Unit 3 :

latent variable : hidden characterstic
for eg. you have an image of nature
if sky is blue in colour :- daytime image
if black :- night image
so this daytime and night are the latent variable

Now,
we have
observed variables $V = \{V_1, V_2, \ldots V_{1024}\})$.
and hidden variables $H = \{H_1, H_2 \ldots Hn\})$.

we will be discussing two concepts.
1. abstraction 2. generation

1 abstraction :-
i.e finding hidden variables when observed variables are given.

$$P(H|V) = \frac{P(H,V)}{\sum_{H} P(V,H)}$$

why there is need of abstraction.
for e.g. you have the image of beach.
and if you have observed pixels or variable how would you describe image :- ? I am looking at an image with pixel 1 is blue pixel 2 is blue and so on but we would not understood what this means.
so we are finding latent variable so we can say that I am looking at image of sunny beach with sand

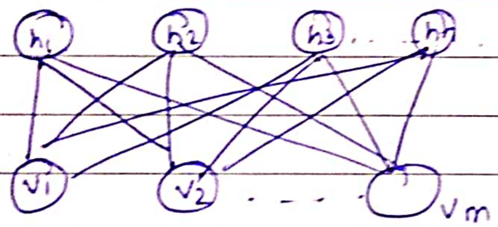
latent variable.

generation :-
knowing latent variable find hidden variable

$$P(V|H) = \frac{P(V,H)}{\Sigma_V P(V,H)}.$$

Now we have to calculate value of $P(V,H)$.
For that :
we are assuming few things
i.e. all our variables take booleam value
The vector $V = \{0,1\}^m$ total 2^m.
hidden $H = \{0,1\}^n$ total 2^n.

Restricted Boltzman machine



- we have edges between hidden, visible variable.
- we do not have edges between (hidden - hidden) or (visible-visible).

- joint probabily can be written as product of factors
  max clique = $m \times n$

$$P(V,H) = \frac{1}{Z} \prod_i \prod_j \phi_{ij}(v_i, h_j)$$

But here we are assuming that v_i and h_i have their own clique

v_i h_i

so $P(V,H) = \frac{1}{Z} \prod_i \prod_j \phi_{ij}(v_i h_j) \prod_i \psi_i(v_i) \prod_j \xi_j(h_j)$

Z is partition function and is given by

$$\sum_{V} \sum_{H} \prod_{i} \prod_{j} \phi_{ij}(v_i, h_j) \prod_{i} \psi_i(v_i) \prod_{j} \xi_j(h_j)$$

Que ppt. p.g. 42/71

$$P(V = \langle 0,0,0 \rangle, H = \langle 1,1 \rangle)$$

$$= \frac{1}{Z} \phi_{11}(0,1)\, \phi_{12}(0,1)\, \phi_{21}(0,1), \phi_{22}(0,1)\, \phi_{31}(0,1)\, \phi_{32}(0,1)\, \psi_1(0)\, \psi_2(0)\, \psi_3(0)\, \xi_1(1)\, \xi_2(1).$$

$$= \frac{1}{Z}\, 3 \times 20 \times 3 \times 1 \times 3 \times 1 \times 30 \times 100 \times 1 \times 1 \times 10.$$

Aur isme Z jaise value nikale then unit 2 me.
to table aisa banega

v_1	v_2	v_3	h_1	h_2
0	0	0	0	0
0	0	0	0	1
0	0	0	1	0

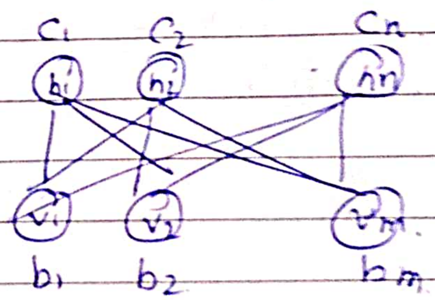
aise.

specific parametric form chosen by RBMs is

$$\phi_{ij}(v_i, h_j) = e^{w_{ij} v_i h_j}$$

$$\psi_i(v_i) = e^{b_i v_i}$$

$$\xi(h_j) = e^{c_j h_j}$$



we have to calculate joint probability

$$P(V,H) = \frac{1}{Z} \prod_i \prod_j \phi\, \omega_{ij} v_i h_j \prod_i \varphi_i(v_i) \prod_j \xi_j(h_j).$$

$$= \frac{1}{Z} \underbrace{\prod_i \prod_j e^{\omega_{ij} v_i h_j} \prod_i e^{b_i v_i} \prod_j e^{c_j h_j}}$$

$$e^{x_1} e^{x_2} e^{x_3} = e^{x_1 + x_2 + x_3}.$$

$$= \frac{1}{Z} e^{\sum_i \sum_j \omega_{ij} v_i h_j} e^{\sum b_i v_i} e^{\sum c_j h_j}$$

$$= \frac{1}{Z} e^{\overbrace{\sum_i \sum_j \omega_{ij} v_i h_j + \sum b_i v_i + \sum c_j h_j}^{\downarrow\, -E(V,H)}}$$

$$= \frac{1}{Z} e^{-E(V,H)}.$$

$$E(V,H) = -\sum_i \sum_j \omega_{ij} v_i h_j - \sum b_i v_i - \sum c_j h_j$$

Because of network restriction which we assumed it it is known as Boltzman machine.

And this term comes from stastical mechanic. which as eq^n.

$$F(state) \propto e^{\frac{-E}{kt}}.$$

which is called as Boltzman distribution or Gibbs distribution.

RBMs as stocastic neural networks

Connecting RBM to neural network

Let us take l-th visible variable

$P(V_l = l \mid H)$

so we will define V_{-l} as the state of all visible units except the l th unit.

$$\alpha_l(H) = -\sum_{i=1}^{n} w_{il} h_i - b_l$$

$$\beta(V_{-l}, H) = -\sum_{i=1}^{n} \sum_{j=1, j \neq l}^{m} w_{ij} h_i v_j - \sum_{i=1, j \neq l}^{m} b_i v_i - \sum_{i=1}^{n} c_i h_j$$

$$E(V, H) = V_l \alpha(H) + \beta(V_{-l}, H)$$

$$P(V_l = 1 \mid H) = P(V_l = 1 \mid V_{-l}, H)$$

$$= \frac{P(V_l = 1, V_{-l}, H)}{P(V_{-l}, H)}$$

$$= \frac{e^{-E(V_l = 1, V_{-l}, H)}}{e^{-E(V_l = 1, V_{-l}, H)} + e^{-E(V_l = 0, V_{-l}, H)}}$$

$$= \frac{e^{-\beta(V_{-l}, H)} \cdot e^{-1\alpha_l(H)}}{e^{-\beta(V_{-l}, H)} \cdot e^{-\alpha_l(H)} + e^{-\beta(V_{-l}, H)}}$$

$$= \frac{e^{-1\alpha(H)}}{e^{-\alpha_l(H)} + 1}$$

$$= \frac{1}{1 + e^{\alpha_l(H)}} \qquad \frac{1}{1 + e^{-x}} = \sigma(e^{-x})$$

$$= \text{Sigmoid}(\alpha_l(H)) = \sigma\left(\sum_{i=1}^{n} w_{il} h_i + b_l\right)$$

$$p(v_i = 1|H) = \sigma\left(\sum_{i=1}^{n} w_{il} h_i + b_l\right).$$

similarly.

$$p(h_i = 1|v) = \sigma\left(\sum_{i=1}^{m} w_{il} v_i + c_l\right).$$

Unsupervised Learning in RBMs.

we have to find $p(x, h)$.

Objective function to be use..

$$\text{maximize} \prod_{i=1}^{N} P(X = x_i).$$

or log likelihood.

$$\ln \mathcal{L}(\theta) = \ln \prod_{i=1}^{l} p(x_i|\theta) = \sum_{i=1}^{l} \ln p(x_i|\theta).$$

where θ is parameters

Computing gradient of log likelihood.

$$\ln \mathcal{L}(\theta) = \ln p(x_i|\theta)$$

$$= \ln p(V|\theta). = \ln \frac{1}{Z} \sum_H e^{-E(V,H)}.$$

$$= \ln \sum_H e^{-E(V,H)} - \ln \sum_{V,H} e^{-E(V,H)}.$$

$$\frac{\delta \ln \mathcal{L}(\theta)}{\delta(\theta)} = \frac{\delta}{\delta\theta}\left(\ln \sum_H e^{-E(V,H)} - \ln \sum_{V,H} e^{-E(V,H)}\right).$$

$\delta \ln(x) = \frac{1}{x}$

$$= -\frac{1}{\sum_H e^{-E(V,H)}} \times e^{-E(V,H)} \frac{\delta E(V,H)}{\delta\theta} + \frac{1}{\sum_{V,H} e^{-E(V,H)}} e^{-E(V,H)} \frac{\delta E(V,H)}{\delta\theta}$$

$$= \frac{-e^{-E(V,H)}}{\sum_H e^{-E(V,H)}} \frac{\delta E(V,H)}{\delta\theta} + \frac{e^{-E(V,H)}}{\sum_{V,H} e^{-E(V,H)}} \frac{\delta E(V,H)}{\delta\theta}$$

$$\frac{e^{-E(V,H)}}{\sum_{V,H} e^{-E(V,H)}} = P(V,H).$$

$$\frac{e^{-E(V,H)}}{\sum_{H} e^{-E(V,H)}} = \frac{\frac{1}{Z} e^{-E(V,H)}}{\frac{1}{Z}\sum_{H} e^{-E(V,H)}}.$$

$$= \frac{P(V,H)}{P(V)} = P(H|V).$$

$$\cancel{\frac{\delta \ln L(\theta)}{\delta \theta} = P(V,H) + P(H|V)}$$

$$\frac{\delta \ln L(\theta)}{\delta \theta} = -\sum_{H} p(H|V) \frac{\delta E(V,H)}{\delta \theta} + \sum_{V,H} P(V,H) \frac{\delta E(V,H)}{\delta \theta}.$$

θ is collection of all parameters in our model.
w_{ij}, b_i, c_i

now, compute partial derivation with respt to weight w_{ij}.

$$\frac{\delta L(\theta)}{\delta w_{ij}} = -\sum_{H} p(H|V) \frac{\delta E(V,H)}{\delta w_{ij}} + \sum_{V,H} p(V,H) \frac{\delta E(V,H)}{\delta w_{ij}}$$

$$= \sum_{H} p(H|V) h_i v_j - \sum_{V,H} p(V,H) h_i v_j.$$

$$= E_{p(H|V)}[v_i h_j] - E_{p(V,H)}[v_i h_j].$$

sum of two expectation.