Machine Learning - 2, Lecture 6

Prob. Given a dataset $S = \{v_1, v_2, \dots, v_N\}$, $v_i \in \{0,1\}^{m \times 1}$

Learn the data distribution $P_\theta(v)$ using these samples.

$$\max_\theta \log P_\theta(v) = \max_\theta \log\left(\sum_h P_\theta(v,h)\right) \qquad P(V) = \sum_h P(v,h)$$

Restricted Boltzmann Machine:

Parametrization of a distribution:

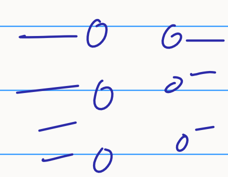
$$1 \geq P_\theta(x) \geq 0, \qquad \int_x P_\theta(x)\,dx = 1$$

$$x \rightarrow \boxed{f_\theta(x)} \rightarrow f_\theta(x)$$

$$P_\theta(x) = \frac{e^{-f_\theta(x)}}{Z} \qquad Z = \int f_\theta(x)\,dx$$

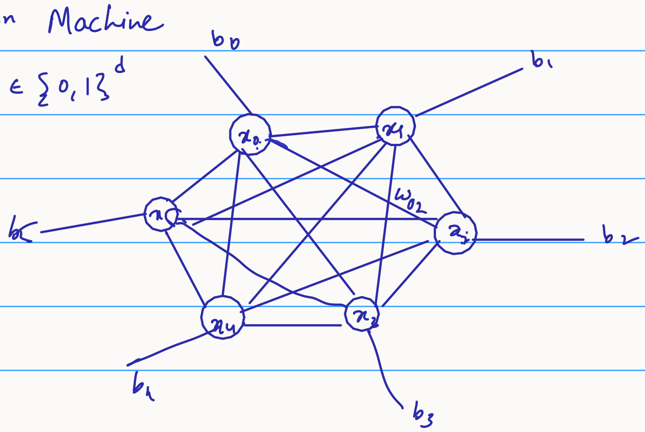
Energy Based Modes

$$P_\theta(x) = \frac{e^{-E_\theta(x)}}{Z} \qquad Z = \int e^{-E_\theta(x)}\,dx$$



Boltzmann Machine

$x \in \{0,1\}^d$

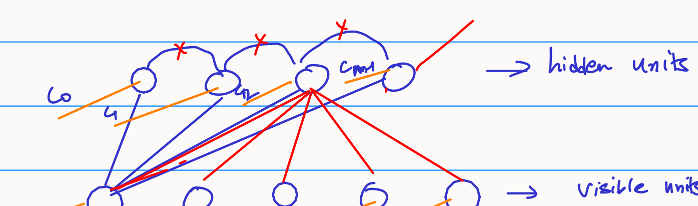


$$E(x) = -x^T W x - b^T x$$

$$= \sum_{i=0}^{5} \sum_{j=0}^{5} x_i x_j w_{ij} - \sum_{i=0}^{5} b_i x_i$$

$$P(x) = \frac{e^{-E(x)}}{Z}$$

$$x^T W x = [v^T\ h^T] \begin{bmatrix} U & W \\ W & R \end{bmatrix} \begin{bmatrix} v \\ h \end{bmatrix}$$

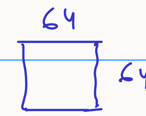


~~Restrict~~ We do not allow intra-layer connections

$$X = (v, h)$$

$$P(x) = \underbrace{P(v, h)} = \frac{e^{-E(v,h)}}{Z}$$

$$\underbrace{Z} = \sum_{v \in \{0,1\}^m} \sum_{h \in \{0,1\}^n} e^{-E(v,h)}$$



$$m = 64 \times 64$$

$$\underbrace{2^{(64 \times 64)}}$$

$$2^m \quad 2^n \rightarrow 2^{m+n} \;\&\; 2^{m+n} / \exp(\cdot)$$

$$E(v,h) = -v^T W h - b^T v - c^T h \underbrace{\cancel{\underbrace{-v^T U v}} \cancel{\underbrace{- h^T R h}}}_{\text{only in BM}}$$

$$p(h|v) = \frac{P(v,h)}{P(v)} = \frac{e^{-E(v,h)}}{Z\, P(v)}$$

$$P(v) = \sum_h P(v,h)$$

$$= \frac{e^{-E(v,h)}}{Z \sum_h P(v,h)}$$

$$= \frac{e^{-E(v,h)}}{\cancel{Z} \sum_h \frac{e^{-E(v,h)}}{\cancel{Z}}}$$

$$= \frac{e^{-E(v,h)}}{\sum_h e^{-E(v,h)}}$$

$$= \frac{e^{v^T W h + \underbrace{b^T v} + c^T h}}{\sum_h e^{+v^T W h + \underbrace{b^T v} + c^T h}}$$

$$= \frac{e^{v^T W h + c^T h}}{\sum_h e^{v^T W h + c^T h}}$$

$$\underbrace{v^T W h} = \sum_{i=0}^{n+} v^T W_i h_i$$

$$= \frac{e^{\sum_i v^T W_i h_i + c_i h_i}}{\sum_h e^{\sum_i v^T W_i h_i + c_i h_i}}$$

$$e^{a+b+c} = e^a e^b e^c$$

$$= \frac{\prod_i e^{+v^T w_i h_i + b_i h_i}}{\sum_h \prod_i e^{v^T w_i h_i + b_i h_i}}$$

$$= \frac{\prod_i e^{v^T w_i h_i + b_i h_i}}{\sum_{h_0 \in \{0,1\}} \sum_{h_1 \in \{0,1\}} \cdots \sum_{h_{n-1} \in \{0,1\}} e^{v^T w_0 h_0 + b_0 h_0} \cdot e^{v^T w_1 h_1 + b_1 h_1} \cdots e^{v^T w_{n-1} h_{n-1} + b_{n-1} h_{n-1}}}$$

$$= \frac{\prod_i e^{v^T w_i h_i + b_i h_i}}{\left(\sum_{h_0 \in \{0,1\}} e^{v^T w_0 h_0 + b_0 h_0}\right)\left(\sum_{h_1 \in \{0,1\}} e^{v^T w_1 h_1 + b_1 h_1}\right) \cdots \left(\sum_{h_{n-1} \in \{0,1\}} e^{v^T w_{n-1} h_{n-1} + b_{n-1} h_{n-1}}\right)}$$

$$= \frac{\prod_i e^{v^T w_i h_i + b_i h_i}}{\prod_i \left(\sum_{h_i \in \{0,1\}} e^{v^T w_i h_i + b_i h_i}\right)}$$

$$= \prod_i \frac{e^{v^T w_i h_i + b_i h_i}}{\left(\sum_{h_i \in \{0,1\}} e^{v^T w_i h_i + b_i h_i}\right)}$$

$$= \prod_i \frac{e^{v^T w_i h_i + b_i h_i}}{1 + e^{v^T w_i + b_i}}$$

$$P(h|v) = \prod_i P(h_i|v)$$

$$\Rightarrow P(h_0, h_1, h_2, \ldots, h_{n-1}|v) = P(h_0|v) P(h_1|v) \cdots P(h_{n-1}|v)$$

$\rightarrow$ given v, h_i's are independent of others

$$P(x, y|z) = P(x|z) P(y|z)$$

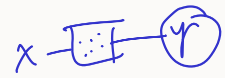
Here, $P(h_i|v) = \dfrac{e^{v^T w_i h_i + b_i h_i}}{1 + e^{v^T w_i + b_i}}$

$$\Rightarrow P(h_i = 1|v) = \frac{e^{v^T w_i + b_i}}{1 + e^{v^T w_i + b_i}}$$

$$= \sigma(v^T w_i + b_i)$$

$$P(h_i = 0|v) = 1 - \sigma(v^T w_i + b_i)$$

Gradient of LL w.r.t θ: $\theta = (W, b, c)$



$\|x - y\|^2$

$$\log P_\theta(v) = \log \frac{1}{Z} \sum_h e^{-E_\theta(v,h)}$$

$$\log P_\theta(v) = \log \sum_h e^{-E_\theta(v,h)} - \log \sum_h \sum_v e^{-E_\theta(v,h)} \quad \Big| \quad Z = \sum_h \sum_v$$

$$\frac{\partial}{\partial \theta} \log P_\theta(v) = ?$$