Multi-Prediction Deep Boltzmann Machines

Goodfellow, Mirza, Courville, Bengio

Vipul Venkataraman
Nov 29, 2016
Outline

• Goal of the paper [1]

• A primer on RBMs and DBMs
  • Training DBMs

• Proposed method: motivations and intuitions

• Results

• Conclusions
Goal of the paper

Make training unsupervised models great again!
Goal of the paper

Make training unsupervised models great again!

Deep Boltzmann Machines
Preliminaries

1 + 1 = 2
Deep Boltzmann Machines
Deep Boltzmann Machines

Training
Deep Boltzmann Machines

- Unsupervised
- Generative model
- Feature learning algorithm
Deep Boltzmann Machines

- Unsupervised
- Generative model
- Feature learning algorithm
Deep Boltzmann Machines

- Unsupervised
- Generative model
- Feature learning algorithm
Deep Boltzmann Machines
Training Methods
Deep Boltzmann Machines

Classification

• Exact inference is intractable

• Use mean field expectations of the hidden units
Training DBMs

Steps [2]:

1. Layer-wise pre-training
   • Unsupervised
   • RBMs as building blocks

2. Discriminative fine-tuning
   • Supervised
   • Back-propagation

Good reference: https://www.youtube.com/watch?v=Oq38plNmddk
Pre-training

RBM
Pre-training

RBM
Fine-tuning

MLP
Pre-training

- *Deep*: good features
- Can use any unsupervised algorithm
  - RBM (w/ CD)
  - Auto-encoder

Fine-tuning

- Won’t make drastic changes
- Need less labelled data
- Can use a lot of unlabelled data
• Greedy training, not considering global interactions
  • Many models, criteria
  • Extra classifier as well
• CD-\(k\): we don’t know \(k\)
  • Gradient approximation may be bad
An Aside: CD Intuition

$$\Delta w_{ij} = \epsilon(< v_i h_j >^0 - < v_i h_j >^1)$$
An Aside: CD Intuition

\[ \Delta w_{ij} = \epsilon (\langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1) \]
An Aside: CD Intuition

\[ \Delta w_{ij} = \epsilon (< v_i h_j >^0 - < v_i h_j >^1) \]

- Far away ‘holes’
- May want our particles to move many steps [3]
  - The mixing may get slower
- CD-1 -> CD-3 -> CD-10
Proposed method

• Mantra: Simplify
Proposed method

• Mantra: *Simplify*

• Many models -> one model
Proposed method

• Mantra: *Simplify*

• Many models -> one model

• Many criteria -> one criterion
Proposed method

- Mantra: *Simplify*
- Many models -> one model
- Many criteria -> one criterion
- Extra classification layer at the top -> unified model
Quick recap
Multi-Prediction Training
Random bit-mask
Example 1
Example 1, update 1
Example 1, update 2

Two mean-field fixed point updates
Example 2, all updates
Example 3, all updates
One iteration
One iteration

Minibatch
One iteration

Minibatch

Backprop
Performance

• Works well (results in a bit)

• Expensive though

• Needing to run several iterations for convergence
Multi-Inference Trick

Mean field
Multi-Inference Trick

Mean field

Multi-inference

average with the mean-field estimate
Multi-Inference Trick

Mean field

Multi-inference

average with the mean-field estimate

Nesterov's accelerated gradient descent
Can someone find me a suitable picture?
Multi-Inference Trick

Image: Goodfellow's defense
Setting

- Dataset: MNIST
- First layer: 500 hidden units
- Second layer: 1000 hidden units
- Minibatch size: 100 examples
- Test set: 10000 examples
- For more related results: [1]
Classification

<table>
<thead>
<tr>
<th>Model</th>
<th>Test error with fine-tuning</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;H 2009*</td>
<td>0.95</td>
</tr>
<tr>
<td>Centered DBM</td>
<td>1.22</td>
</tr>
<tr>
<td>MP-DBM</td>
<td>0.99</td>
</tr>
</tbody>
</table>

MP-DBM with 2X hidden units: 0.91
Robustness

Variation across hyperparameters

Validation set misclassification rate

Centering  Centering+  Multi-Prediction
Missing inputs

MNIST classification with missing inputs

- Standard DBM (no fine tuned stage)
- Centered DBM
- Standard DBM (+ fine tuned stage)
- MP-DBM
- MP-DBM (2X hidden units)

Test set misclassification rate vs. Probability of dropping each input unit
Conclusions

• Simpler, more intuitive methodology for training Deep Boltzmann Machines

• Improved accuracy for approximate inference problems
References


RAISE YOUR HAND

Questions?

ONE MORE TIME