Online Structure Learning for Feed-Forward and Recurrent Sum-Product Networks

Motivation

Feature engineering replaced by architecture engineering
• But architecture design is an art (trial and error)
• Need automated way to learn structure
Contribution: online structure learning algorithm for Recurrent and Feed-forward SPNs

Online Structure Learning with Running Average Update (oSLRAU)

At each product node, monitor covariance in the scope of the product node
If the correlation of two variables exceeds a threshold, introduce correlation
e.g. correlation(X₁, X₂) > threshold

Options:
1) create multivariate leaf distribution
2) create mixture distribution

Proof of Concept

Structure learned after 200 data points
Structure learned after 500 data points

Conclusions

Contributions:
• New online structure learning algorithm for both feed forward and recurrent SPNs
• Code available: github.com/kalraa/splrau

Future work
• Reduce complexity w.r.t. # of features from quadratic to linear
• Discriminative structure learning

Experiments

Large Continuous Datasets: average log likelihood comparison

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Rank</th>
<th>ILSPN</th>
<th>eSLRAU</th>
<th>RealNVP</th>
<th>OSLRAU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>1</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Fashion MNIST</td>
<td>2</td>
<td>0.19</td>
<td>0.20</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>Cifar10</td>
<td>3</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
</tr>
</tbody>
</table>

O SLRAU is better than
• Incremental LearnSPN (ILSPN)
• Real-Valued Preserving (RealNVP)
• Random Structures for gaussian SPNs

RSPN + oSLRAU is much faster and more accurate than
• RSPN + Search and Score.

With Pruning

Proof of Concept

Each node computes a probability over its scope
Scope of a node: set of variables in sub-SPN rooted at that node
Decomposable product node: children with disjoint scopes
Complete/smooth sum node: children with identical scopes
Decomposability + completeness \implies valid distribution

Semantics

Sum Product Network

Leaves: base distributions (e.g., Gaussians)
Interior nodes: sums and products
Edges:
• Unweighted below product nodes
• Non-negative weights below sum nodes

Evaluation

Probasillstic Inference

SPN represents a joint distribution, e.g., \( p(x) = 0.3 \cdot x = 0.5 \cdot 0.5 = 0.125 \)

O SLRAU is the new state of the art for online structure learning in both recurrent and regular SPNs

RSPN + oSLRAU is much faster and more accurate than
• RSPN + Search and Score.

Adapting to Non-Stationary datasets -- Trained on MNIST sorted from 0-9, then generated samples from the distribution. Pruning generates all 9s, No pruning generates many digits

1) Unroll network with as many template copies as length of data sequence
2) Share parameters across all template copies
3) Online parameter update: same as for feedforward networks
4) Online structure update:
   a) relabel scope of input interface nodes to binary hidden variables that allow scopes in different template copies to be the same
   b) detect correlations and update structure across all template copies in the same way