Deep Generative Models

Anup Deshmukh
Seminar-2

Outline

- Boltzmann Machines for Real-Valued Data
  - Gaussian-Bernoulli RBMs
  - Mean and Covariance RBM’s
- Convolutional Boltzmann Machines
- Boltzmann Machines for Structured or Sequential Outputs

1.1 Boltzmann Machines for Real Valued Data

Non-real Images

- Treat grayscale images in the training set as defining [0,1] probability values.
- This is a common procedure for evaluating binary models on grayscale image datasets.

1.2 Gaussian-Bernoulli RBMs

The energy function

$$E(\mathbf{v},\mathbf{h}) = \sum_{i \,\varepsilon\, vis} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j \,\varepsilon\, hid} b_j h_j - \sum_{i,j} \frac{v_i}{\sigma_i} h_j w_{ij}$$

- Parabolic containment function
- Energy gradient produced by the total input to a visible input

2.1 Undirected Models of Conditional Covariance

- Motivation: It is the relationships between pixels and not their absolute values where most of the useful information in images resides.
- The Mean and Covariance RBM:
  - Hidden layer: Mean units and Covariance units
  - Gaussian RBM and Covariance RBM
- The Energy function for mcRBM model:

$$E_{\mathrm{mc}}(\boldsymbol{x}, \boldsymbol{h}^{(m)}, \boldsymbol{h}^{(c)}) = E_{\mathrm{m}}(\boldsymbol{x}, \boldsymbol{h}^{(m)}) + E_{\mathrm{c}}(\boldsymbol{x}, \boldsymbol{h}^{(c)}),$$

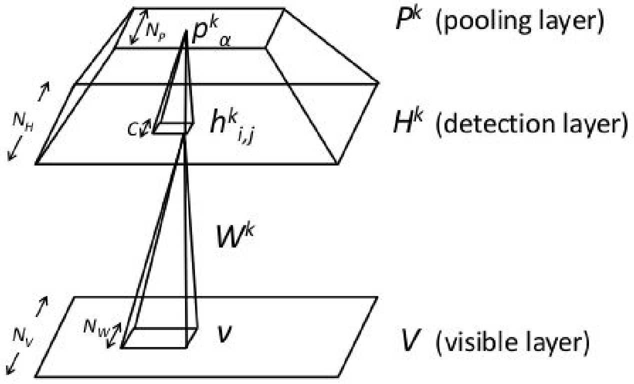
2.2 Mean and Variance RBM

$$E_{\mathrm{m}}(\boldsymbol{x}, \boldsymbol{h}^{(m)}) = \frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{x} - \sum_j \boldsymbol{x}^{\top}\boldsymbol{W}_{:,j} h_j^{(m)} - \sum_j b_j^{(m)} h_j^{(m)},$$

$$E_{\mathrm{c}}(\boldsymbol{x}, \boldsymbol{h}^{(c)}) = \frac{1}{2}\sum_j h_j^{(c)} \left(\boldsymbol{x}^{\top}\boldsymbol{r}^{(j)}\right)^2 - \sum_j b_j^{(c)} h_j^{(c)}.$$

$$p_{\mathrm{mc}}(\boldsymbol{x}, \boldsymbol{h}^{(m)}, \boldsymbol{h}^{(c)}) = \frac{1}{Z}\exp\left\{-E_{\mathrm{mc}}(\boldsymbol{x}, \boldsymbol{h}^{(m)}, \boldsymbol{h}^{(c)})\right\},$$

3.1 Convolutional Boltzmann Machines



The detection layer H^k is partitioned into blocks of size C x C.

For each k in {1,2,...K}, the pooling layer P^k shrinks the representation of the detection layer H^k by a factor of C along each dimension.

$$\sum_{(i,j)\in B_\alpha} h^k_{i,j} \leq 1, \quad \forall k, \alpha.$$

3.2 Convolutional Boltzmann Machines

$$
\begin{aligned}
E(\mathbf{v},\mathbf{h},\mathbf{p},\mathbf{h}') \; = \; & -\sum_{k} v \bullet (W^k * h^k) - \sum_{k} b_k \sum_{ij} h^k_{ij} \\
& -\sum_{k,\ell} p^k \bullet (\Gamma^{k\ell} * h'^{\ell}) - \sum_{\ell} b'_{\ell} \sum_{ij} h'^{\ell}_{ij}
\end{aligned}
$$

- Unary terms for each of the groups in the detection layers
- Interaction terms between V and H and between P and H'

4.1 Boltzmann Machines for Structured or Sequential Outputs

- The structured output
    - Speech synthesis task
- Sequence modeling
    - Video game and Film industry

Questions/Comments?

Thank You