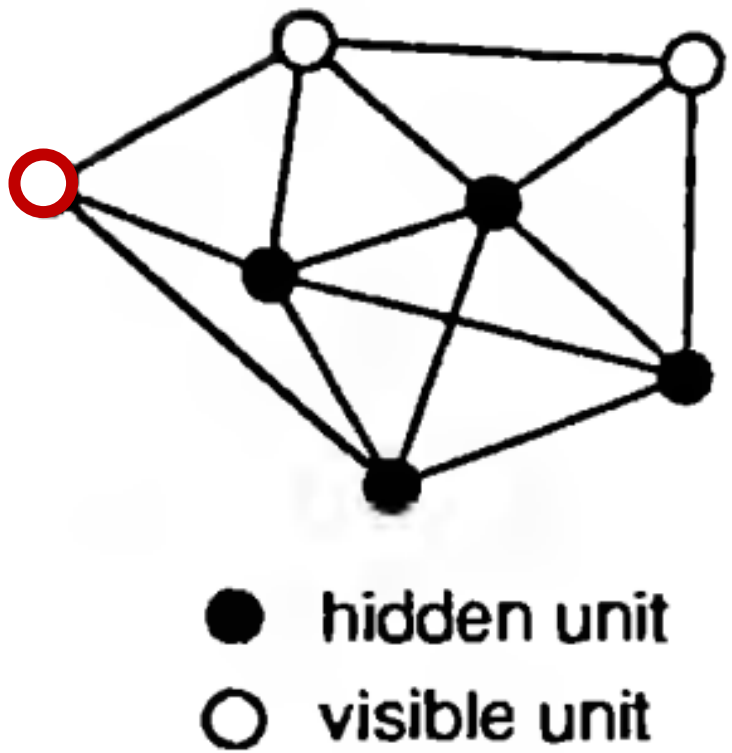
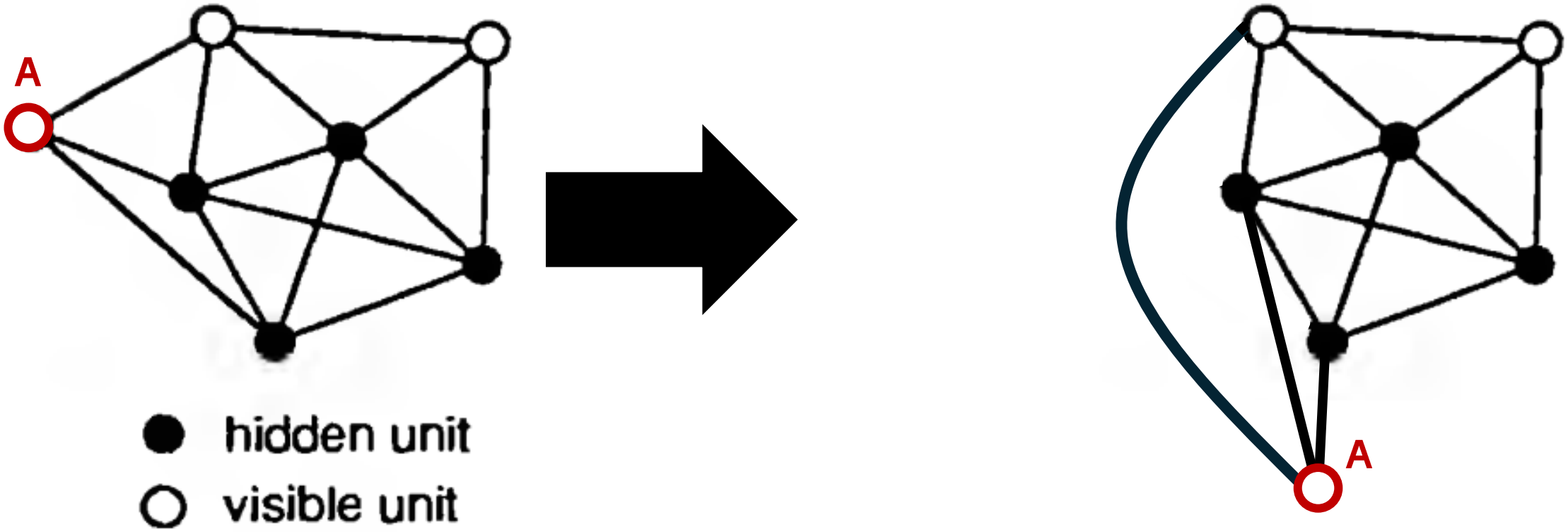
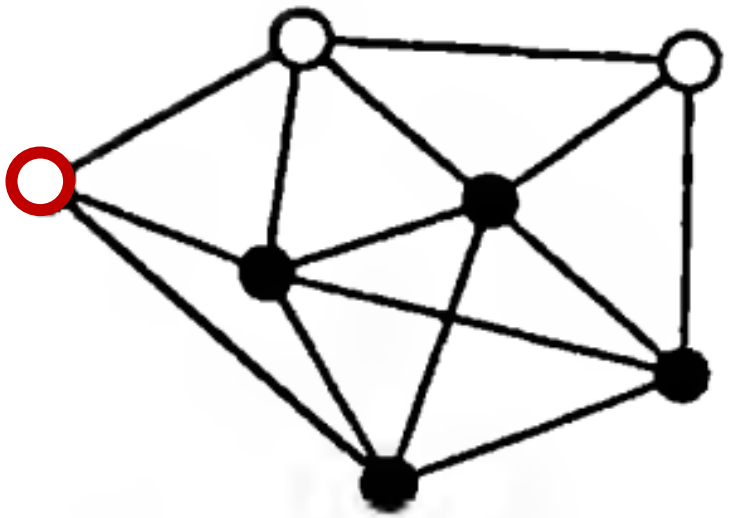
볼츠만 머신

볼츠만 머신 전제

- 대칭 연결 ($w_{ij} = w_{ji}$)
  - 이유: 안정적인 상태로 도달 가능 , 즉 $\Delta H \leq 0$ 임.
- 각각의 유닛은 +1, -1의 상태를 가짐. ($S_i \in \{-1, +1\}$)
- **유닛**은 가시 유닛과 은닉 유닛으로 나뉨.
- 가시 유닛을 입력 유닛과 출력 유닛으로 나뉨.


hidden unit
visible unit

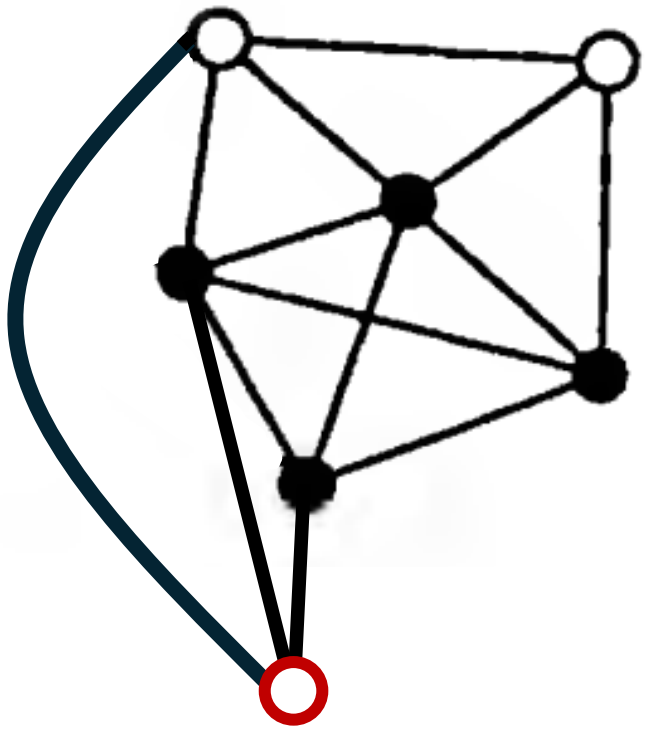

A
hidden unit
visible unit
A

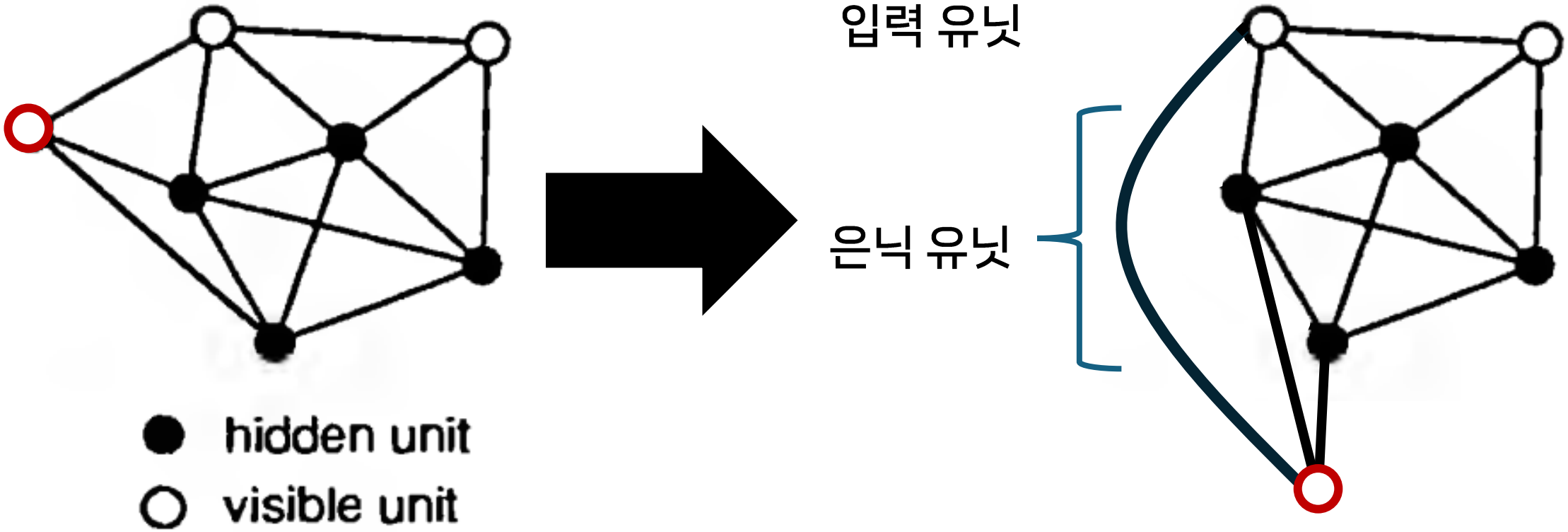


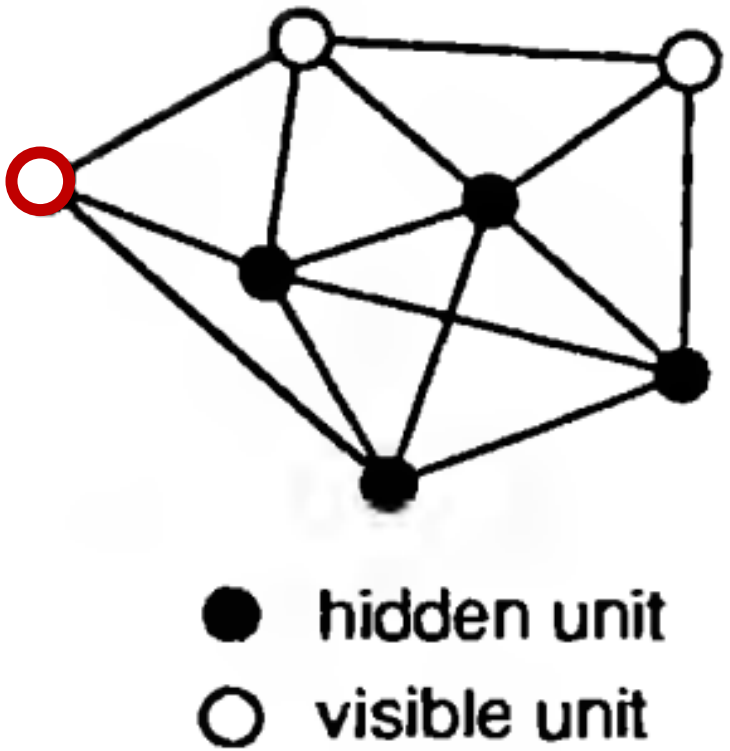
hidden unit

visible unit

입력 유닛



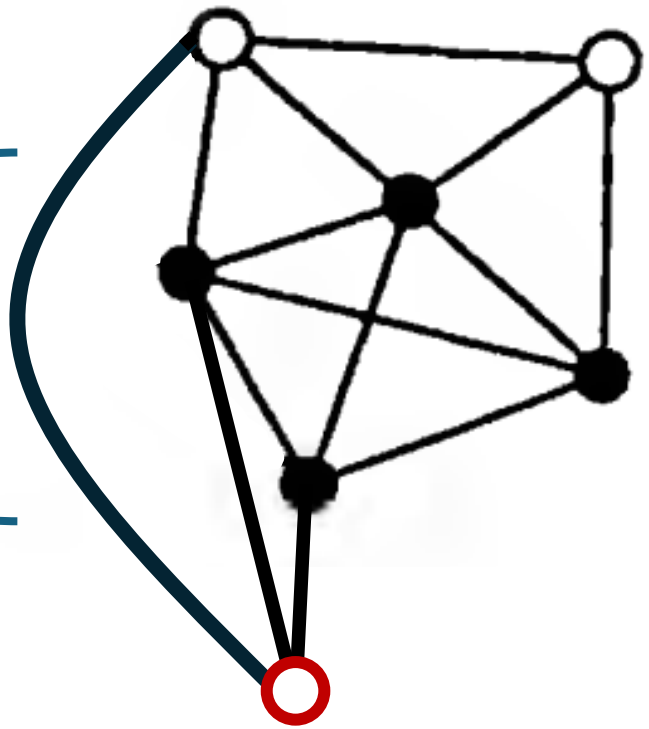

hidden unit
visible unit
입력 유닛
은닉 유닛



입력 유닛

은닉 유닛

출력 유닛



입력 유닛

은닉 유닛

출력 유닛



볼츠만 머신 수식

총 입력 $h_i = \sum_j w_{ij} S_j$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}$

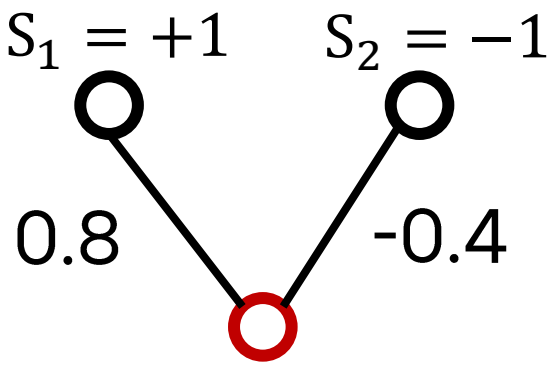
확률 $\mathrm{P}(\mathrm{S_i} = +1) = g(h_i), \quad \mathrm{P}(\mathrm{S_i} = -1) = 1 - g(h_i)$

볼츠만 머신 수식

총 입력 $h_i = \sum_j w_{ij} S_j$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$ T는 온도

확률 $\mathrm{P}(\mathrm{S_i} = +1) = g(h_i), \quad \mathrm{P}(\mathrm{S_i} = -1) = 1 - g(h_i)$

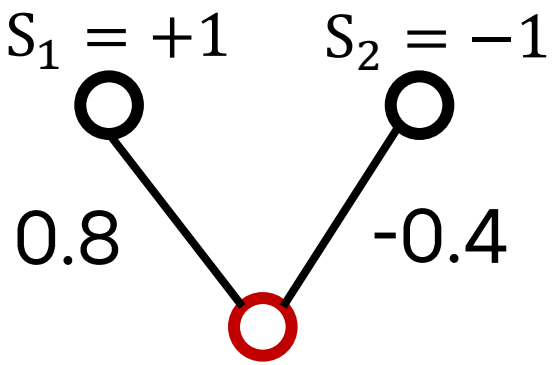


볼츠만 머신 수식

총 입력 $h_i = \sum_j w_{ij} S_j$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$ T는 온도

확률 $\mathrm{P}(\mathrm{S_i} = +1) = g(h_i), \quad \mathrm{P}(\mathrm{S_i} = -1) = 1 - g(h_i)$



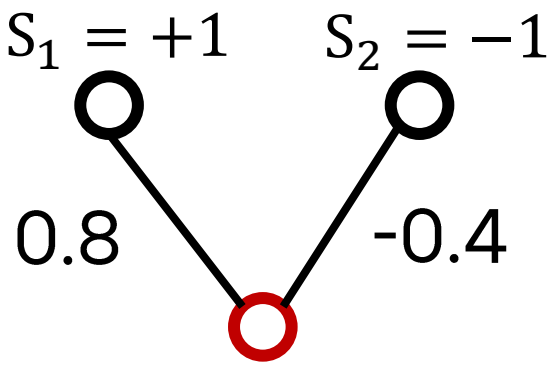
$h_3 = (0.8 * 1) + (-0.4 * -1) = 1.2$

볼츠만 머신 수식

총 입력 $h_i = \sum_j w_{ij} S_j$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$ T는 온도

확률 $\mathrm{P}(\mathrm{S_i} = +1) = g(h_i), \quad \mathrm{P}(\mathrm{S_i} = -1) = 1 - g(h_i)$



$h_3 = (0.8 * 1) + (-0.4 * -1) = 1.2$

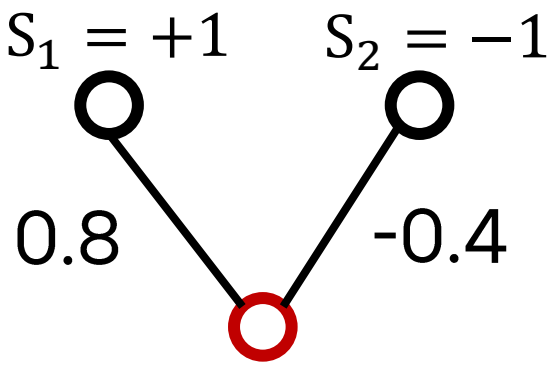
$\beta = 1, g(h_3) \approx 0.9$

볼츠만 머신 수식

총 입력 $h_i = \sum_j w_{ij} S_j$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$ T는 온도

확률 $\mathrm{P}(\mathrm{S_i} = +1) = g(h_i), \quad \mathrm{P}(\mathrm{S_i} = -1) = 1 - g(h_i)$

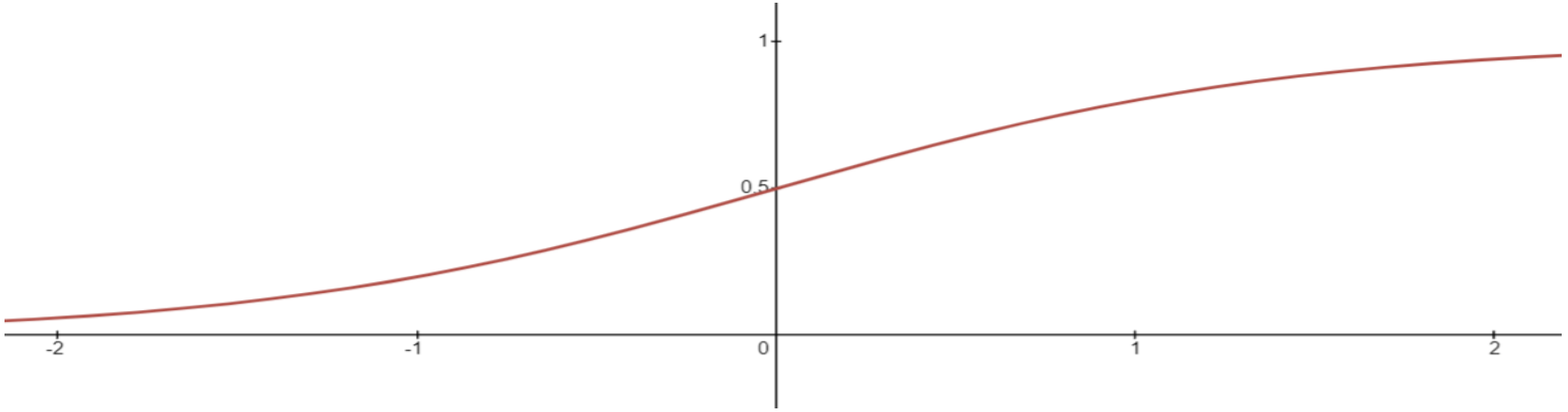


$h_3 = (0.8 * 1) + (-0.4 * -1) = 1.2$

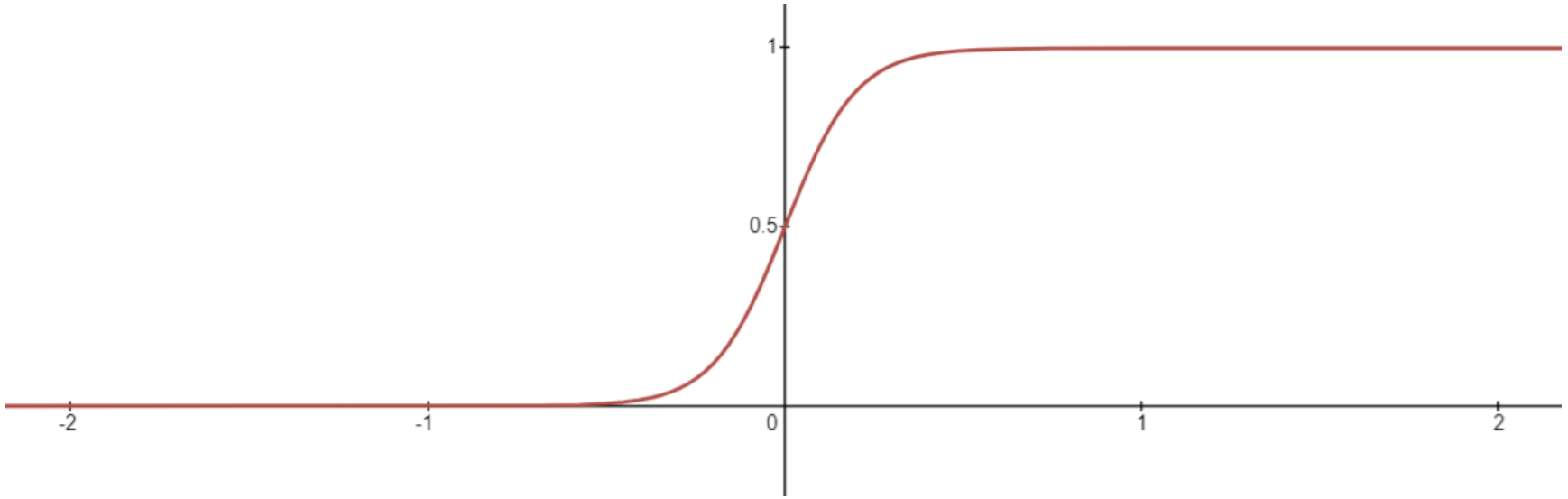
$\beta = 1, g(h_3) \approx 0.9$

$\mathrm{P}(\mathrm{S_3} = +1) = g(h_3) = 0.9$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$

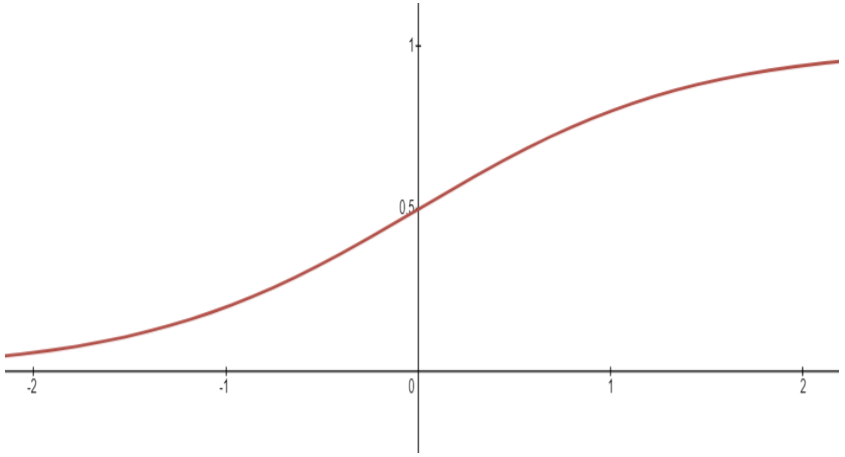


$\beta = 0.5$



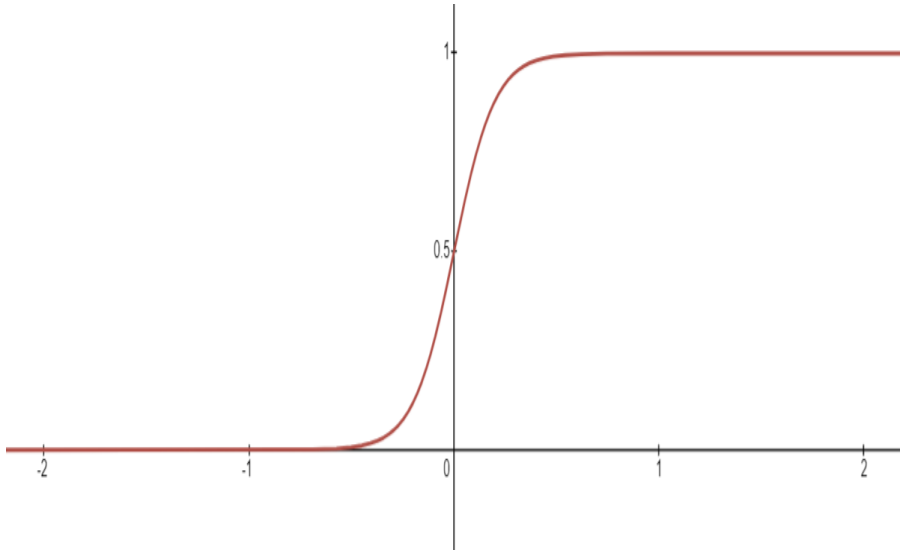
$\beta = 5$

활성화 함수 $g(h_i) = \frac{1}{1+\exp(-2\beta h_i)}, \beta = \frac{1}{T}$



$\beta = 0.5$

출력 무작위 ▲



$\beta = 5$

출력 무작위 ▼

$T \uparrow \Rightarrow \beta \downarrow$
$\Rightarrow$ 출력 무작위

$T \downarrow \Rightarrow \beta \uparrow$
$\Rightarrow$ 출력 결정론적

에너지 함수, 상태 확률 예시

$$H(\{S_i\}) = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j$$

에너지 함수, 상태 확률 예시

$$H(\{S_i\}) = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j$$

$$P(\{S_i\}) = \frac{e^{-\beta H(\{S_i\})}}{Z}$$

에너지 함수, 상태 확률 예시

$$H(\{S_i\}) = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j$$

$$P(\{S_i\}) = \frac{e^{-\beta H(\{S_i\})}}{Z}$$

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

에너지 함수, 상태 확률 예시

$w_{12} = 2, \beta = 1$

$S = (S_1, S_2)$

$$H(S) = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j = -2S_1 S_2$$

$$P(S) = \frac{e^{-\beta H(S)}}{Z} = \frac{e^{2S_1 S_2}}{Z} = \frac{e^{2S_1 S_2}}{Z}$$

$$Z = 2e^2 + 2e^{-2}$$

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

에너지 함수, 상태 확률 예시

$$w_{12} = 2, \beta = 1$$

$$S = (S_1, S_2)$$

$$P(S) = \frac{e^{-\beta H(S)}}{Z} = \frac{e^{2S_1S_2}}{Z} = \frac{e^{2S_1S_2}}{Z}$$

$$Z = 2e^2 + 2e^{-2}$$

$$P(\{S_1 = +1, S_2 = +1\}) = \frac{e^2}{Z} = 0.49$$

가시 유닛 상태의 확률

$$P_\alpha = \sum_\gamma P_{\alpha\gamma} = \frac{\sum_\gamma e^{-\beta H_{\alpha\gamma}}}{Z}$$

α: 가시 유닛 상태 집합
γ: 은닉 유닛 상태 집합
P_α: 가시 상태가 α일 확률
$P_{\alpha\gamma}$: 가시 상태가 α, 은닉 상태가 γ일 확률
$H_{\alpha\gamma}$: α γ 일 때, 전체 상태의 에너지

오차

$$E = \sum_{\alpha} R_\alpha \ln \frac{R_\alpha}{P_\alpha} \geq 0$$

$$E = 0 \Leftrightarrow R_\alpha = P_\alpha$$

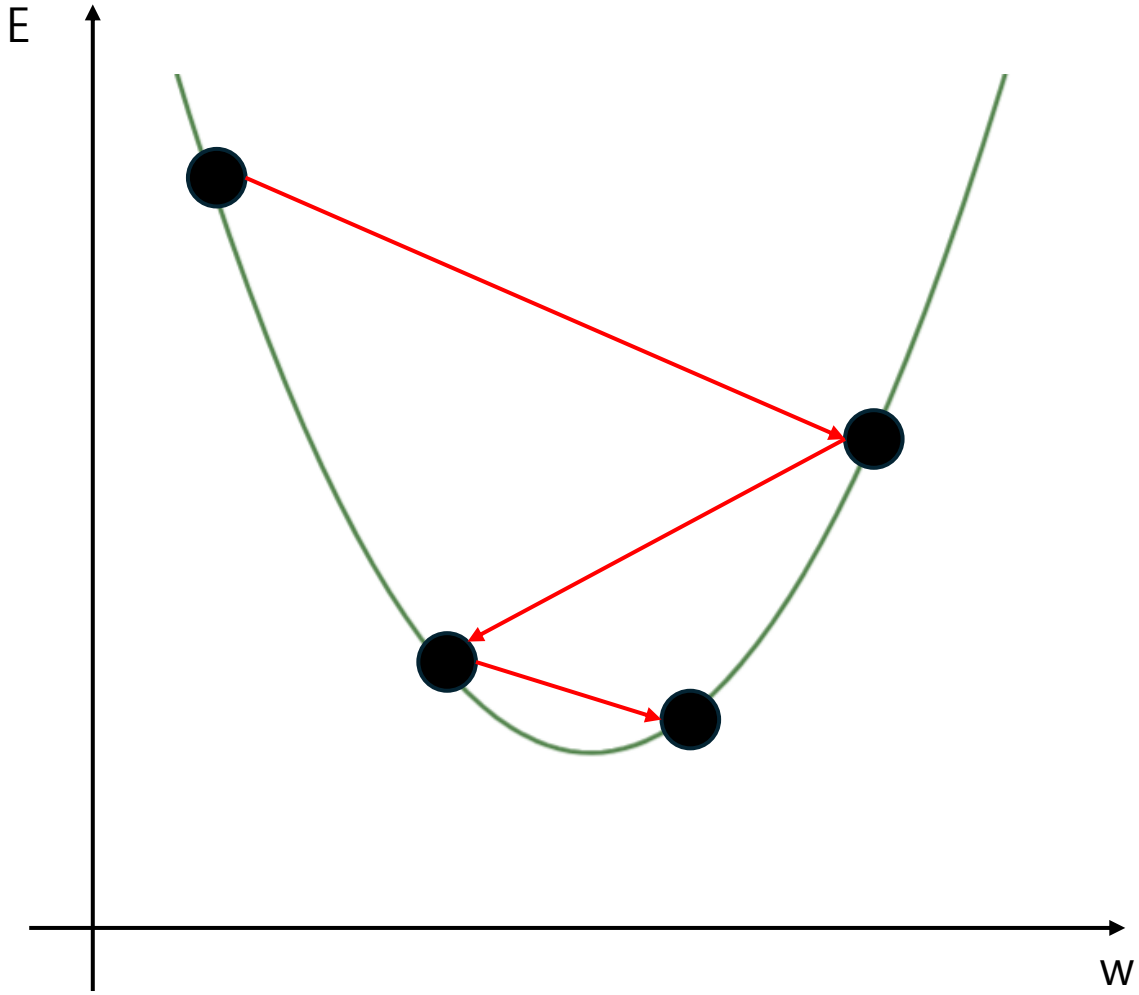
R_α: 원하는 목표 확률분포

α: 가시 유닛 상태 집합
P_α: **현재 가중치로 구한** 가시 상태가 α일 확률

가중치 학습 방법-경사 하강법

경사 하강법

$$w_{new} = w_{old} - \eta \frac{dE}{dw}$$



가중치 학습 방법-경사 하강법

경사 하강법

$$w_{new} = w_{old} - \eta \frac{dE}{dw}$$

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} = -\eta \sum_{\alpha} \frac{R_\alpha}{P_\alpha} \frac{\partial P_\alpha}{\partial w_{ij}}$$

$$\frac{\partial P_\alpha}{\partial w_{ij}} = \beta [\sum_{\gamma} S_i^{\alpha\gamma} S_j^{\alpha\gamma} P_{\alpha\gamma} - P_\alpha \{\sum_{\alpha,\gamma} P_{\alpha\gamma} S_i^{\alpha\gamma} S_j^{\alpha\gamma}\}]$$

가중치 학습 방법-경사 하강법

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} = -\eta \sum_{\alpha} \frac{R_\alpha}{P_\alpha} \frac{\partial P_\alpha}{\partial w_{ij}} = \cdots$$

$$= \eta\beta[< S_i S_j >_{clamped} - < S_i S_j >_{free}]$$

$< S_i S_j >_{clamped}$: 가시 유닛을 목표 상태에 고정한 후, $S_i S_j$의 평균

$< S_i S_j >_{free}$: 현재 가중치에서의 $S_i S_j$의 평균

가중치 학습 방법-경사 하강법

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} = -\eta \sum_{\alpha} \frac{R_\alpha}{P_\alpha} \frac{\partial P_\alpha}{\partial w_{ij}} = \cdots$$

$$= \eta\beta[< S_i S_j >_{clamped} - < S_i S_j >_{free}]$$

$< S_i S_j >_{clamped}$: 가시 유닛을 목표 상태에 고정한 후, $S_i S_j$의 평균

$< S_i S_j >_{free}$: 현재 가중치에서의 $S_i S_j$의 평균

가중치 학습 방법 예제



목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

초기 상태: 모든 가중치 값은 0
β=1, η=0.2

가중치 학습 방법 예제

$$\Delta w_{12} = \eta\beta[< S_1 S_2 >_{clamped} - < S_1 S_2 >_{free}]$$



초기 상태: 모든 가중치 값은 0
β=1, η=0.2

목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

가중치 학습 방법 예제

$$\Delta w_{12} = 0.2[< S_1 S_2 >_{clamped} - < S_1 S_2 >_{free}]$$



초기 상태: 모든 가중치 값은 0
β=1, η=0.2

목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

가중치 학습 방법 예제



$$< S_1 S_2 >_{clamped} = 1$$

S1	S2	S3	확률
1	1	1	0.5
1	1	-1	0.5
-1	-1	1	0.5
-1	-1	-1	0.5

초기 상태: 모든 가중치 값은 0
β=1, η=0.2

목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

가중치 학습 방법 예제

$$< S_1 S_2 >_{free} = 0$$



S1	S2	S3	확률
1	1	1	0.125
1	1	-1	0.125
1	-1	1	0.125
...			0.125
-1	-1	-1	0.125

초기 상태: 모든 가중치 값은 0
β=1, η=0.2

목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

가중치 학습 방법 예제

$$\Delta w_{12} = 0.2[< S_1 S_2 >_{clamped} - < S_1 S_2 >_{free}]$$
$$= 0.2(1 - 0) = 0.2$$

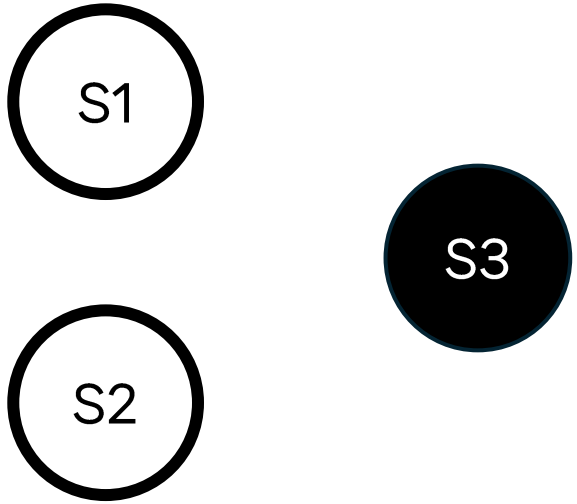


$$w_{12\,new} = w_{12\,old} + 0.2 = 0 + 0.2 = 0.2$$

초기 상태: 모든 가중치 값은 0
β=1, η=0.2

목표 가시 상태:
P(S1, S2) = P(+1, +1) = P(-1, -1) = 0.5

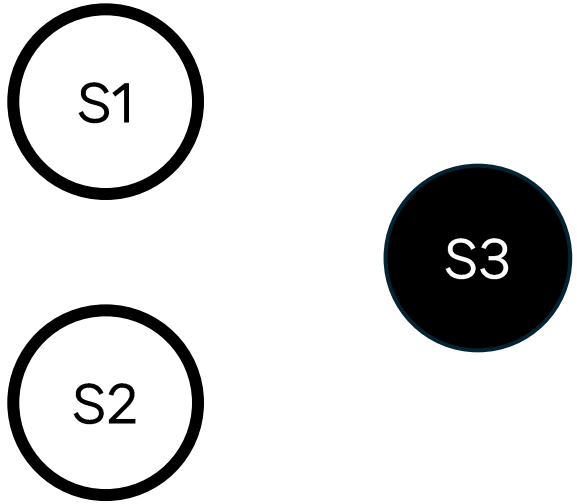
가중치 학습 방법 예제



$$w_{12\,new} = 0.2$$

초기 상태: 모든 가중치 값은 0
β=1, η=0.2

가중치 학습 방법 예제



S1	S2	확률
1	1	0.29
1	-1	0.20
-1	1	0.20
-1	-1	0.29

$w_{12\,new} = 0.2$

β=1, η=0.2

상태 변경

$$P(S_i \rightarrow -S_i) = \frac{1}{1 + e^{\beta \Delta H_i}}$$

에너지 감소 → 확률 증가

에너지 증가 → 확률 감소

상태 변경

홉필드 모델에서는 에너지 변화량이 항상 0 이하였다. (ΔH≤0)
하지만 볼츠만 머신은 아니다.

S_1	S_2	H	ΔH
+1	+1	-2	2-(-2)=4
-1	+1	2	

$w_{12} = 2, \beta = 1$

$S = (S_1, S_2)$

$$H(S) = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j = -2 S_1 S_2$$

상태 변경

홉필드 모델에서는 에너지 변화량이 항상 0 이하였다. ($\Delta H \leq 0$)
하지만 볼츠만 머신은 아니다.

S_1	S_2	H	ΔH
+1	+1	-2	2-(-2)=4
-1	+1	2	

확률은 약 0.01

$\Delta H > 0$ 일때 상태 변경이 될 가능성은 적지만 불가능 한 것은 아님
이러한 특성 때문에 국소 최솟값에서 벗어나 다른 상태도 탐색할 수 있음

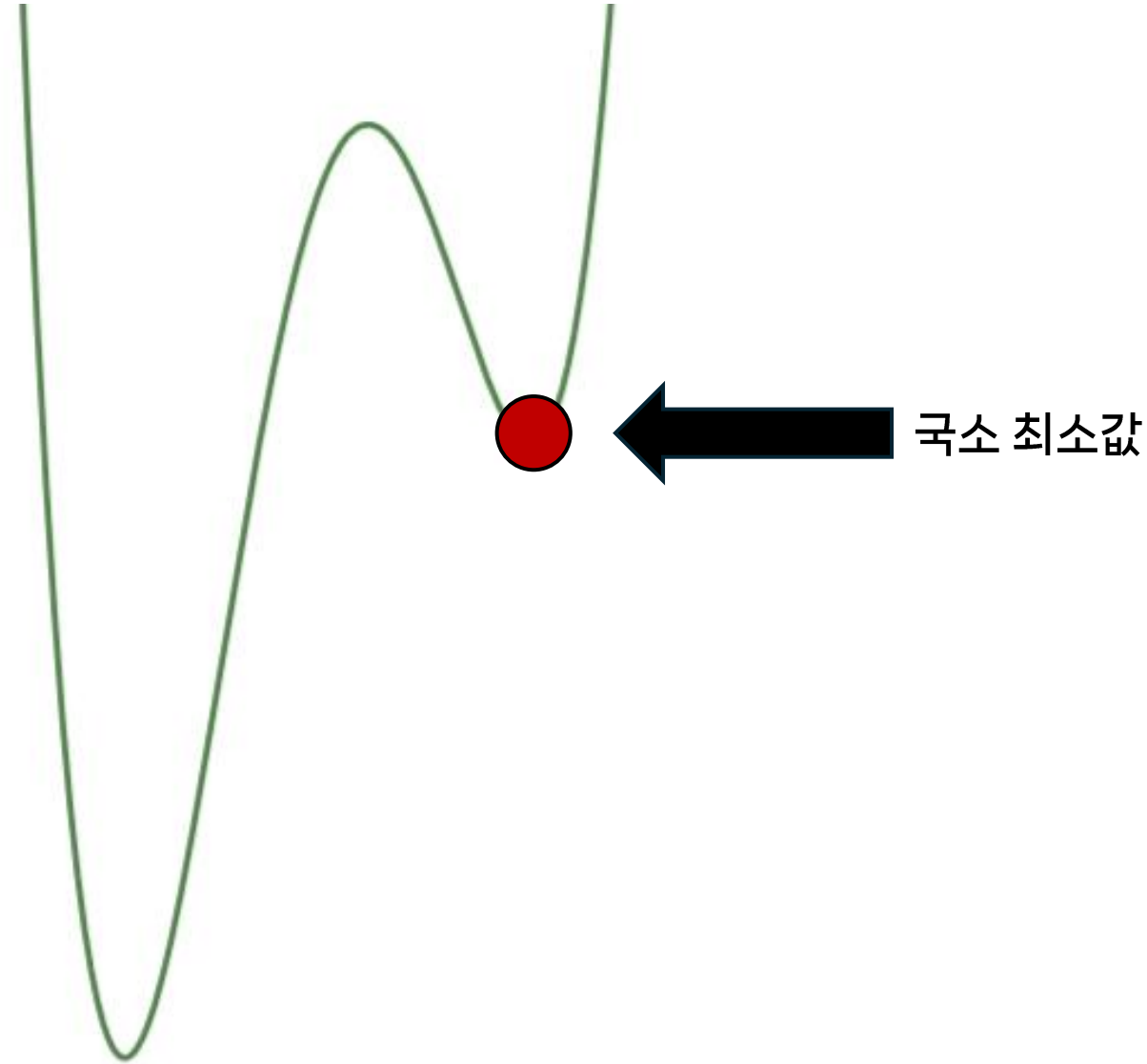
시뮬레이티드 어닐링

국소 최솟값에서 벗어나 다른 상태도 탐색할 수 있음
하지만 β값이 높으면, 즉 T(온도)가 낮으면 ΔH_i>0 일때, 상태가 바뀔 확률이 매우 적어짐

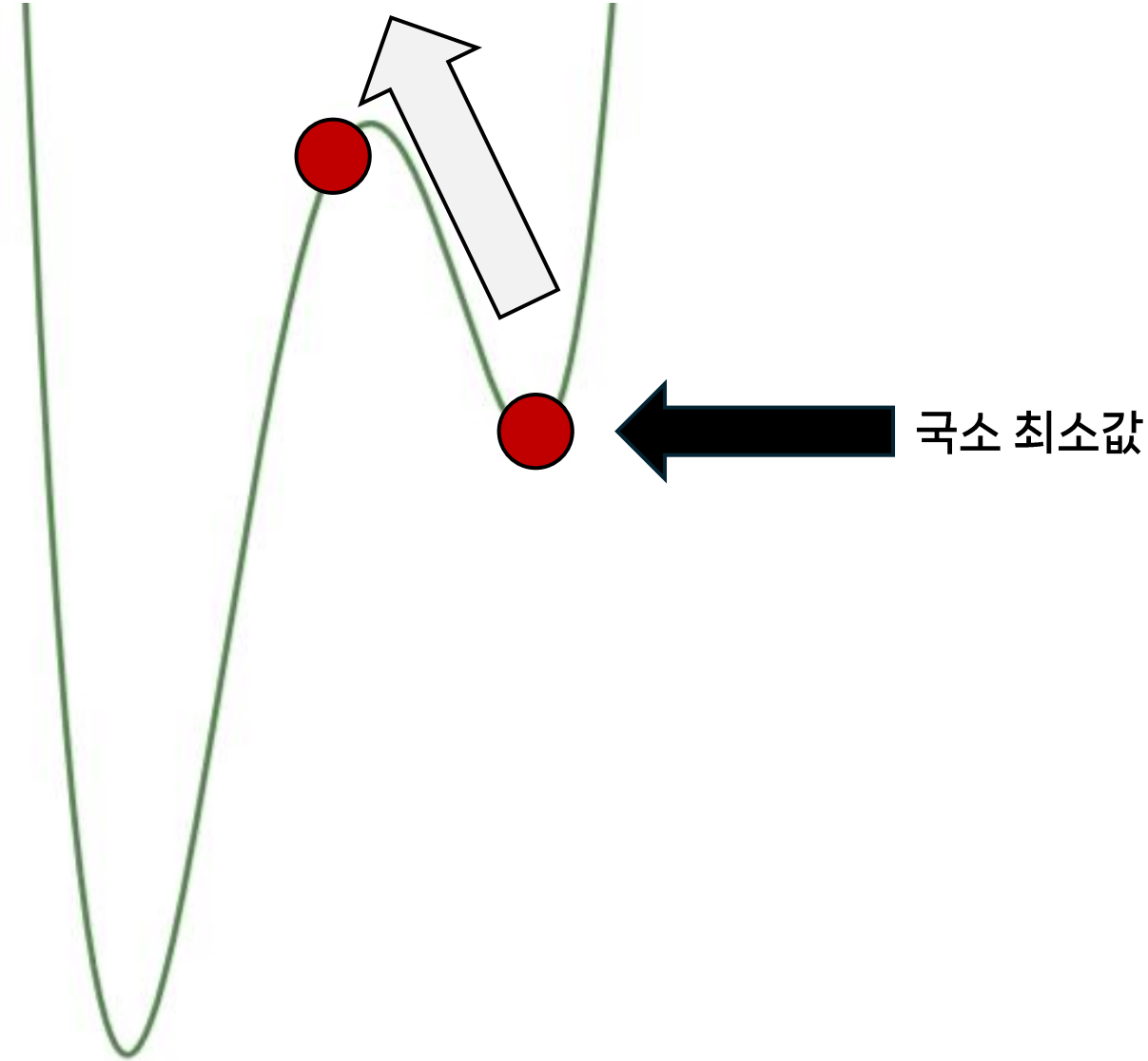
$$P(S_i \rightarrow -S_i) = \frac{1}{1 + e^{\beta \Delta H_i}}$$

그래서 높은 온도에서 시작하여 점점 온도를 낮춤

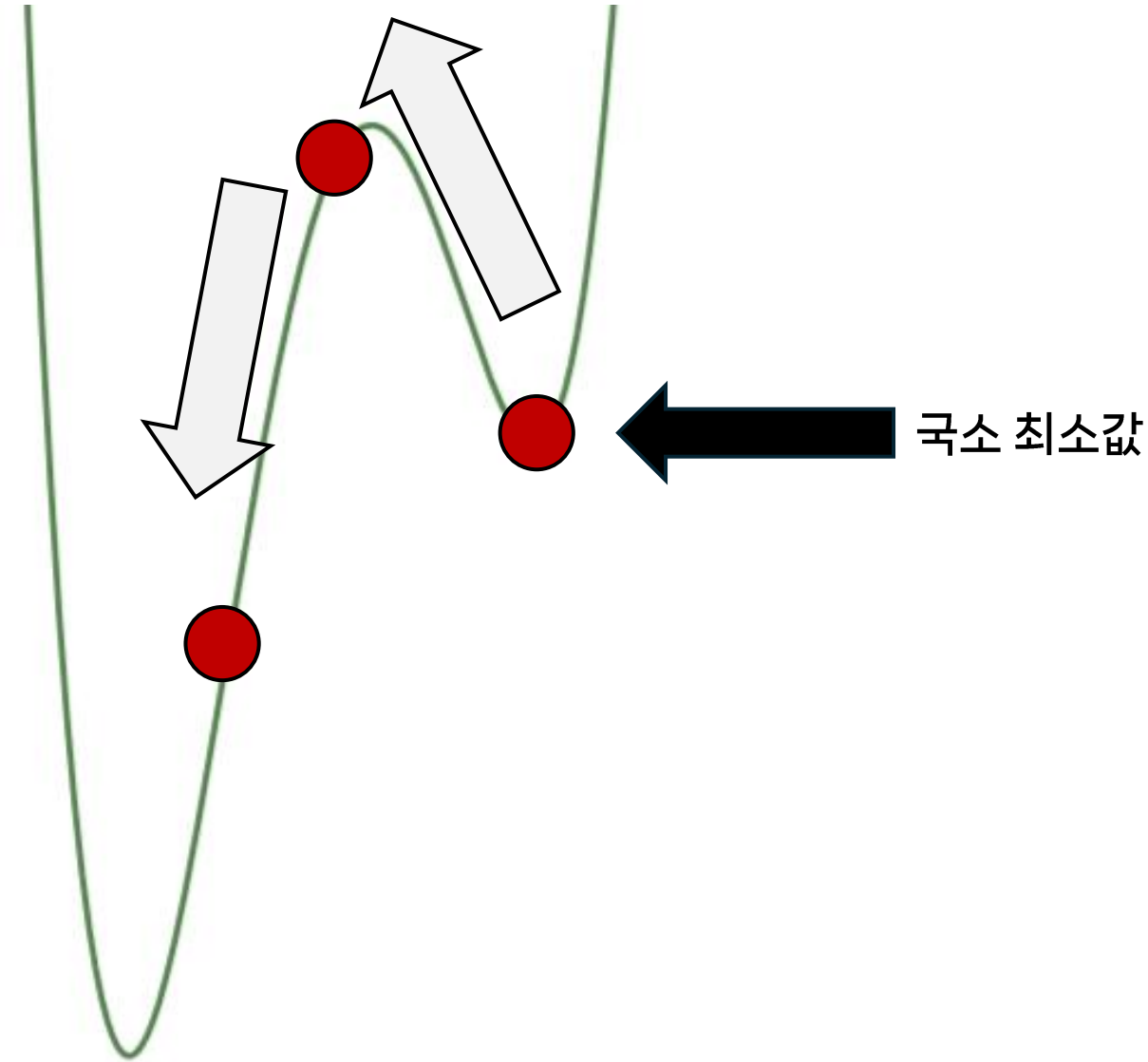
△H > 0 이면 넘어갈 수 있음



△H > 0 이면 넘어갈 수 있음



△H > 0 이면 넘어갈 수 있음



볼츠만 머신 변형

- 가시 유닛을 입력, 출력으로 구분하는 경우

$$E = \sum_{\gamma} Q_{\gamma} \sum_{\alpha} \mathrm{R}_{\alpha|\gamma} \ln \frac{R_{\alpha|\gamma}}{P_{\alpha|\gamma}}$$

$\boldsymbol{Q_{\gamma}}$: 입력 $\boldsymbol{\gamma}$이 나타날 확률

$\mathrm{R}_{\alpha|\gamma}$: 입력 $\boldsymbol{\gamma}$가 주어졌을 때 출력 α이 나올 확률 (목표)

$P_{\alpha|\gamma}$: 입력 $\boldsymbol{\gamma}$가 주어졌을 때 출력 α이 나올 확률 (현재)

볼츠만 머신 한계

속도가 느림

계산할 항이 많음

$$\Delta w_{ij} = \eta\beta[< S_i S_j >_{clamped} - < S_i S_j >_{free}]$$

1. w 여러 번 계산
2. 각 w마다 clamped, free 계산
3. 각 clamped, free 계산마다 모든 S, w를 사용

볼츠만 머신 한계 극복

- 평균장 근사 (10-30 배 빠르게 학습할 수 있음)

$$< S_i S_j > \approx < S_i >< S_j > = m_i m_j$$

clamped에서는 ±1로 고정, free에서는 다음 식 이용

$$m_i = \tanh\left(\beta \sum_j w_{ij} m_j\right) \in [-1, 1]$$

다음 과정 반복

$$m_i^{new} = \tanh\left(\beta \sum_j w_{ij} m_j^{old}\right)$$

정리

- 볼츠만 머신의 목표
  - 가중치를 조정함으로써 유닛 상태 패턴 확률이 목표 분포와 비슷해지도록 만드는 것

- 홉필드 모델과 차이점
  - 에너지 함수 변화량이 양수일 수 있음 → 시뮬레이티드 어널링을 통해 국소 최솟값을 벗어날 수 있음

- 은닉층 추가 → 고차원적인 패턴 학습 가능
  - XOR 문제 해결

"Boltzmann Machines as Generalized Hopfield Networks: A Review of Recent Results and Outlooks"

논문 내용

- 볼츠만 머신 = 홉필드 네트워크 ?

비교 항목	볼츠만 머신	홉필드 모델
특징	데이터의 확률분포 학습	불완전하거나 잡음이 포함된 입력에서 저장된 패턴을 복원
구성	가시 유닛 N개 은닉 유닛 P개 연결 가중치 $w_{i\mu}$	뉴런 N개 저장패턴 P개 패턴 요소 ξ_i^μ
동작	학습 데이터의 통계적 구조로 수렴	저장된 패턴으로 수렴

홉필드 모델

이진 뉴런 N개: $\sigma_i \, \epsilon \, \{-1, +1\}$
저장 패턴 P개: $\mu = 1, \ldots, P$
μ번째 패턴에서 i번째 뉴런이 가져야 하는 값: ξ_i^μ

헤브 학습 규칙: $J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu$

에너지 함수: $H_H(\sigma) = -\frac{1}{2} \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j$

확률 분포: $P_H(\sigma) = \frac{1}{Z_H} e^{-\beta H_H(\sigma)}$, $P_H(\sigma) \propto \exp\left[\frac{\beta}{2N} \sum_{i,j,\mu} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j\right]$

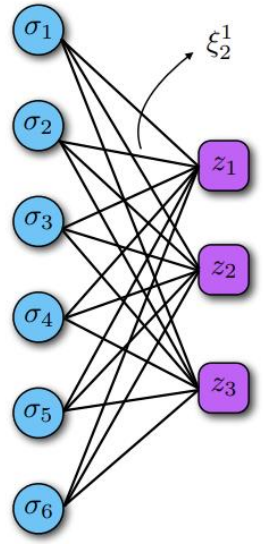
볼츠만 머신

$$H_B(\sigma, z) = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{\mu=1}^{P} W_{i\mu} \sigma_i z_\mu$$

전제: 가시 유닛끼리 연결되어 있지 않음,

$$z_\mu \sim \mathcal{N}(0, \beta^{-1})$$

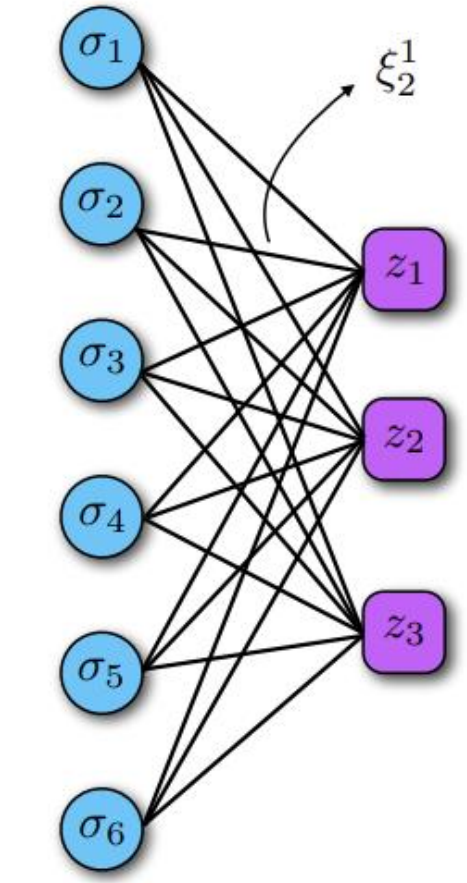
z: 은닉 유닛
σ: 가시 유닛



Boltzmann Machine

RBM

- RBM (restricted 볼츠만 머신)
  - 두 개의 층으로 구성
  - 하나의 층 내부에서는 서로 연결되지 않음



Boltzmann Machine

Marginalization of Hidden Units

$$P_B(\sigma, z) \propto \exp\left[-\frac{\beta}{2}\sum_{\mu=1}^{P} z_\mu^2 + \frac{\beta}{\sqrt{N}}\sum_{i=1}^{N}\sum_{\mu=1}^{P} W_{i\mu}\sigma_i z_\mu\right]$$

원하는 것: 가시 유닛의 확률 분포 → 은닉 유닛 z 제거

z 제거 방법: 적분 이용 $P_B(\sigma) = \int P_B(\sigma, z)\, dz$

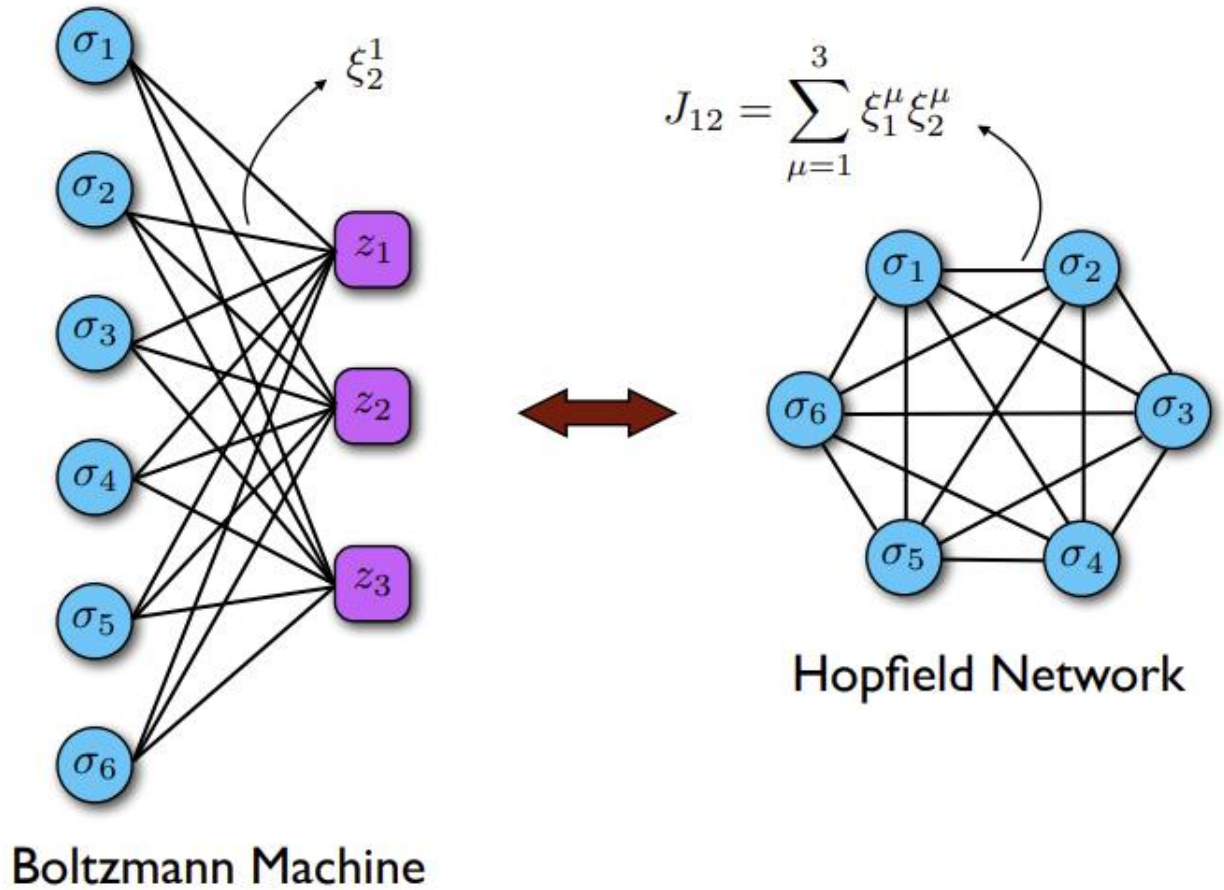
recap: $P_\alpha = \sum_\gamma P_{\alpha\gamma}$

Marginalization of Hidden Units

$$P_B(\sigma) \propto \exp\left[\frac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{P} W_{i\mu}W_{j\mu}\sigma_i\sigma_j\right]$$

이 식은 홉필드 모델의 식과 동일하다.

$$P_H(\sigma) \propto \exp\left[\frac{\beta}{2N}\sum_{i,j,\mu} \xi_i^\mu \xi_j^\mu \sigma_i\sigma_j\right]$$


σ_1
σ_2
σ_3
σ_4
σ_5
σ_6
ξ_2^1
z_1
z_2
z_3
$J_{12} = \sum_{\mu=1}^{3} \xi_1^\mu \xi_2^\mu$
σ_1
σ_2
σ_3
σ_4
σ_5
σ_6
Boltzmann Machine
Hopfield Network