Lecture FYS4411, April 5, 2024

FYS4411/9411, APRIL 5, 2024

Generative ML & Energy based model (AKA Boltzmann machines)

– Neural network

$$NN(x;\theta) = \psi_T(x;\theta) \Rightarrow$$

Probability

$$P(x;\theta) \propto |NN(x;\theta)|^2$$

$$\mathbb{E}[E(x;\theta)] ; \nabla_\theta \mathbb{E}[E(x;\theta)] = 0$$

$$\arg\min_{\theta \in \mathbb{R}^p} \mathbb{E}\left[E(x;\theta)\right]$$

software : NetKet (G. Carleo)
Science 2017

– NN with Boltzmann distribution

$$|\psi(x;\theta)|^2 = p(x;\theta) = \frac{e^{-\beta \tilde{E}(x;\theta)}}{Z}$$

Model : $\tilde{E}(x;\theta)$

$\tilde{E}(x;\theta)$ Not the energy of the system

$\tilde{E}(x,h;\theta) =$

(i) Binary-Binary Boltzmann machine

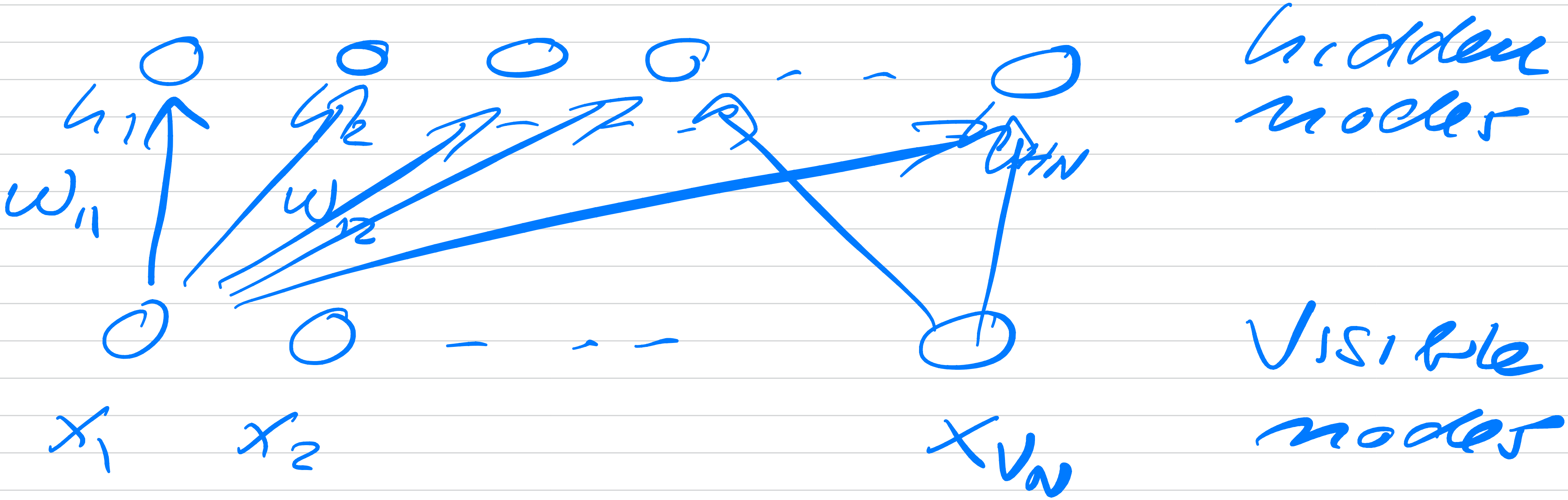
visible nodes

$$\tilde{E}(x,h;\theta) = \sum_{i=1}^{V_N} a_i x_i + \sum_{j=1}^{H_N} b_j h_j$$

$$+ \sum_{ij}^{V_N H_N} w_{ij} x_i h_j$$

hidden nodes

Neural Network (RBM)



$N = 2$ particles in 2D $\Rightarrow$

$$V_N = 4$$

$$\left.\begin{array}{l} a_i,\ b_j = \text{biases} \\ w_{ij} = \text{weights} \end{array}\right\} \text{parameter } \Theta$$

$$x_i' = \{-1, 1\} \qquad h_j' = \{-1, 1\}$$

(ii) Gaussian-Binary

$$\tilde{E}(x, h; \theta) = \sum_{i=1}^{V_N} \frac{(x_i' - a_i)^2}{\sigma_i^2} + \sum_{j=1}^{H_N} b_j h_j' + \sum_{ij}^{V_N H_N} w_{ij}' x_i' h_j'$$

binary

Technicality

$$p(x, h; \theta) = \frac{e^{-\beta E(x, h; \theta)}}{Z} \quad \text{?, 1}$$

$$Z = \begin{cases} \int dx \int dh \; e^{-\beta E(x, h; \theta)} \\ \sum_{x_i \in D} \sum_{h_j} e^{-\beta E(x_i; h_j; \theta)} \end{cases}$$

we want

Marginal distribution

$$|\psi_T(x; \theta)|^2 = p(x; \theta)$$

$$P(x_i;\theta) = \sum_{h_j} P(x_i, h_j; \theta)$$

From project 1

$$p(x,h;\theta) = \frac{e^{-\tilde{E}(x,h;\theta)}}{Z}$$

$$p(x,h;\theta) = \prod_{i=1}^{V_N} \prod_{j=1}^{H_N} p(x_i, h_j; \theta)$$

$$\arg\max_{\theta \in \mathbb{R}^P} \log p(x,h;\theta)$$

Minimize $\mathbb{E}[E(x;\theta)]$

$$= \int dx\, p(x;\theta)\, E_L(x;\theta)$$

$$\nabla_\theta \mathbb{E}[E(x;\theta)] = 0$$

$$\frac{\partial \mathbb{E}[E]}{\partial \theta_i} = 2\Big[\Big\langle E_L \frac{1}{\psi_T}\frac{\partial \psi_T}{\partial \theta_i}\Big\rangle$$

$$- \langle E_L\rangle \Big\langle \frac{1}{\psi_T}\frac{\partial \psi_T}{\partial \theta_i}\Big\rangle\Big]$$

$$\frac{1}{\Psi_T}\frac{\partial \Psi_T}{\partial \theta_i} = \frac{\partial \ln \Psi_T}{\partial \theta_i}$$

Gaussian - binary

$$\tilde{E}(x, h; \theta) = \sum_{i=1}^{VN} \frac{(x_i - a_i)^2}{\sigma_i^2} + \sum_{j=1}^{HN} b_j h_j$$

$\sigma_i^2 = 1$

parameter

$$+ \sum_{ij}^{VNHN} w_{ij} h_j x_i$$

Can assume

$$|\psi_T(x_j;\theta)|^2 = \frac{e^{-\tilde{E}(x_j\theta)}}{Z} = p(x_j\theta)$$

$$\psi_T(x_j;\theta) = \sqrt{p(x_j;\theta)}$$

$$= \frac{1}{\sqrt{Z}} \sqrt{\sum_{\{h_j\}} e^{-E(x_j, h_j;\theta)}}$$

$h_j = \{-1, 1\}$ or $\{0, 1\}$

$$\ln \Psi_T(x;\theta) =$$

$$-\frac{1}{2}\ln Z - \sum_{i=1}^{V_N}(x_i' - a_i')^2$$

$$+\frac{1}{2}\sum_{j=1}^{V_H}\ln\left(1+e^{b_j + \sum_{i=1}^{V_N} x_i' w_{ij}'}\right)$$

in the optimization we need

$$\frac{\partial \ln \Psi_T}{\partial a_i'} \; ; \; \frac{\partial \ln \Psi_T}{\partial b_j'} \; ; \; \frac{\partial \ln \Psi_T}{\partial w_{ij}'}$$

$$E_L = \frac{1}{\Psi_T} H \Psi_T$$

$$E_L = \frac{1}{2} \sum_{i=1}^{n} \left(-\left[\frac{\partial \ln \Psi_T}{\partial x_i} \right]^2 - \frac{\partial^2}{\partial x_i^2} \ln \Psi_T + \omega^2 r_i^2 \right)$$

$$H = \sum_{i=1}^{n} \left(-\frac{1}{2} \nabla_i^2 + \frac{1}{2} \omega^2 r_i^2 \right) + \sum_{i<j} v(r_{ij})$$

$$+ \sum_{i<j} v(r_{ij})$$