Helmholtz Machine & The ML criterion

Dong Wang

2019/11/11

The ML criterion

- Maximum likelihood the data, with a supposed model
- Very good principle

Difficulties with hidden variable

- Gaussian mixture model
- An EM algorithm, by computing posterior of hidden
- An energy idea: Helmholtz free energy

$$E_\alpha(\theta, d) = -\log p(\alpha|\theta)p(d|\alpha,\theta) \qquad P_\alpha(\theta, d) = \frac{p(\alpha|\theta)p(d|\alpha,\theta)}{\sum_{\alpha'} p(\alpha'|\theta)p(d|\alpha',\theta)} = \frac{e^{-E_\alpha}}{\sum_{\alpha'} e^{-E_{\alpha'}}}$$

$$\log p(d|\theta) = -\left[\sum_\alpha P_\alpha E_\alpha - \left(-\sum_\alpha P_\alpha \log P_\alpha\right)\right]$$

Difficulties with hidden variable

- How if the posterior not possible?
- Sampling by Markov chain
- Variational lower-bound by simple function (mean field)

$$\begin{aligned} \log p(d|\theta) &= -\sum_\alpha Q_\alpha E_\alpha - \sum_\alpha Q_\alpha \log Q_\alpha + \sum_\alpha Q_\alpha \log[Q_\alpha/P_\alpha] \\ &= -F(d;\theta,Q) + \sum_\alpha Q_\alpha \log[Q_\alpha/P_\alpha] \end{aligned}$$

Helmholtz Machine & wake-sleep method

- Variational by complex function

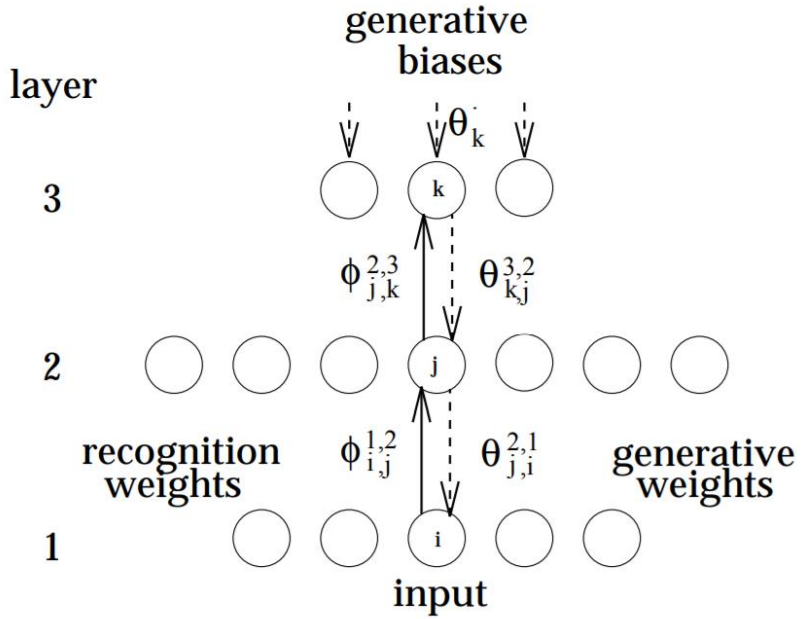

Figure 3: A Helmholtz Machine.
A simple three layer Helmholtz machine modelling the activity of 5 binary inputs (layer 1) using a two stage hierarchical model. Generative weights (θ) are shown as dashed lines, including the generative biases, the only such input to the units in the top layer. Recognition weights (ϕ) are shown with solid lines. Recognition and generative activation functions are described in the text.

Connected to AE/VAE and flow

- Maximum likelihood with two path

Is ML sufficient?

- Seems not, in particular with a compact assumption like Gaussian
- More criterion is important for particular tasks, e.g., maximum mutual information, high-order decorrelation, minimum entropy, no-gaussianlity…
- Pons and Pros for Gaussian assumption