FYS5429/9429, lectures March 27

FYS5429/9429 March 27

i.i.d = independent and identically distributed

Domain of events

$$X = \{x_0\ x_1 \ldots\ x_{n-1}\}$$

$$P(x; \theta) = \prod_{i=0}^{n-1} P_X(x_i; \theta)$$

parameters to optimize

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ P(X;\theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ \log P(X;\theta)$$

or

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\left(-\log P(X;\theta)\right)$$

Our first encounter:

Boltzmann distribution

$\Longrightarrow$ ENERGY MODELS

$$P(E_i ; \beta) = \frac{e^{-\beta E_i}}{Z(\beta)}$$

energy

inverse temp $1/k_B T$

normalization constant / partition function

$$Z(\beta) = \sum_{E_i} e^{-\beta E_i}$$

Example

$$E(x_i;\theta) \longrightarrow E(x,h;\theta)$$

$$x_i = \{0,1\} \qquad h_i = \{0,1\}$$

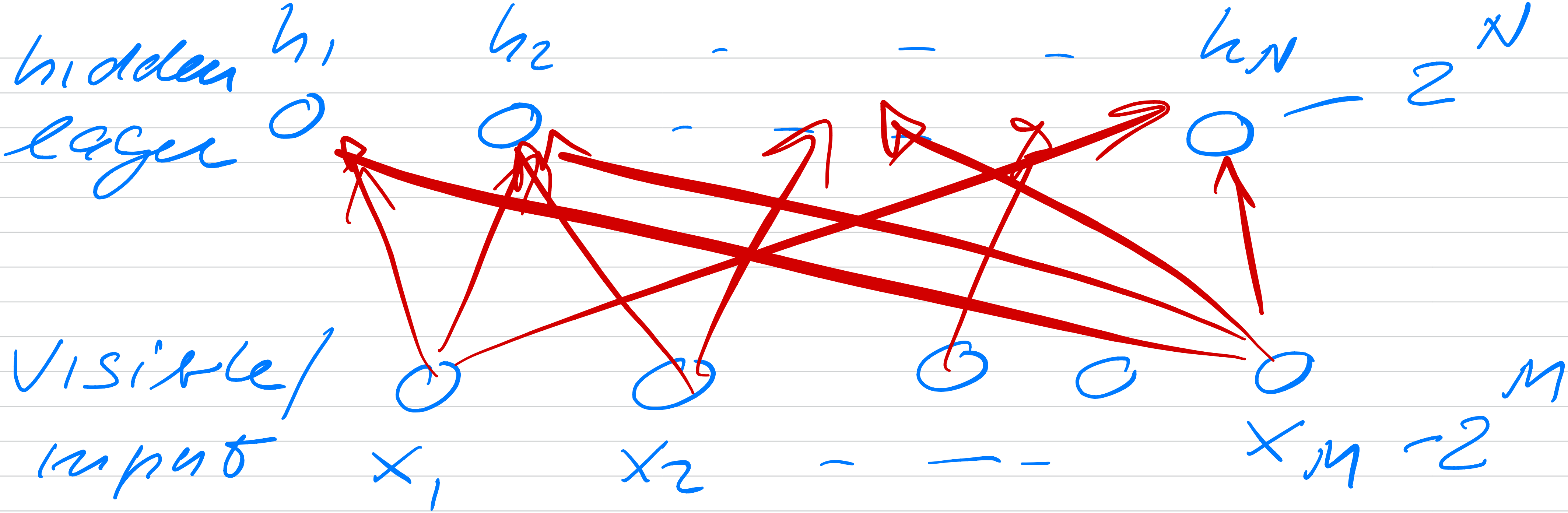
$$E(x,h;\theta) = \sum_{i=1}^{M} x_i a_i + \sum_{j=1}^{N} b_j h_j$$

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

weight

biases

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



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

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

$M = 2 \Rightarrow X: \{0,0\}, \{0,1\},$

$\{1,0\}, \{1,1\}$

Total configs = $2^M = 4$

$N = 2 \Rightarrow h: \{0,0\}, \{0,1\}$

$\{1,0\} \quad \{1,1\}$

Total config $2^N = 4$

Marginal probability

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

$$f(x;\theta)$$

$$p(x;\theta) \geq 0$$

$$\hat{\theta} = \arg\max_{\theta} \log p(x;\theta)$$

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

$$P(x;\theta) = \prod_{x_i' \in D} P(x_i';\theta)$$

$$\log P(x;\theta) = \sum_{x_i'} \log P(x_i';\theta)$$

$$= \sum_{x_i'} \log f(x_i;\theta)$$

$$- \log Z(\theta)$$

Take derivative:

$$\nabla_\theta \log Z(\theta) = \frac{\nabla_\theta Z(\theta)}{Z(\theta)}$$

$$= \frac{\nabla_\theta \sum_{x_i'} f(x_i'; \theta)}{Z(\theta)}$$

$$= \frac{\sum_{x_i'} \nabla_\theta f(x_i'; \theta)}{Z(\theta)}$$

$$f(x_i;\theta) = \exp\left[\log f(x_i;\theta)\right]$$

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

$$= \frac{\sum_{x_i} \underbrace{\exp(\log f(x_i;\theta))}_{f(x_i;\theta)} \nabla_\theta \log f(x_i;\theta)}{Z(\theta)}$$

$$p(x_i;\theta) = \frac{f(x_i;\theta)}{Z(\theta)}$$

$$= \sum_{x_i \in D} p(x_i;\theta) \nabla_\theta \log f(x_i;\theta)$$

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

optimization part

$$\nabla_\theta P = 0 = \nabla_\theta \log f(x;\theta)$$

$$- \boxed{\mathbb{E}\left[\nabla_\theta \log f(x_i;\theta)\right]}$$

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

$$\simeq \frac{1}{M}\sum_{i=1}^{M} \nabla_\theta \log f(x_i;\theta)$$

solved with Markov chain Monte Carlo sampling (MC²)

- Markov chain
- sampling
  - Metropolis
  - Gibbs
  - Langevin

Markov chains

$$p(x_i; t) \Rightarrow P_i(t)$$

$$\sum_{i \in D} P_i(t) = 1$$

probability for making a transition

$$W_{ij} \rightarrow W(j \rightarrow i)$$

$$\sum_j W_{ij} = 1 \,,$$ stochastic matrix

$$|\lambda_0| > |\lambda_1| \; - \; - \; > |\lambda_{M-1}|$$

$$\lambda_0 = \underline{1}$$

$$P_i(t+1) = \sum_j W(j \to i) P_j(t)$$

$$P(t+1) = W P(t)$$

$$P(t+1) = P(t) = P$$

stationary state.

$$W v_i = \lambda_i v_i$$

$$P_i(t=0) = \sum_j \alpha_j v_j$$

$$\alpha_j = \langle P_i(t=0) | v_j \rangle$$

$$W P_i(t=0) = \sum_j \alpha_j W v_j =$$

$$\sum_j \alpha_j \lambda_j v_j$$

$$p_i(t+1) = W p_i(t=0)$$
$$\overset{\shortparallel}{0}$$

if we do this τ-steps

$$p_i(\tau) = W^{\tau} p_i(t=0)$$
$$= \sum_j \alpha_j \lambda_j^{\tau} v_j$$
$$= \alpha_0 v_0 \underset{\underset{1}{\shortparallel}}{\lambda_0^{\tau}} + \sum_{j \neq 0} \alpha_j \lambda_j^{\tau} v_j$$

$$\lim_{x \to \infty} P_1'(x) = P_1' = \kappa_0 \sigma_0$$