Deep Belief Networks for Representational Learning

Sudhir Raman
Quick Recap

- Representational Learning.
- Helmholtz machine
Motivation

- Representational Learning.
- Biologically motivated models.
- Advances in machine learning approaches.
Background

- Neural Networks
  - Back-propagation

- Not a generative model

- Can be painfully slow

- Does not always converge
Background

- Stochastic Belief Networks

\[
\phi_i = \sum_j w_{ji} x_j \\
p_i = \frac{1}{1 + exp(-\phi_i)}
\]
Background

Boltzmann Machines:

- Restricted Boltzmann Machines (RBM)

- Gradient learning is often very slow in practice.

\[ \Delta w_{ij} = \eta[< x_i x_j >^0 - < x_i x_j >^\infty] \]
Deep Belief Networks (DBN)

- Combine the previous models.
  - Top 2 layers form a RBM.
  - Rest of the layers form a Belief net.

- Greedy approach –
  - Learn parameters one layer at a time.
  - Recognition and Generative parameters are symmetric.
Classification

- Add labels as a part of the generative model.

- Additional layers can be added to model noise.
Inference Algorithm

- Learn each layer (starting from the bottom)
  - Based on contrastive divergence.
    \[ \Delta w_{ij} = \eta[<v_i h_j>^0 - <v_i h_j>^1] \]

- Once a layer is learnt, sample hidden states from the dataset and learn the next layer using this as input.

- At a higher layer, add the classification label to the input and proceed as before.

- Fine tune the weights.
Project Goals

- Implement DBN for Handwritten digits:
  - Learn a generative model for handwritten digits.
  - Learn a classifier for pair-wise classification of handwritten digits.
MNIST dataset

- Handwritten digits [0-9]:
  - Size normalized and centered.
  - 28x28 Images.
  - Training set = 60000 , Test set = 10000

http://yann.lecun.com/exdb/mnist/index.html
Generative Model

- Experiment: $H1 = 50$ units, $H2 = 100$ units.

- Samples from the generative model for “2” and [0 1 2 3]
Pair-wise classification results:

- Comparison with units per layer - $H_1 = 50$ units, $H_2 = 200$ units.

- 4 Vs 9: Effect of increasing the number of units per layer - $H_1 = 200$ units, $H_2 = 800$ units.
Classification: 4 Vs 9

- Effect of increasing training set size.
- Effect of adding additional layers.

Error Snapshot
Final Pair-wise Classification

- With 2 layers: H1-200 units, H2-800 units.
Observations

- Easy to implement.

- Training the model is slow with increasing units.
  - More suitable for a parallel architecture.

- Learns well for the MNIST dataset.

- Performs better for larger number of hidden units.

- Adding more layers did not show improvement.
  - In principle, adding layers improves the variational bound.
  - Probably influenced by a minimum threshold for hidden units.
Summary

- Biologically inspired model.

- Learns well for the MNIST dataset.

- Simplified Inference –
  - Greedy approach.

- Future Work:
  - Further experiments with other datasets (objects, faces).
  - Defining other tasks with this model – Regression, Semi-supervised learning, Dim reduction, Incomplete data.
  - Fine tuning.
  - Investigation of the layer issue.


--- The End ---