FYS4411/9411, APRIL 4, 2025

Generative ML and Energy based models (AKA Boltzmann machines)

we want to train a PDF

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

$$\hat{\theta} = \underset{\theta}{\arg\max}\; p(x;\theta)$$

- Network

  - Boltzmann machine
    (Restricted $\Rightarrow$ RBM)

$$|\Psi_T(x;\theta)|^2 \sim p(x;\theta)$$

optimization

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

$$p(x;\theta) \rightarrow p(x,h;\theta)$$

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

- Neural network

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

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

(FYS4429)

optimization

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

$$\theta = \{ w, b \}$$

weights biases

Boltzmann machine

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

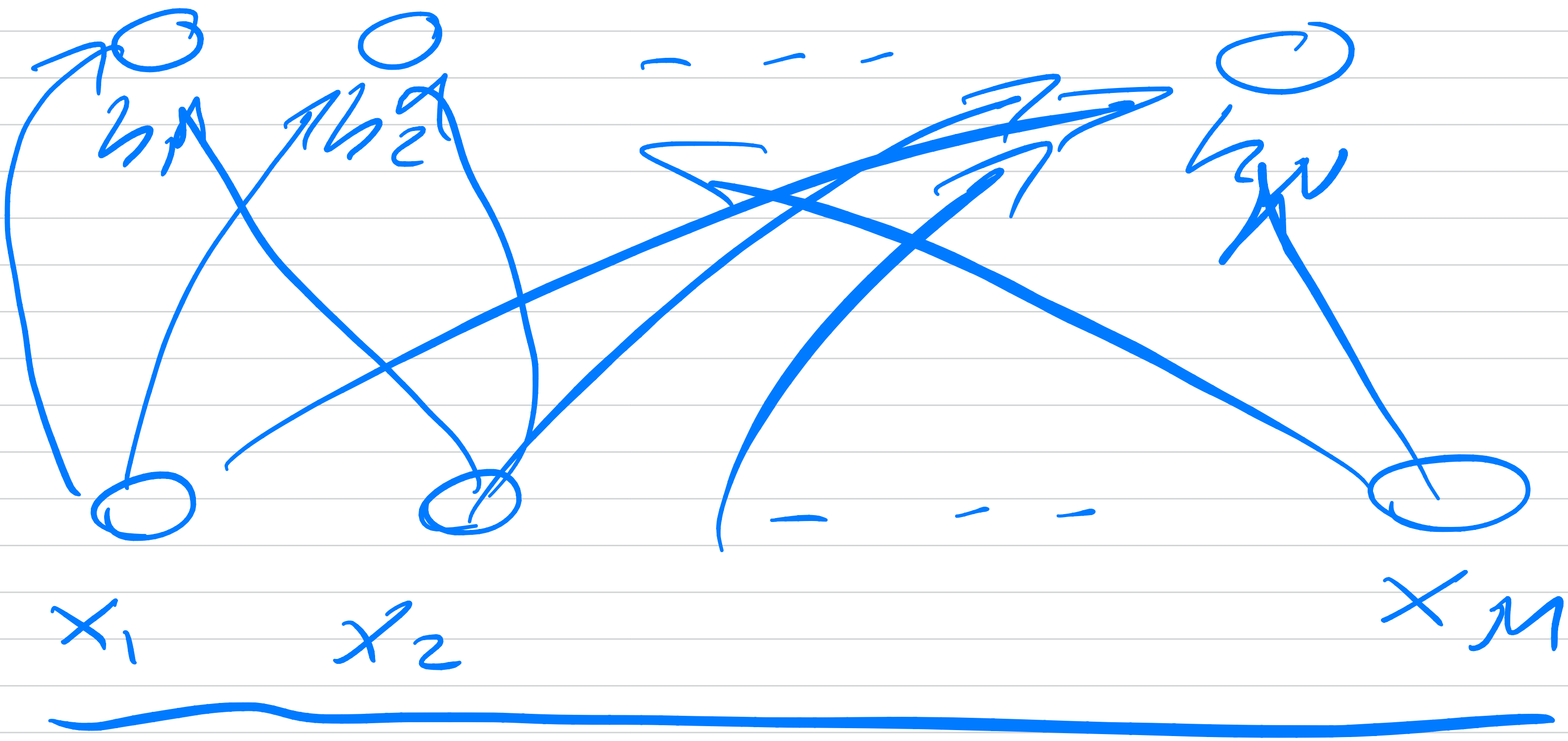
Model for a network

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

RBM - example :

RBM

hidden layer



input layer

$$\tilde{E}(x, h; \theta) = \sum_{i=1}^{M} a_i x_i + \sum_{j=1}^{N} b_j h_j + \sum_{i,j}^{MN} x_i W_{ij} h_j$$

(Binary-binary

$x_i = \{0, 1\}$

$h_j = \{0, 1\}$

biases

weights

$$\theta = \{ W, a, b \}$$

$M = 10 \quad N = 10$

$$\tilde{E}(x, h; \theta) = a^T x + b^T h + x^T W h$$

$\boxed{\text{Gaussian - binary}}$ continuous

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

(H/O : $e^{-\alpha x^2}$

Technicalities

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

normalization or partition function.

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

Example: Binary-Binary

$M = 10$ $\qquad\qquad$ $N = 10$ $\qquad$ $2^N = 1024$

$\downarrow$

sum over x_i' $= 2^M = 1024$

we want

$$|\psi_T(x; \theta)|^2 \propto \underbrace{p(x; \theta)}_{\text{marginal probability}}$$

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

How can we use project 1 in project 2 with RBMs?

Intermezzo: standard BM approach.

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

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

marginality probabilities

$$p(x;\theta) = \sum_{h_j} p(x, h_j; \theta)$$

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

conditional probabilities

$$p(x \mid h;\theta) = \frac{p(x, h;\theta)}{p(h;\theta)}$$

$$p(h/x;\theta) = p(x,h;\theta)/p(x;\theta)$$

We want to optimize

$$p(x;\theta) = \frac{1}{Z(\theta)} \prod_{x_i \in D} f(x_i;\theta)$$

$$f(x_i;\theta) = e^{-E(x_i;\theta)}$$

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^p}{\arg\max} \log p(x;\theta)$$

$$\nabla_\theta \log p(x;\theta) = 0$$

$$\nabla_\theta \left[\sum_{x_i} \log f(x_i;\theta) \right]$$

$$- \left(E \left[\nabla_\theta \log f(x_i;\theta) \right] \right) = 0$$

Binary-Binary

$$P(x;\theta) = \frac{1}{Z(\theta)} e^{a^T x} \prod_{j=1}^{N} \left(1+e^{b_j + x^T w_{*j}}\right)$$

$$f(x_i, h_j;\theta) = a^T x + b^T h + x^T W h$$

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

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

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

$$\frac{\partial \mathbb{E}[E]}{\partial \theta_j} = 2\Bigg[\left\langle E_L \frac{1}{\psi_T}\frac{\partial \psi_T}{\partial \theta_i}\right\rangle$$

$$- \langle E_L \rangle \left\langle \frac{1}{\psi_T}\frac{\partial \psi_T}{\partial \theta_i} \right\rangle\Bigg]$$

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

Model in RBM

continuous

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

binary

parameters

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

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

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

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

$$\text{or } \{0, 1\}$$

$$\ln \Psi_T(x;\theta) = -\frac{1}{2}\ln Z - \sum_{i=1}^{M}(x_i - a_i)^2 + \frac{1}{2}\sum_{j=1}^{N}\ln\left(1 + e^{b_j + \sum_{i=1}^{M} x_i w_{ij}}\right)$$

in the optimization we need

$$\frac{\partial \ln \Psi_T}{\partial a_i}; \quad \frac{\partial \ln \Psi_T}{\partial b_j};$$

$$\frac{\partial \ln \Psi_T}{\partial w_{ij}}$$

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

$$r_{ij} = |\vec{r}_i - \vec{r}_j| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

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

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

$$\frac{1}{2}\sum_{i=1}^{n} w^2 v_i^2 + \sum_{i<j} v(v_i v_j)$$

Binary - Binary

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

$$= -(a^T x + b^T h + x^T W h)$$

$$v_i = \{0, 1\} \qquad h_j = \{0, 1\}$$

$$Z(\theta) = \sum_{x h} e^{+(a^T x + b^T h + x^T W h)}$$

$$P(x;\theta) = \frac{1}{Z(\theta)} \sum_h e^{a^T x + b^T h + x^T W h}$$

$$= \frac{e^{a^T x}}{Z(\theta)} \sum_h e^{b^T h + x^T W h}$$

$$= \frac{e^{a^T x}}{Z(\theta)} \sum_h e^{\sum_{j=1}^{N} (b_j + x^T W_{*j}) h_j}$$

$$= \frac{e^{a^T x}}{Z(\theta)} \prod_{j=1}^{N} e^{(b_j + x^T W_{*j}) h_j}$$

$$= \frac{1}{Z(\theta)} e^{a^T x} \left(\sum_{h_1 = \{0,1\}} e^{(b_1 + x^T W_{*1}) h_1} \right)$$

$$\times \left(\sum_{h_2 = \{0,1\}} e^{(b_2 + x^T W_{*2}) h_2} \right)$$

$$\vdots$$

$$\times \underbrace{\left(\sum_{h_N = \{0,1\}} e^{(b_N + x^T W_{*N}) h_N} \right)}_{1 + e^{(b_N + x^T W_{*N})}}$$