Restricted Boltzmann Machines

Sargur N. Srihari
srihari@cedar.buffalo.edu

Topics

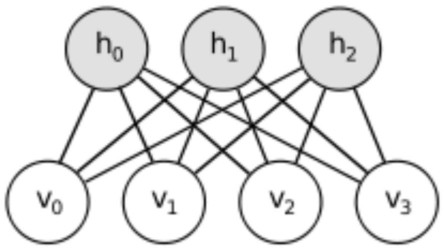
1. Boltzmann machines
    2. Restricted Boltzmann machines
    3. Deep Belief Networks
    4. Deep Boltzmann machines
    5. Boltzmann machines for continuous data
    6. Convolutional Boltzmann machines
    7. Boltzmann machines for structured and sequential outputs
    8. Other Boltzmann machines
9. Backpropagation through random operations
10. Directed generative nets
11. Drawing samples from autoencoders
12. Generative stochastic networks
13. Other generative schemes
14. Evaluating generative models
15. Conclusion

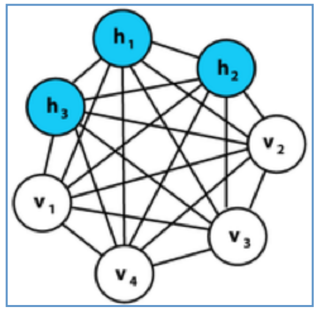
2. Restricted Boltzmann Machines

- Some of the most common building blocks of deep probabilistic models
- Units are organized as a layer of observed variables and a single layer of latent variables
- RBMs can be stacked on top of another

RBM is a bipartite graph

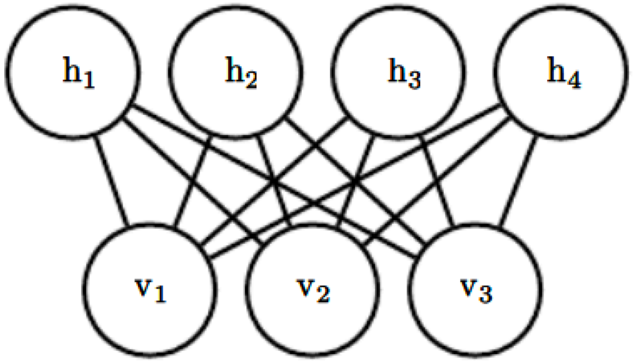
- RBM is a special case of Boltzmann machines and Markov networks
- No visible-visible and hidden-hidden connections– Bipartite graph

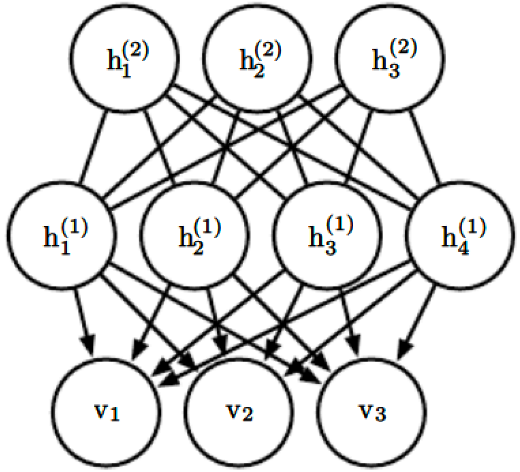


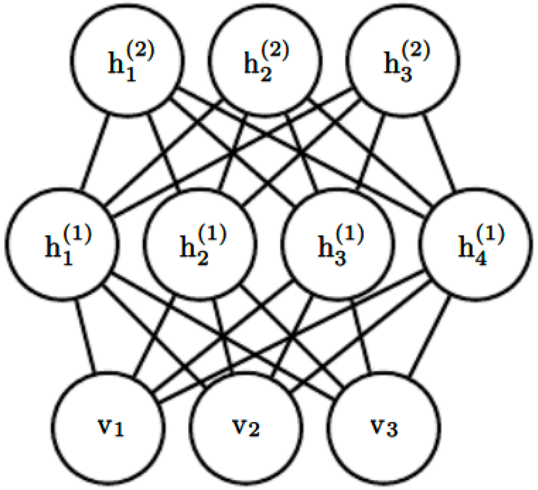


General BM

Some examples of RBMs







RBM is an undirected graphical model based on a bipartite graph
Typically no intralayer connections

Deep belief network
Is a hybrid graphical model
Involving both directed and
Undirected connections
No intralayer connections
Has multiple hidden layers

Deep Boltzmann Machine is an undirected graphical model with several layers of latent variables

Binary version of RBM

- Observed layer: set of $n_{\boldsymbol{v}}$ binary r.v.s, $\boldsymbol{v}$
- Latent or hidden set of n_h binary r.v.s $\boldsymbol{h}$
- Its energy function is

$$\mathrm{E}(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{h} - \boldsymbol{v}^{\mathrm{T}}\mathrm{W}\boldsymbol{h}$$

  - where $\boldsymbol{b}$, $\boldsymbol{c}$ and W are unconstrained, real-valued learnable parameters

- Thus model is divided into two groups of units $\boldsymbol{v}$ and $\boldsymbol{h}$ and the interaction between them is described by matrix W

RBM: an energy-based model

- RBM is an energy-based model, like the Boltzmann machine
- Joint-probability distribution is specified by the energy function:

$$P(\mathbf{v}=\boldsymbol{v},\mathbf{h}=\boldsymbol{h})=(1/\mathrm{Z})\exp(-\boldsymbol{E}(\boldsymbol{v},\boldsymbol{h}))$$

  – The energy function for an RBM is
  – $E(\boldsymbol{v},\boldsymbol{h})= -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{h} - \boldsymbol{v}^{\mathrm{T}}\mathrm{W}\boldsymbol{h}$
  – Z is the partition function

$$Z = \Sigma_{\boldsymbol{v}} \Sigma_{\boldsymbol{h}} E(\boldsymbol{v},\boldsymbol{h})$$

  – Since Z is intractable $P(\boldsymbol{v})$ is also intractable

But, RBM conditionals are tractable

- Although $P(\boldsymbol{v})$ is intractable,
  - Bipartite structure of RBM has special property that
    - Conditionals $P(\boldsymbol{h}|\boldsymbol{v}), P(\boldsymbol{v}|\boldsymbol{h})$ are factorial & easily computed:

$$P(\boldsymbol{h} \mid \boldsymbol{v}) = \frac{P(\boldsymbol{h},\boldsymbol{v})}{P(\boldsymbol{v})} = \frac{1}{P(\boldsymbol{v})}\frac{1}{Z}\exp\left\{\boldsymbol{b}^T\boldsymbol{v} + \boldsymbol{c}^T\boldsymbol{h} + \boldsymbol{v}^T W\boldsymbol{h}\right\} = \frac{1}{Z'}\exp\left\{\boldsymbol{c}^T\boldsymbol{h} + \boldsymbol{v}^T W\boldsymbol{h}\right\}$$
$$= \frac{1}{Z'}\exp\left\{\sum_{j=1}^{n_h} c_j h_j + \sum_{j=1}^{n_h} \boldsymbol{v}^T W_{:,j}\boldsymbol{h}_j\right\} = \frac{1}{Z'}\prod_{j=1}^{n_h}\exp\left\{\boldsymbol{c}_j\boldsymbol{h}_j + \boldsymbol{v}^T W_{:,j}\boldsymbol{h}_j\right\}$$

- Normalizing the distributions over individual binary $\boldsymbol{h}$

$$P(h_j = 1 \mid \boldsymbol{v}) = \frac{\tilde{P}(h_j = 1 \mid \boldsymbol{v})}{\tilde{P}(h_j = 0 \mid \boldsymbol{v}) + \tilde{P}(h_j = 1 \mid \boldsymbol{v})} = \frac{\exp\left\{c_j + \boldsymbol{v}^T W_{:,j}\right\}}{\exp\left\{0\right\} + \exp\left\{c_j + \boldsymbol{v}^T W_{:,j}\right\}} = \boldsymbol{\sigma}\left(c_j + \boldsymbol{v}^T W_{:,j}\right)$$

- We now express full conditional as a factorial distribution

$$P(\boldsymbol{h} \mid \boldsymbol{v}) = \prod_{j=1}^{n_h}\boldsymbol{\sigma}\left(\left(2\boldsymbol{h}-1\right)\odot(\boldsymbol{c}+W^T\boldsymbol{v})\right)_j$$

and similarly

$$P(\boldsymbol{v} \mid \boldsymbol{h}) = \prod_{j=1}^{n_v}\boldsymbol{\sigma}\left(\left(2\boldsymbol{v}-1\right)\odot(\boldsymbol{b}+W^T\boldsymbol{h})\right)_i$$

Training RBMs

- RBM admits
  - efficient evaluation and differentiation of P(v)
  - efficient MCMC sampling: block Gibbs sampling
- So it can be trained using techniques for training models with intractable partition functions: CD, SML (PCD) etc
- Compared to other undirected models in deep learning, RBM is relatively straightforward to train because $P(\boldsymbol{h}|\boldsymbol{v})$ can be computed in closed form
  - Deep Boltzmann machines are harder