

# 볼츠만 머신

👤 생성자	👤 재환 김
🏷️ 태그	엔지니어링

볼츠만 머신(Boltzmann Machine)은 통계 물리학에서 기원한 **에너지 기반의 확률 신경망 모델**로, 확률 분포를 학습하고 샘플링하는 데 매우 유용합니다. 볼츠만 머신은 복잡한 데이터의 분포를 학습하는 데 유리하며, 확률론적 추론과 최적화 문제에도 활용됩니다. 이 모델은 모든 뉴런이 서로 상호작용하면서 시스템의 **에너지를 최소화**하고, 이를 통해 데이터의 잠재적인 구조와 패턴을 찾아내는 데 중점을 둡니다.

## 1. 볼츠만 머신의 배경: 통계 물리학과 볼츠만 분포

볼츠만 머신은 물리학에서 볼츠만 분포(Boltzmann Distribution)와 **에너지 최소화 원리**를 바탕으로 만들어졌습니다. 볼츠만 분포는 온도에 따라 입자들이 어떤 상태에 있을 확률을 나타내는 확률 분포입니다. 이 분포는 고온에서는 모든 상태에 고르게 분포되지만, 온도가 낮아지면 에너지가 낮은 상태로 수렴하는 특성을 가집니다. 볼츠만 머신은 이 원리를 차용해 각 뉴런의 상태가 에너지 최소화 상태로 점진적으로 수렴하도록 설계되었습니다.

### 볼츠만 분포

볼츠만 분포에서 특정 상태  $s$ 가 나타날 확률은 다음과 같이 정의됩니다.

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

- $P(s)$ : 시스템이 상태  $s$ 에 있을 확률
- $E(s)$ : 상태  $s$ 의 에너지
- $\beta = \frac{1}{k_B T}$ : 온도에 대한 역수, 여기서  $k_B$ 는 볼츠만 상수,  $T$ 는 절대 온도
- $Z$ : 분배 함수(정규화 상수)로, 모든 가능한 상태의 에너지를 정규화합니다.

이때, **온도가 낮을수록 에너지가 낮은 상태가 선택될 가능성이 높아** 시스템이 점점 안정된 상태로 수렴합니다. 볼츠만 머신은 이 원리를 통해 시스템의 뉴런 상태를 업데이트하며, 최종적으로 특정 확률 분포에 수렴하도록 합니다.

## 2. 볼츠만 머신의 구조

볼츠만 머신은 **완전 연결 이진 신경망**으로, 모든 뉴런이 서로 연결된 형태로 구성됩니다. 각 뉴런은 **0** 또는 **1**의 이진 값을 가지며, 상호작용을 통해 시스템의 에너지가 결정됩니다. 볼츠만 머신의 뉴런은 크게 두 종류로 나눌 수 있습니다.

- **가시 노드(Visible Units):** 데이터가 주어지는 입력층으로, 관찰된 데이터를 입력하고 학습하는 부분입니다.
- **숨겨진 노드(Hidden Units):** 데이터의 잠재적인 패턴을 학습하여 가시 노드가 더 복잡한 분포를 모델링할 수 있도록 합니다.

이 두 종류의 노드 간 상호작용으로 인해 볼츠만 머신은 데이터를 더 정교하게 표현할 수 있으며, 숨겨진 노드는 특히 입력 데이터 간의 **잠재적인 관계**를 학습하는 데 중요한 역할을 합니다.

### 3. 에너지 함수와 확률 분포

볼츠만 머신의 **에너지 함수**는 뉴런들 간의 연결 가중치와 각 뉴런의 상태에 의해 정의됩니다. 에너지가 낮은 상태일수록 더 안정적이며, 시스템은 점차적으로 에너지를 낮추는 상태로 수렴합니다. 볼츠만 머신에서의 에너지는 다음과 같은 함수로 표현됩니다.

$$E(s) = - \sum_{i < j} w_{ij} s_i s_j - \sum_i b_i s_i$$

- $E(s)$ : 특정 상태  $s$ 의 에너지
- $w_{ij}$ : 뉴런  $i$ 와  $j$  간의 가중치
- $s_i$ : 뉴런  $i$ 의 상태(0 또는 1)
- $b_i$ : 뉴런  $i$ 에 대한 바이어스

이 에너지는 네트워크 상태가 주어졌을 때 시스템의 안정성을 측정하며, 시스템은 에너지를 점차 최소화하여 특정 확률 분포를 형성합니다. 이 확률 분포는 에너지가 낮은 상태일수록 높은 확률을 가지며, 네트워크는 이러한 확률 분포에 수렴하도록 뉴런 상태를 업데이트합니다.

### 4. 상태 갱신과 기브스 샘플링

볼츠만 머신은 네트워크의 상태를 반복적으로 갱신하여 에너지가 최소화되는 상태로 수렴하는데, 이때 **기브스 샘플링(Gibbs Sampling)** 기법을 사용합니다. 기브스 샘플링은 각 뉴런의 상태를 확률적으로 갱신하면서 시스템이 최종적으로 안정 상태에 도달하도록 합니다.

기브스 샘플링을 통한 각 뉴런의 상태 갱신은 다음과 같은 조건부 확률로 이루어집니다.

$$P(s_i = 1 | s_{-i}) = \sigma \left( \sum_j w_{ij} s_j + b_i \right)$$

- $P(s_i = 1 | s_{-i})$ : 다른 뉴런들의 상태가 고정되었을 때, 뉴런  $i$ 가 1일 확률
- $\sigma(x) = \frac{1}{1+e^{-x}}$ : 시그모이드 함수

기브스 샘플링을 통해 시스템은 반복적으로 뉴런 상태를 업데이트하며, 이 과정을 통해 네트워크는 에너지 최소화 상태에 도달하게 됩니다. 이 상태에서 네트워크는 학습한 확률 분포를

기반으로 데이터의 특성을 모델링할 수 있게 됩니다.

## 5. 볼츠만 머신의 학습

볼츠만 머신의 학습은 네트워크의 가중치와 바이어스를 조정하여 입력 데이터의 분포를 반영하는 과정입니다. 일반적으로 볼츠만 머신의 학습은 기브스 샘플링을 통해 각 뉴런 상태를 업데이트한 후, 에너지를 최소화하도록 가중치를 조정하는 과정을 포함합니다.

## 대조 발산 (Contrastive Divergence, CD)

볼츠만 머신 학습의 주요 알고리즘 중 하나인 **\*\*대조 발산(CD)\*\***은 제한된 볼츠만 머신(RBM)에서 주로 사용됩니다. 이는 가중치 업데이트를 위한 효율적인 학습 방법으로, 다음과 같은 과정을 따릅니다.

1. **포워드 패스**: 가시 노드에 데이터를 입력하여 숨겨진 노드의 초기 상태를 계산합니다.
2. **재구성(Reconstruction)**: 숨겨진 노드의 상태를 기반으로 가시 노드를 재구성합니다.
3. **가중치 업데이트**: 원본 데이터와 재구성된 데이터 간의 차이에 기반하여 가중치를 업데이트합니다.

CD는 기브스 샘플링을 간소화하여 볼츠만 머신의 학습을 가속화하며, 이 과정을 여러 번 반복하여 가중치가 데이터 분포에 맞게 수렴하도록 합니다.

## 6. 볼츠만 머신의 변형: 제한된 볼츠만 머신(RBM)

기본적인 볼츠만 머신은 계산 복잡도가 매우 높으며, 모든 노드가 서로 연결된 구조로 인해 학습이 매우 어렵습니다. 이를 개선한 모델이 바로 **\*\*제한된 볼츠만 머신(RBM)\*\***입니다.

## 제한된 볼츠만 머신의 구조

RBM은 가시 노드와 숨겨진 노드 간에만 연결된 구조를 가지며, **가시 노드 간, 숨겨진 노드 간의 연결을 제거**한 모델입니다. 이 구조는 학습의 효율성을 높여주며, 대조 발산 알고리즘을 통해 비교적 빠르게 학습할 수 있습니다. RBM은 특히 **심층 신경망**에서 특징 추출 또는 사전 학습 단계로 많이 사용됩니다.

## 7. 볼츠만 머신의 응용

볼츠만 머신과 그 변형 모델들은 여러 분야에서 응용되고 있습니다.

- **추천 시스템**: 사용자의 과거 데이터 패턴을 학습하여 맞춤형 추천을 제공합니다.
- **이미지 복원**: 손상된 이미지의 패턴을 학습하여 원본 이미지로 복원합니다.
- **차원 축소 및 특징 추출**: RBM을 사용하여 데이터의 중요한 특성을 추출하고, 이를 고차원 데이터의 차원 축소에 활용합니다.

- **최적화 문제:** 볼츠만 머신은 조합 최적화 문제에서 전역 최적해를 찾기 위해 사용될 수 있습니다. 예를 들어, 여행자 문제(TSP)와 같은 최적화 문제에서 상태를 샘플링하여 최적의 해를 찾습니다.