</td>
<td>-57.15 250050</td>
<td><strong>-56.74</strong> 51115</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09 277550</td>
<td><strong>-35.36</strong> 56615</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91 337550</td>
<td><strong>-10.89</strong> 68615</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76 407550</td>
<td><strong>-30.79</strong> 82615</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td><strong>-98.41</strong> 450050</td>
<td>-98.49 91115</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93 475050</td>
<td><strong>-10.89</strong> 96115</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40 735050</td>
<td><strong>-9.90</strong> 148115</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01 1250050</td>
<td><strong>-34.90</strong> 251115</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99 1250050</td>
<td><strong>-51.32</strong> 251115</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91 2097550</td>
<td><strong>-158.60</strong> 420615</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14 2222550</td>
<td><strong>-85.65</strong> 445615</td>
</tr>
<tr>
<td>20Newsgroup</td>
<td>910</td>
<td>-154.37 2275050</td>
<td><strong>-152.49</strong> 456115</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01 2645050</td>
<td><strong>-254.44</strong> 530115</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23 3890050</td>
<td><strong>-28.25</strong> 779115</td>
</tr>
</tbody>
</table>

better log-LH  fewer # params

HyperSPNs
- regularize by encoding parameters with small NN
## Results

- Better log-LH
- Fewer # params

### HyperSPNs

- Regularize by encoding parameters with small NN
- Better generalization
- More memory efficient
- Keeps tractability of SPNs

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay Log-LH</th>
<th>Adam HyperSPN Log-LH</th>
<th>Variables</th>
<th>Adam Weight Decay # Params</th>
<th>Adam HyperSPN # Params</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLTCS</td>
<td>16</td>
<td>-6.02</td>
<td>-6.01</td>
<td>9115</td>
<td>40050</td>
<td>9115</td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05</td>
<td>-6.05</td>
<td>9615</td>
<td>42550</td>
<td>9615</td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14</td>
<td>-2.13</td>
<td>33115</td>
<td>160050</td>
<td>33115</td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36</td>
<td>-13.26</td>
<td>35615</td>
<td>172550</td>
<td>35615</td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18</td>
<td>-39.83</td>
<td>51115</td>
<td>250050</td>
<td>51115</td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98</td>
<td>-52.75</td>
<td>51115</td>
<td>250050</td>
<td>51115</td>
</tr>
<tr>
<td>Netfix</td>
<td>100</td>
<td>-57.15</td>
<td>-56.74</td>
<td>51115</td>
<td>250050</td>
<td>51115</td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09</td>
<td>-35.36</td>
<td>56615</td>
<td>277550</td>
<td>56615</td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91</td>
<td>-10.89</td>
<td>68615</td>
<td>337550</td>
<td>68615</td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76</td>
<td>-30.79</td>
<td>82615</td>
<td>407550</td>
<td>82615</td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td>-98.41</td>
<td>-98.41</td>
<td>91115</td>
<td>450050</td>
<td>91115</td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93</td>
<td>-10.89</td>
<td>96115</td>
<td>475050</td>
<td>96115</td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40</td>
<td>-9.90</td>
<td>148115</td>
<td>735050</td>
<td>148115</td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01</td>
<td>-34.90</td>
<td>251115</td>
<td>1250050</td>
<td>251115</td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99</td>
<td>-51.32</td>
<td>251115</td>
<td>1250050</td>
<td>251115</td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91</td>
<td>-158.60</td>
<td>420615</td>
<td>2097550</td>
<td>420615</td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14</td>
<td>-85.65</td>
<td>445615</td>
<td>2222550</td>
<td>445615</td>
</tr>
<tr>
<td>20Newsgroups</td>
<td>910</td>
<td>-154.37</td>
<td>-152.49</td>
<td>456115</td>
<td>2275050</td>
<td>456115</td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01</td>
<td>-254.44</td>
<td>530115</td>
<td>2645050</td>
<td>530115</td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23</td>
<td>-28.25</td>
<td>779115</td>
<td>3890050</td>
<td>779115</td>
</tr>
</tbody>
</table>
**Results**

<table>
<thead>
<tr>
<th>Name</th>
<th>Variables</th>
<th>Adam Weight Decay Log-LH</th>
<th>Adam HyperSPN Log-LH</th>
<th>fewer # params</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLTCS</td>
<td>16</td>
<td>-6.02</td>
<td>-6.01</td>
<td></td>
</tr>
<tr>
<td>MSNBC</td>
<td>17</td>
<td>-6.05</td>
<td>-6.05</td>
<td></td>
</tr>
<tr>
<td>KDDCup2k</td>
<td>64</td>
<td>-2.14</td>
<td>-2.13</td>
<td></td>
</tr>
<tr>
<td>Plants</td>
<td>69</td>
<td>-13.36</td>
<td>-13.26</td>
<td></td>
</tr>
<tr>
<td>Audio</td>
<td>100</td>
<td>-40.18</td>
<td>-39.83</td>
<td></td>
</tr>
<tr>
<td>Jester</td>
<td>100</td>
<td>-52.98</td>
<td>-52.75</td>
<td></td>
</tr>
<tr>
<td>Netfix</td>
<td>100</td>
<td>-57.15</td>
<td>-56.74</td>
<td></td>
</tr>
<tr>
<td>Accidents</td>
<td>111</td>
<td>-36.09</td>
<td>-35.36</td>
<td></td>
</tr>
<tr>
<td>Retail</td>
<td>135</td>
<td>-10.91</td>
<td>-10.89</td>
<td></td>
</tr>
<tr>
<td>Pumsb-star</td>
<td>163</td>
<td>-31.76</td>
<td>-30.79</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>180</td>
<td>-98.41</td>
<td>-98.49</td>
<td></td>
</tr>
<tr>
<td>Kosarek</td>
<td>190</td>
<td>-10.93</td>
<td>-10.89</td>
<td></td>
</tr>
<tr>
<td>MSWeb</td>
<td>294</td>
<td>-10.40</td>
<td>-9.90</td>
<td></td>
</tr>
<tr>
<td>Book</td>
<td>500</td>
<td>-35.01</td>
<td>-34.90</td>
<td></td>
</tr>
<tr>
<td>EachMovie</td>
<td>500</td>
<td>-52.99</td>
<td>-51.32</td>
<td></td>
</tr>
<tr>
<td>WebKB</td>
<td>839</td>
<td>-159.91</td>
<td>-158.60</td>
<td></td>
</tr>
<tr>
<td>Reuters-52</td>
<td>889</td>
<td>-90.14</td>
<td>-85.65</td>
<td></td>
</tr>
<tr>
<td>20Newsgrp</td>
<td>910</td>
<td>-154.37</td>
<td>-152.49</td>
<td></td>
</tr>
<tr>
<td>BBC</td>
<td>1058</td>
<td>-262.01</td>
<td>-254.44</td>
<td></td>
</tr>
<tr>
<td>Ad</td>
<td>1556</td>
<td>-52.23</td>
<td>-28.25</td>
<td></td>
</tr>
</tbody>
</table>

**HyperSPNs**
- regularize by encoding parameters with small NN
- better generalization
- more memory efficient
- keeps tractability of SPNs

**Paper / Code:**

Stanford University