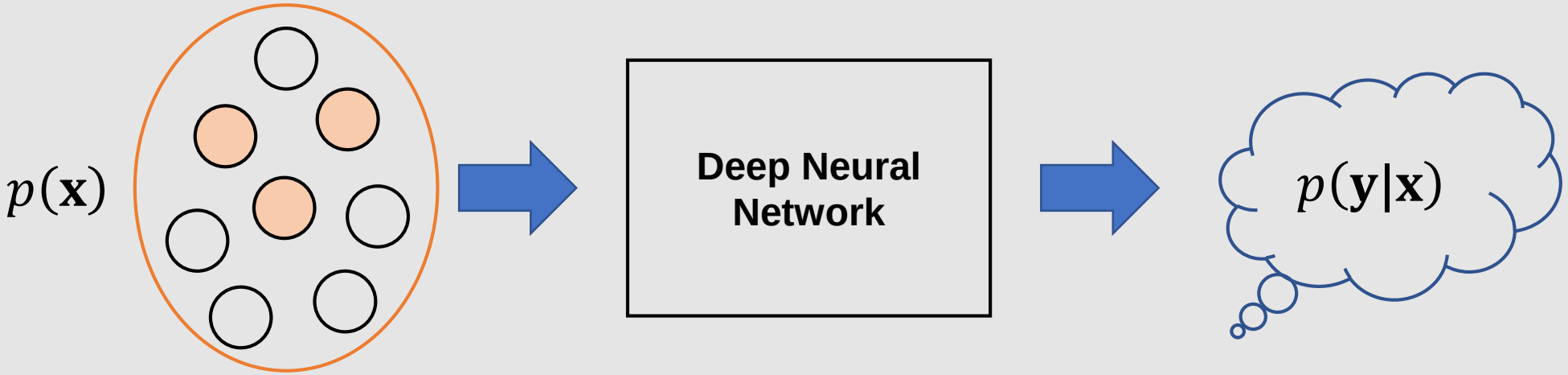
Deep Generative Models
- Boltzmann Machine

2019 MMI Deeplearning Seminar

20141403 김진범

Boltzmann Machine

- MLP(Multilayer Perceptron)의 한계점
  - Classification에 초점이 맞춰져 있다. → $p(\mathbf{y}|\mathbf{x})$



  - 특정 목적을 가진 network를 다른 task에 곧바로 적용하는 것이 힘들다.
  - 대신에 input data 자체를 학습해서 더 많은 정보를 이용하자! → $p(\mathbf{x})$

Boltzmann Machine

- Entropy
  - Entropy는 증가한다.
    → 열이 더운 곳에서 차가운 곳으로 흐른다.
    → 정확하게는 차가운 곳에서 더운 곳으로 흐르는 현상은 확률적으로 극히 낮다.
  - Entropy 자체가 확률을 나타내는 척도

  - Entropy가 증가하면 사용할 수 있는 energy는 감소한다.
  - Boltzmann의 entropy 개념을 도입하여 energy가 최소가 되는 상태가 안정적
  - Energy가 낮을수록 변수들 사이의 연관성이 높다.

Boltzmann Machine

- Energy-Based Models
  - 각 state에 대한 energy function이 최소가 되도록 학습하는 model
  - Energy-Based model에서는 확률 분포를 다음과 같이 정의

$$P(x) = \frac{e^{-E(x)}}{Z}, Z = \sum_{x} e^{-E(x)} \ (x: state)$$

  - Model의 representation을 높이기 위해 visible variable뿐만 아니라 hidden variable이 필요

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

Boltzmann Machine

- Energy-Based Models
  - **계산의 편의를 위해** free energy**를 도입**

$$P(x) = \frac{e^{-F(x)}}{Z}, Z = \sum_{x} e^{-F(x)}$$

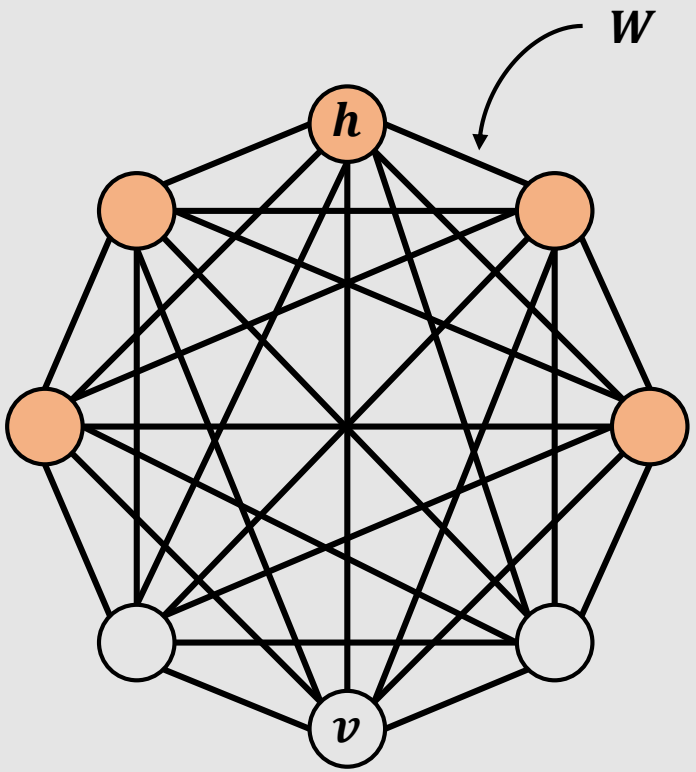
$$F(x) = -\log \sum_{h} e^{-E(x,h)}$$

  - EBM**은** data**의** log-likelihood**를 사용하여** gradient descent **방법으로 학습**

$$\theta := \theta - \rho \frac{\partial \log P(x)}{\partial \theta}$$

Boltzmann Machine

- Boltzmann Machine
  - Visible node**와** hidden(latent) node**로 구분**
  - **모든** node(Binary)**가 서로 연결**
  - Undirected graph model
  - Energy-based model



Boltzmann Machine

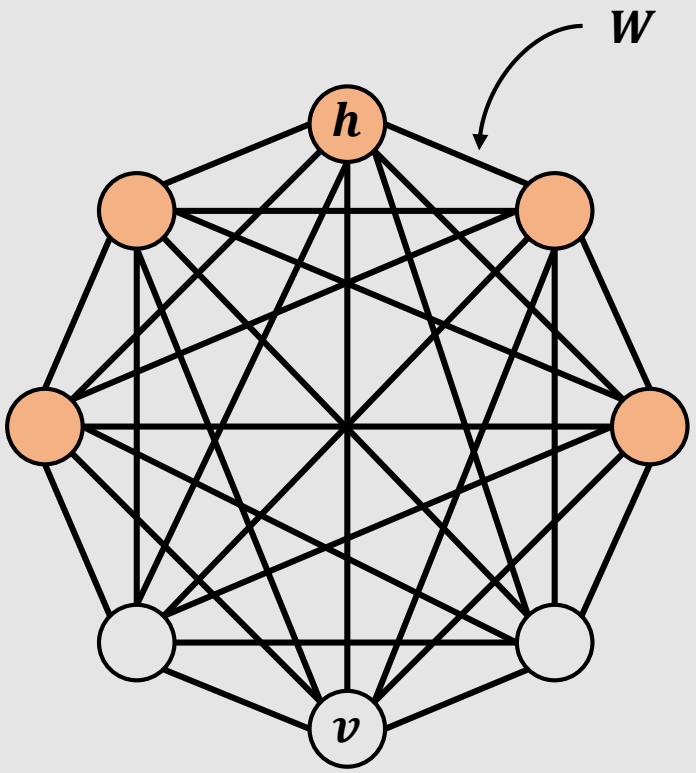
- Boltzmann Machine

$$E(\boldsymbol{v},\boldsymbol{h})$$

$$= -\boldsymbol{h}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{h}$$

$$-\frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{x} - \frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{h}$$

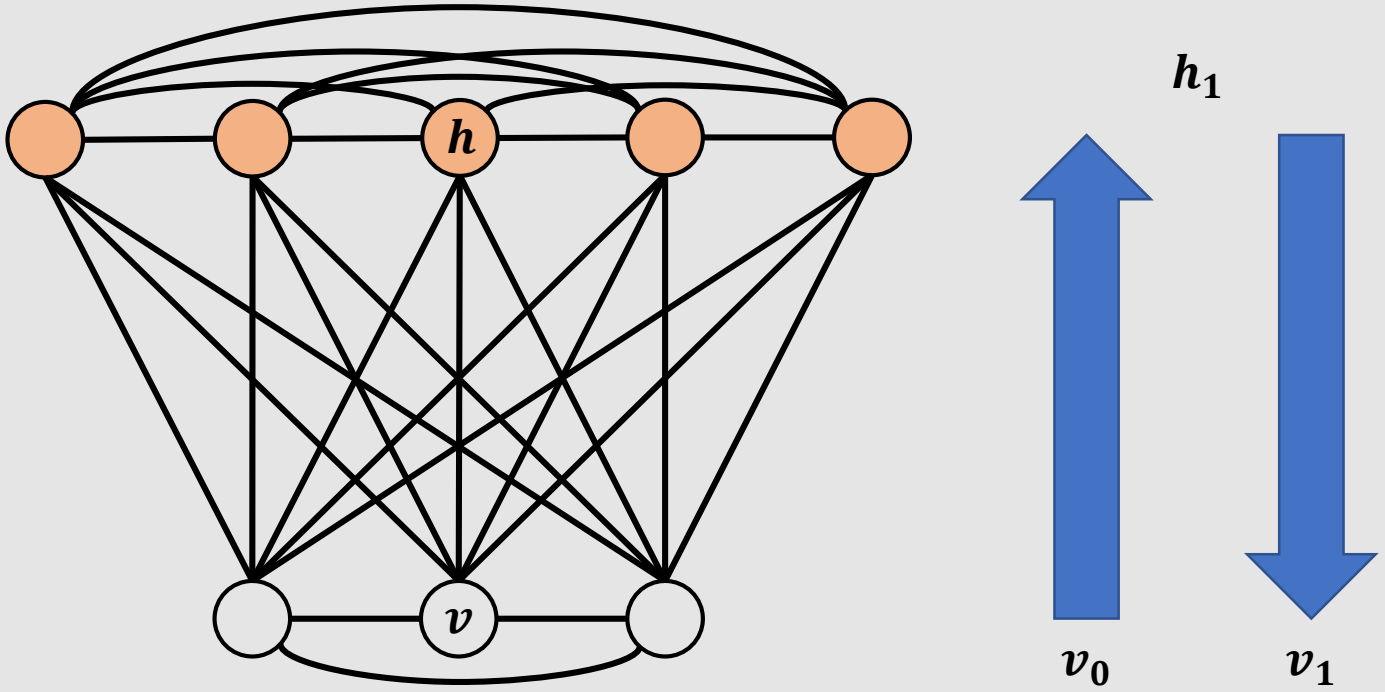
$$P(\boldsymbol{v},\boldsymbol{h}) = \frac{\exp\left(-E(\boldsymbol{v},\boldsymbol{h})\right)}{Z}, P(\boldsymbol{v}) = \sum_{\boldsymbol{h}} P(\boldsymbol{v},\boldsymbol{h})$$



  - Energy가 작을 수록 node가 존재할 확률이 높다.

Boltzmann Machine

- Boltzmann Machine.

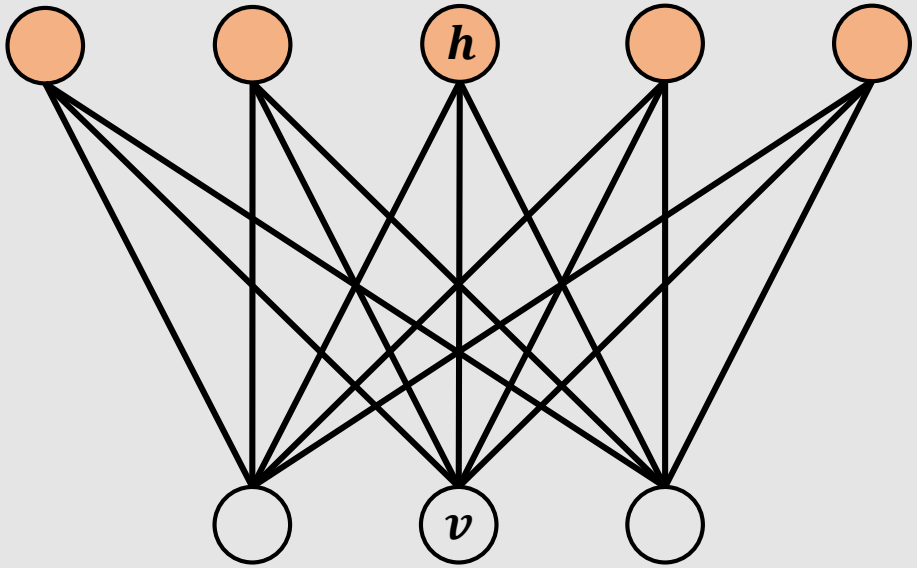


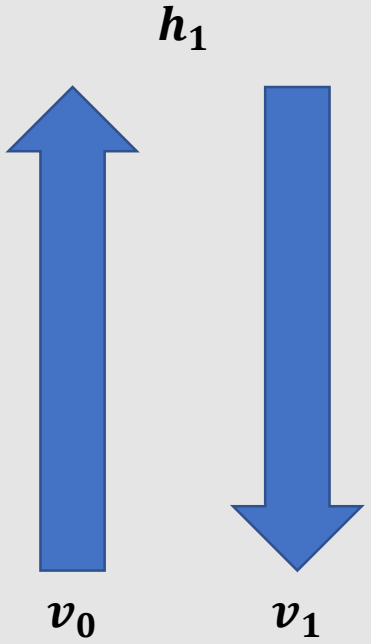
Boltzmann Machine

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\boldsymbol{h}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{h}$$

- Restricted Boltzmann Machine.

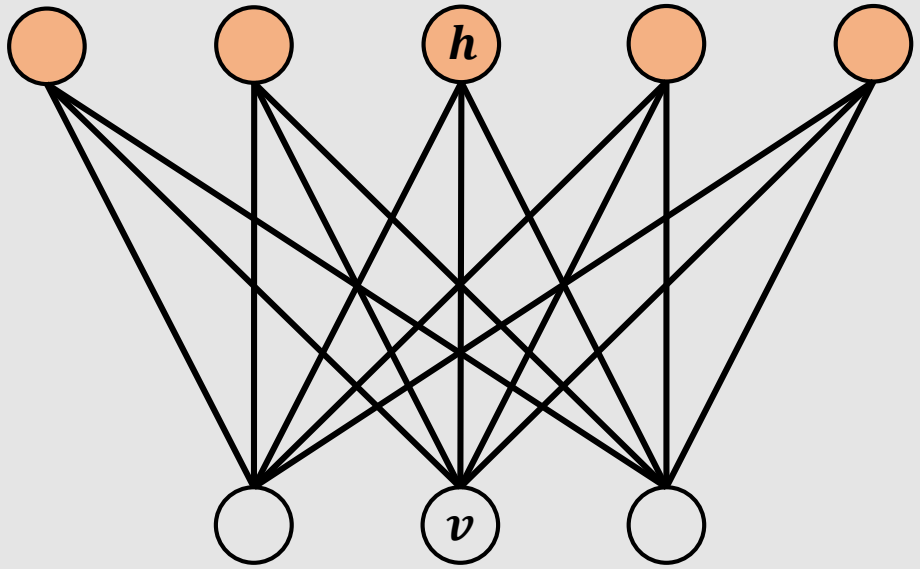
$$P(\boldsymbol{v}, \boldsymbol{h}) = \frac{\exp\left(-E(\boldsymbol{v}, \boldsymbol{h})\right)}{Z}, P(\boldsymbol{v}) = \sum_{\boldsymbol{h}} P(\boldsymbol{v}, \boldsymbol{h})$$





Boltzmann Machine

- Restricted Boltzmann Machine.



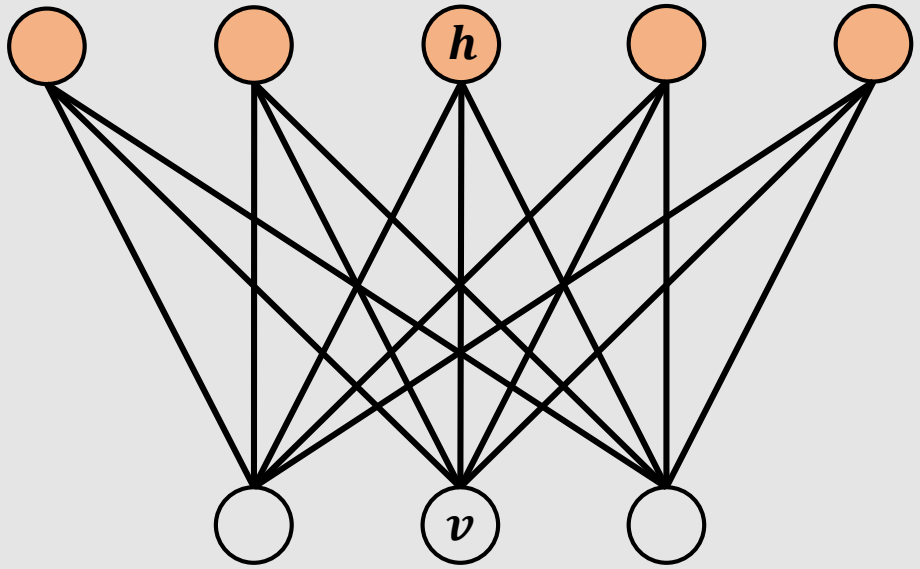
$$P(\boldsymbol{v},\boldsymbol{h}) = \frac{\exp\left(-E(\boldsymbol{v},\boldsymbol{h})\right)}{Z}$$

$$= \frac{1}{Z}\exp\left(\boldsymbol{h}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{v}\right)\exp\left(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}\right)\exp\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{h}\right)$$

$$= \frac{1}{Z}\prod_{j}\prod_{k}\exp\left(W_{jk}h_j v_k\right)\prod_{k}\exp\left(c_k v_k\right)\prod_{j}\exp\left(b_j h_j\right)$$

Boltzmann Machine

- Restricted Boltzmann Machine.



$$P(\boldsymbol{h}|\boldsymbol{v}) = \prod_j P(h_j|\boldsymbol{v})$$

$$P(h_j = 1|\boldsymbol{v}) = \sigma(b_j + \boldsymbol{W}_j\boldsymbol{v})$$

$$P(\boldsymbol{v}|\boldsymbol{h}) = \prod_k P(v_k|\boldsymbol{h})$$

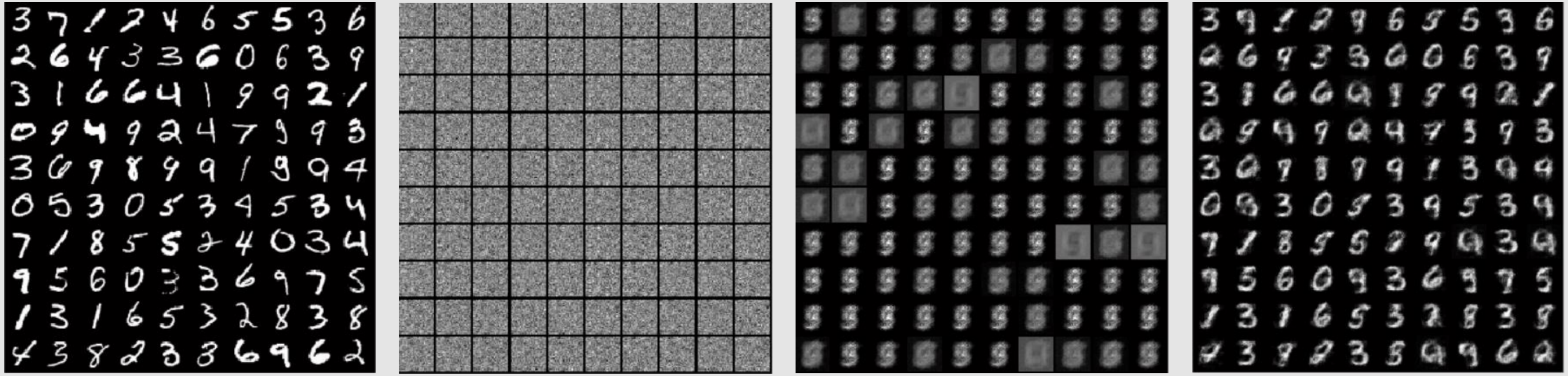
$$P(v_k = 1|\boldsymbol{h}) = \sigma(c_k + \boldsymbol{h}^{\mathrm{T}}\boldsymbol{W}_k)$$

Boltzmann Machine

- Restricted Boltzmann Machine
  - Feature extractor
    → deterministic model보다 다양한 feature를 얻을 수 있다.
  - Pre-training for supervised learning
    → Energy의 감소를 목표로 하지만 Energy가 증가할 확률도 존재
    → Global minima를 찾기 쉬워진다.
  - Data generator
    → Energy가 감소하는 방향으로 학습되기 때문에 input data와 유사한 data를 얻는다.

Boltzmann Machine

- Restricted Boltzmann Machine.



1) Boltzmann Machine

- Restricted Boltzmann Machine.

Boltzmann Machine

- Difference between Autoencoder and Boltzmann Machine
  - Autoencoder는 input data와 output data와의 차이를 그대로 이용하는 deterministic한 방법으로 학습한다.

$$h = f(Wx + b)$$

  - Boltzmann Machine은 완전히 probability만을 이용하여 학습한다.

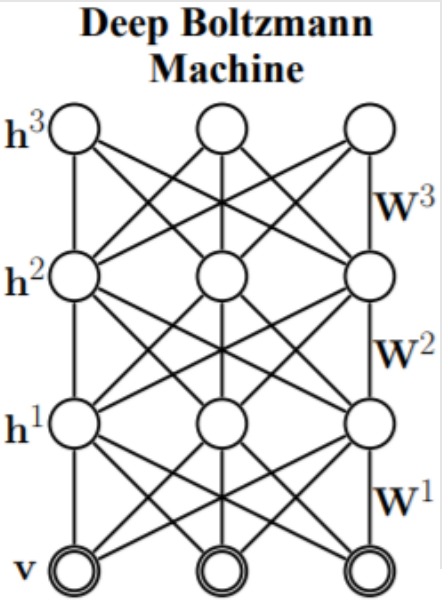
→ 확률 자체를 이용하여 energy를 줄이는 방향으로 학습.

$$P(v, h) = f(Wx + b)$$

Boltzmann Machine

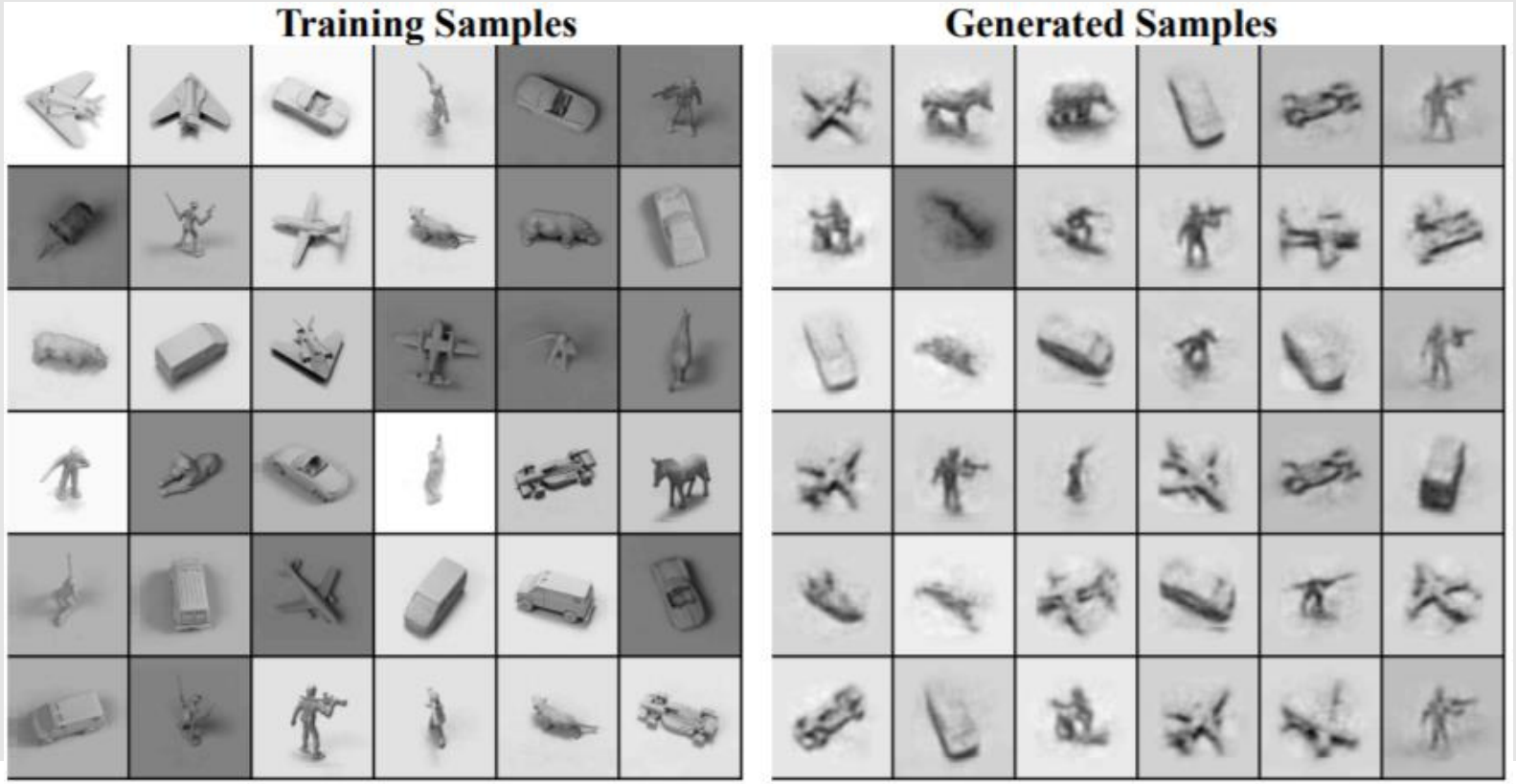
- Deep Boltzmann Machine(DBM)
  - Multilayer Restricted Boltzmann Machine
  - DBM can learn internal representations that become increasingly complex.
  - High-level representations can be built from a large supply of unlabeled inputs and very limited labeled data

  - Autoencoder와 비슷하게 greedy layer-wise pretraining을 사용
  - Stacked autoencoder와 다르게 undirected model



Boltzmann Machine

- Deep Boltzmann Machine(DBM)



B

- D

Image	Given Tags	Generated Tags	Input Tags	Nearest neighbors to generated image features	
	pentax, k10d, kangarooisland, southaustralia, sa, 300mm, australia, australiansealion	beach, sea, surf, strand, shore, wave, seascape, sand, ocean, waves	nature, hill, scenery, green, clouds		
	< no text >	night, lights, christmas, nightshot, nacht, nuit, notte, longexposure, noche, nocturna	flower, nature, green, flowers, petal, petals, bud		
	aheram, 0505, sarahc, moo	portrait, bw, balckandwhite, people, faces, girl, blackwhite, person, man	blue, red, art, artwork, painted, paint, artistic, surreal, gallery, bleu		
	unseulpixel, naturey crap	fall, autumn, trees, leaves, foliage, forest, woods, branches, path	bw, blackandwhite, noiretblanc, bianconero, blancoynegro