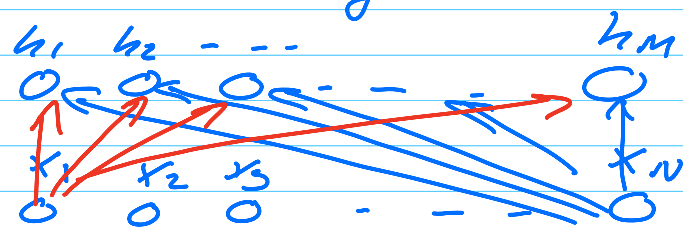
FYS5429/9429, APRIL 26, 2023

RBM

one hidden layer



visible layer

$$\Theta = \{W, b, c\}$$

$x = \{x_1 x_2 \dots x_N\}$

$h = \{h_1 h_2 \dots h_M\}$

$$p(x, h; \Theta) = \frac{\tilde{p}(x, h; \Theta)}{Z(\Theta)}$$

we are interested in the marginal probabilities

$$p(x; \Theta) = \frac{\tilde{p}(x; \Theta)}{Z(\Theta)}$$

$$\tilde{p}(x; \Theta) = \sum_h \tilde{p}(x, h; \Theta)$$

$$Z(\Theta) = \sum_{xh} \tilde{p}(x, h; \Theta)$$

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

$$\beta = 1$$

marginal distribution for h

$$p(h; \theta) \propto \sum_x \tilde{p}(x, h; \theta)$$

we need also conditional probabilities

$$p(x \mid h), \quad p(h \mid x)$$

Maximizing $p(x, h; \theta)$ is done by maximizing

$$\hat{\theta} = \arg\max_{\theta \in R^n} \log p(x; \theta)$$

$$\Rightarrow \vec{\nabla}_\theta \log p(x; \theta) =$$

$$\nabla_\theta \log \tilde{p}(x; \theta) -$$

$$\nabla_\theta \log Z(\theta)$$

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

$$= \sum_x \frac{\nabla_\theta \tilde{p}(x; \theta)}{Z(\theta)}$$

$$\left(\tilde{p}(x) > 0 \right)$$

$$= \sum_x \frac{\nabla_x \exp(\log \tilde{p})}{Z}$$

$$= \sum_x \frac{\exp(\log \tilde{p}(x)) \nabla_\theta \log \tilde{p}(x)}{Z}$$

$$= \sum_x \frac{\tilde{p}(x) \nabla_\theta \log \tilde{p}(x)}{Z}$$

$$= \sum_x p(x) \nabla_\theta \log \tilde{p}(x)$$

$$\left(\mathbb{E}[x^n] = \sum_x p(x) x^n \right)$$

$$= \mathbb{E}\left[\nabla_\theta \log \tilde{p}(x) \right]$$