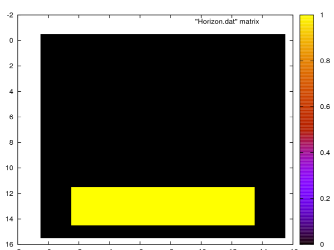
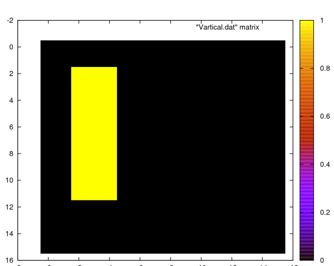
Introduction

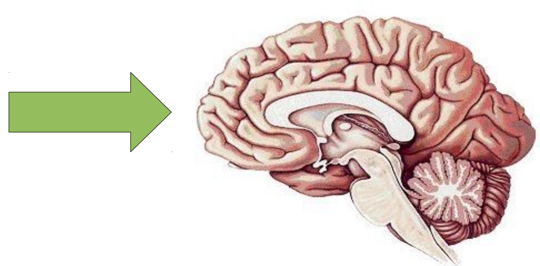
Boltzmann Machine

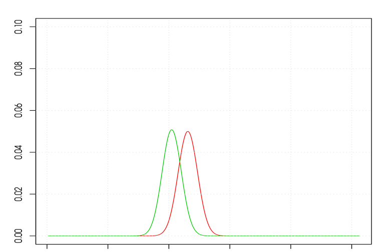
boltzmann machineの目的(1)

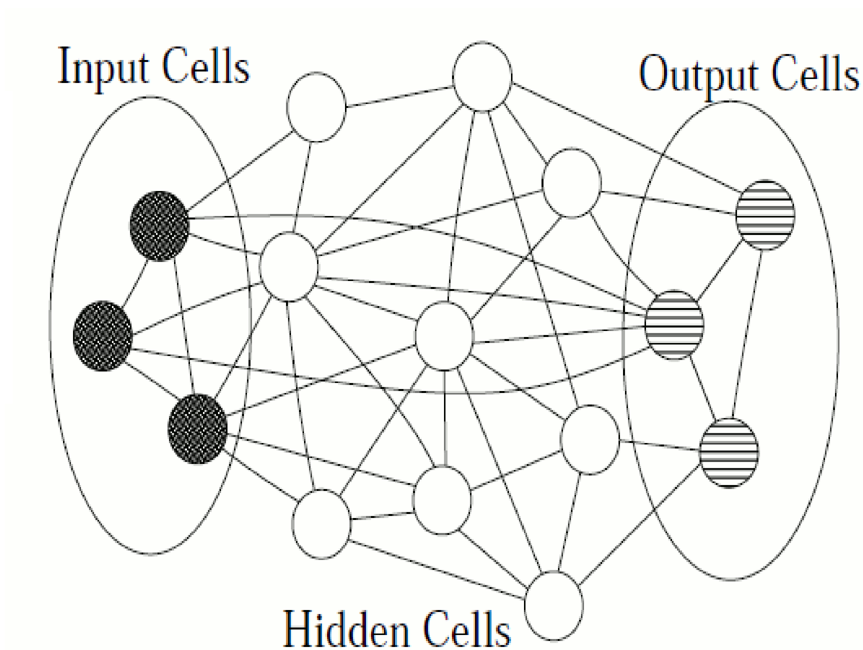
- 神経回路網を模した数理モデル
  - hopfield netなどのコネクショニストモデルの1つ
- 環境からの信号が入力素子 $\mathbf{v}$ へ
  この環境の確率構造Pを神経回路網Wへ写し取る



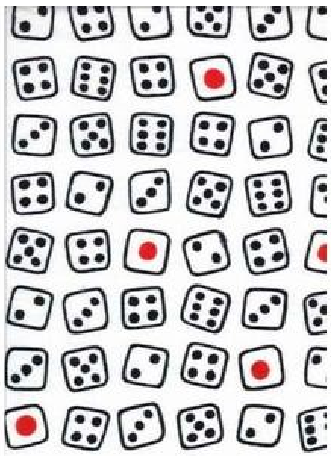




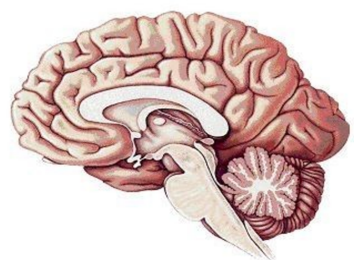


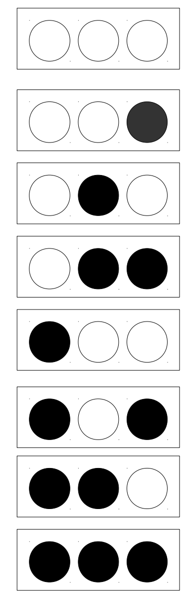


boltzmann machineの目的(2)

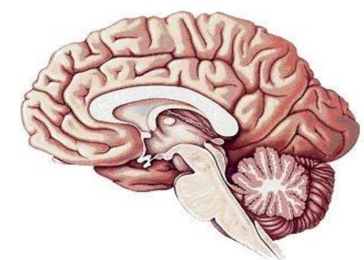


事象





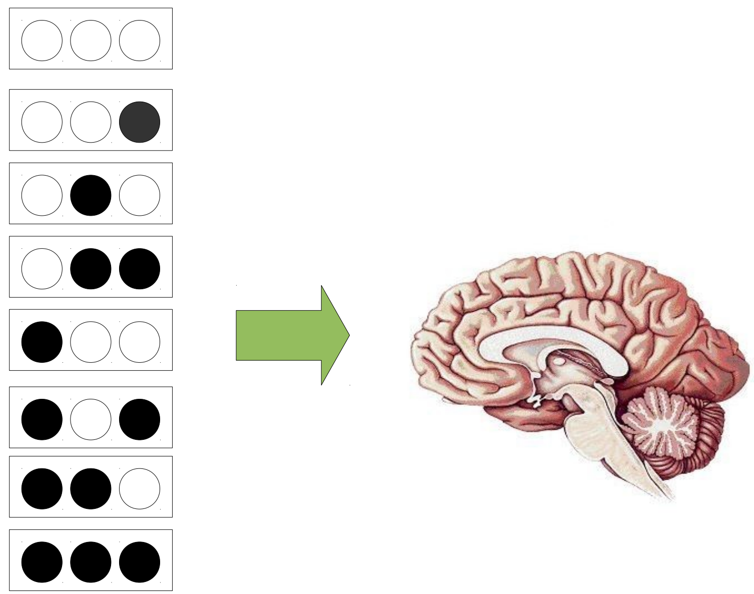
事象



- 神経回路網が置かれている環境の中で事象は確率的に起こる
- 環境での確率的な構造を回路網に写し取りたい
- このとき事象は{0,1}の2値パターンの組合せで表現される

boltzmann machineの目的(3)

P　　　　P’



KL情報量　$$G = \sum_{\alpha} P(V_\alpha) \ln \frac{P(V_\alpha)}{P'(V_\alpha)} \qquad (8)$$

- KL情報量
  - 環境のもつ確率構造Pと神経回路網Wを実行したときの確率P’の近さを測る
- この値を微分して勾配法を用いることで修正する

確率構造を写し取れる仕組み

- 環境がもつ確率構造 pij
- 回路網を自由に走らせることでえられる p'ij
- PとP'のKL情報量を目的関数Gとする
- 目的関数G(P||P')を結合係数wijで微分

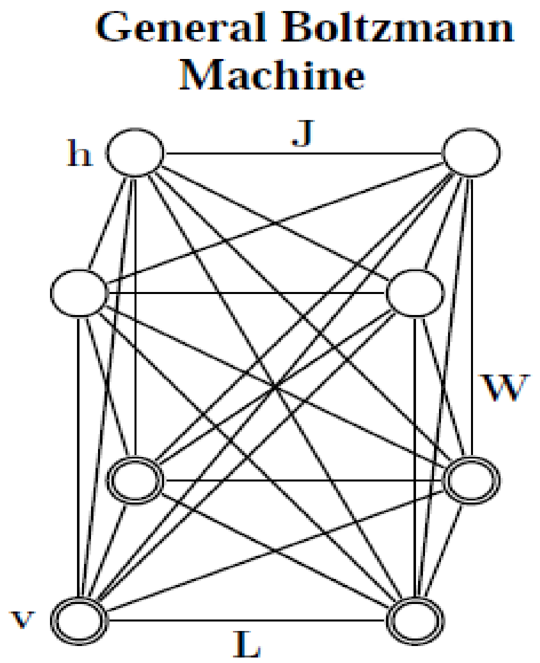
$$\frac{\partial G}{\partial w_{ij}} = -\frac{1}{T}(p_{ij} - p'_{ij}) \tag{9}$$

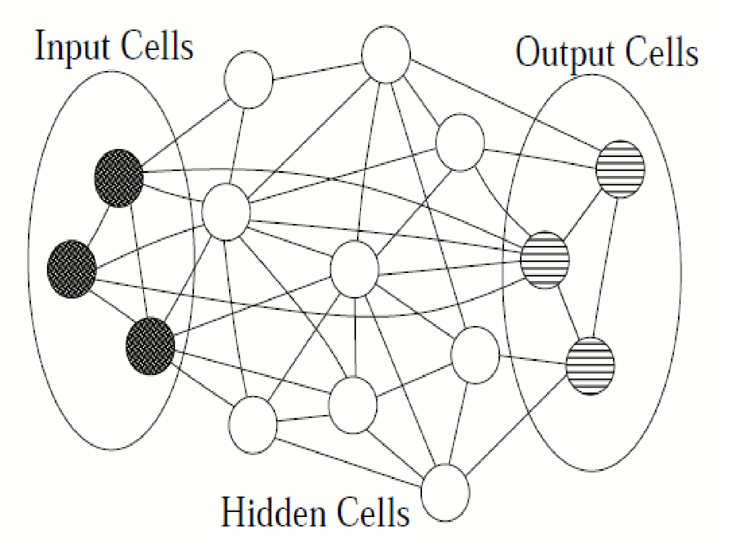
$$\Delta w_{ij} = \epsilon(p_{ij} - p'_{ij}) \tag{10}$$

- 結合係数を少しずつ修正していく

$$W^{t+1} = W^t + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^n (\mu^n)^\top - \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^\top \right)$$

神経回路網の状態



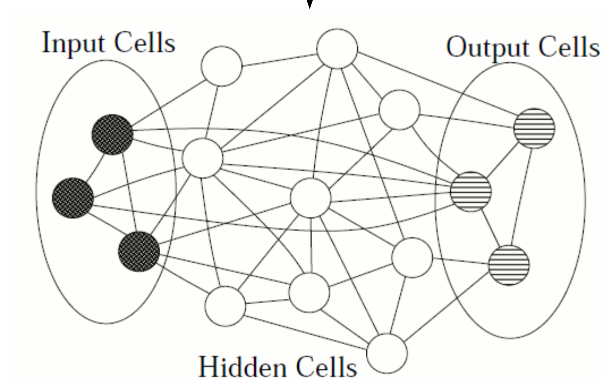


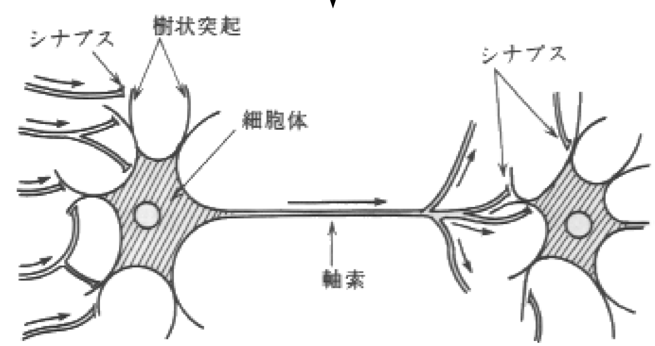
$$E(\mathbf{v}, \mathbf{h}; \theta) = -\frac{1}{2}\mathbf{v}^\top \mathbf{L}\mathbf{v} - \frac{1}{2}\mathbf{h}^\top \mathbf{J}\mathbf{h} - \mathbf{v}^\top \mathbf{W}\mathbf{h}, \qquad (1)$$

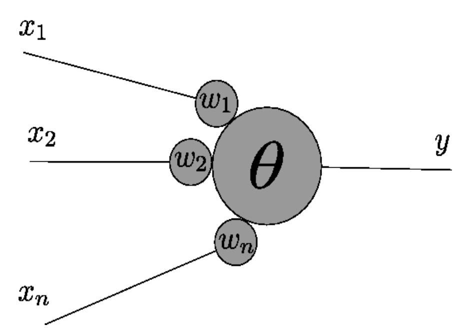
- 回路状態はエネルギー関数で表される

エネルギー関数

$$E = -\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i$$



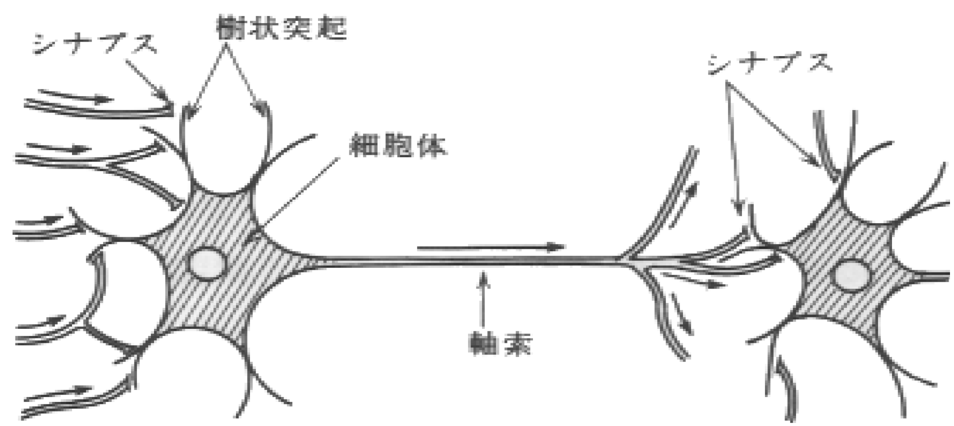


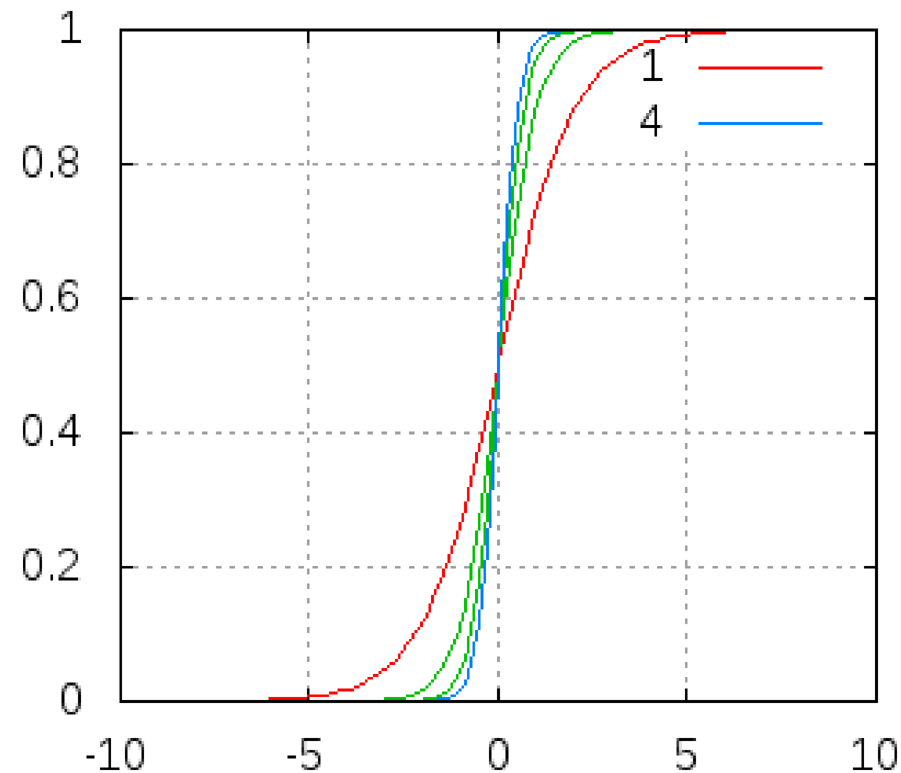


- エネルギー関数
  - 重み付き総和
  - 回路W上の各素子sが同時に1になったとき減少

神経細胞の発火確率

$$\sigma(x) = 1/(1+\exp(-x))$$

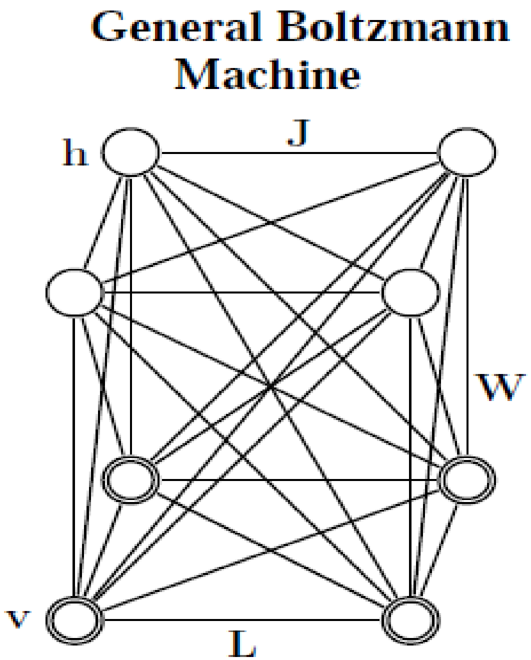




- 非線形関数では微分できない
- 閾値関数を近似する微分可能な関数が必要

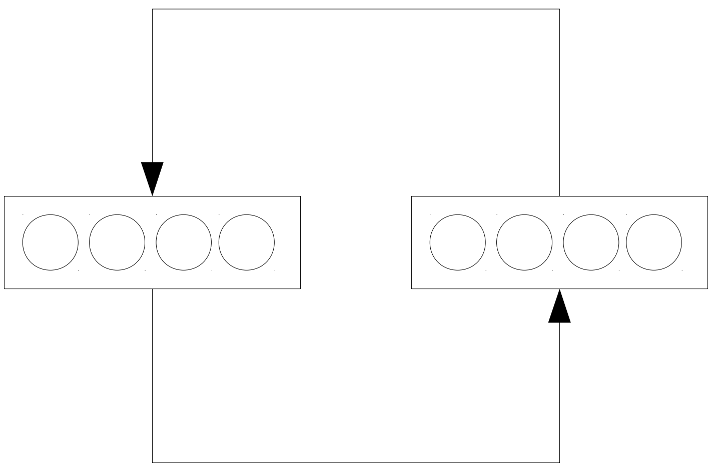
$$p(\mathbf{v};\theta) = \frac{p^*(\mathbf{v};\theta)}{Z(\theta)} = \frac{1}{Z(\theta)} \sum_h \exp\left(-E(\mathbf{v},\mathbf{h};\theta)\right),$$

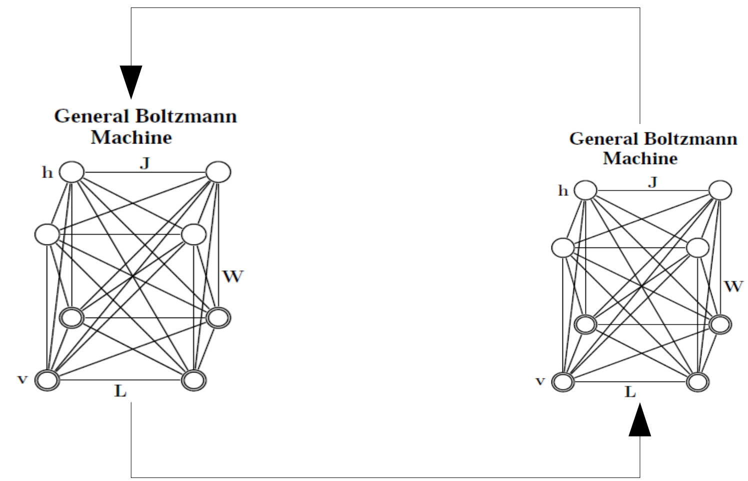
$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp\left(-E(\mathbf{v},\mathbf{h};\theta)\right), \quad (3)$$



- 回路状態は状態数 2^n 個のマルコフ連鎖
- 定常分布はユニークでボルツマン分布となる

マルコフ連鎖



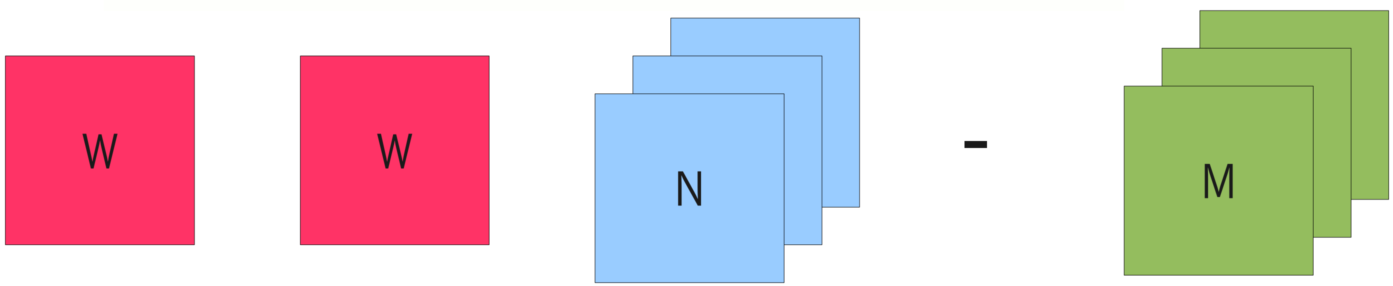


- 回路状態は状態数 2^n 個のマルコフ連鎖
- 定常分布はユニークに定まり, ボルツマン分布となる

神経回路網の更新則

データ依存期待値　　モデル依存期待値

$$W^{t+1} = W^t + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^n (\mu^n)^\top - \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^\top \right)$$



結合状態の更新則

学習係数　データ依存期待値　モデル依存期待値

$$\begin{aligned}
\Delta \mathbf{W} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{h}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{h}^{\top}] \right), \qquad (6) \\
\Delta \mathbf{L} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{v}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{v}^{\top}] \right), \\
\Delta \mathbf{J} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{h}\mathbf{h}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{h}\mathbf{h}^{\top}] \right),
\end{aligned}$$

ここをどうやって推定するか?

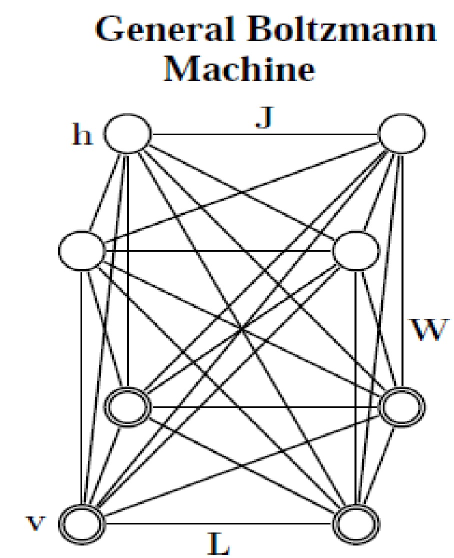
Variational approach

Stochastic approximation procedure

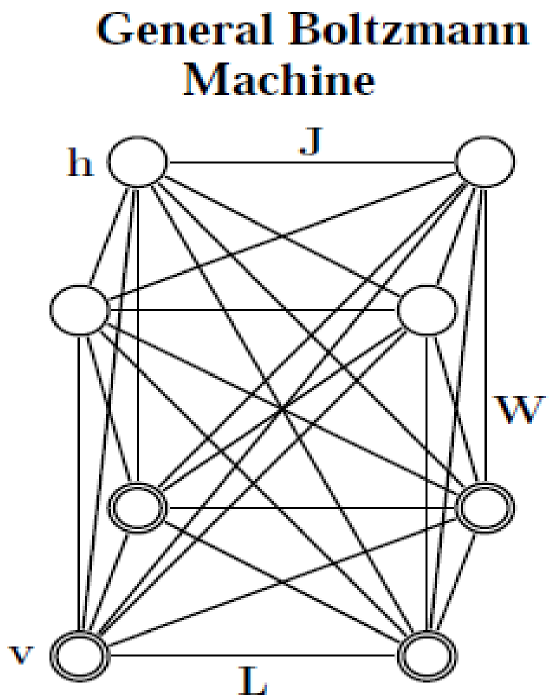
Boltzmann machineのアルゴリズム

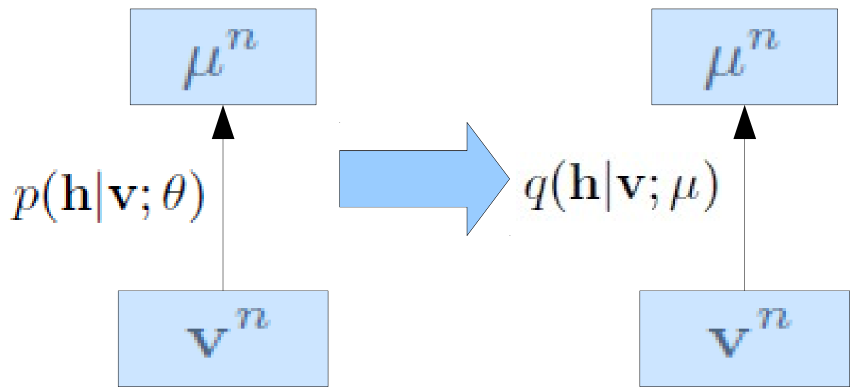
$$W^{t+1} = W^{t} + \alpha_t\left(\overbrace{\frac{1}{N}\sum_{n=1}^{N}\mathrm{v}^{n}(\mu^{n})^{\top}}^{\text{データ依存期待値}} - \overbrace{\frac{1}{M}\sum_{m=1}^{M}\tilde{\mathrm{v}}^{t+1,m}(\tilde{\mathrm{h}}^{t+1,m})^{\top}}^{\text{モデル依存期待値}}\right)$$

- 訓練データ　　の期待値を求める
- モデルの期待値　　を求める
- 神経回路を更新



variational learning



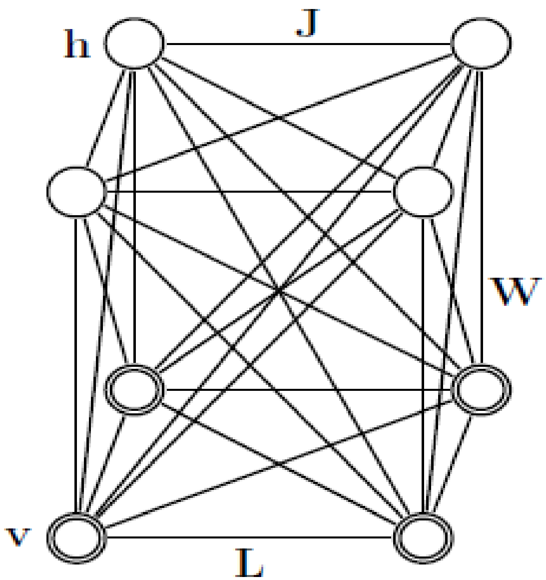


$$\begin{aligned}\ln p(\mathbf{v};\theta) &\geq \sum_{\mathbf{h}} q(\mathbf{h}|\mathbf{v};\mu)\ln p(\mathbf{v},\mathbf{h};\theta) + \mathcal{H}(q) \qquad (7)\\ &= \ln p(\mathbf{v};\theta) - KL[q(\mathbf{h}|\mathbf{v};\mu)||p(\mathbf{h}|\mathbf{v};\theta)],\end{aligned}$$

- 訓練データ　　の対数尤度を最大化
- 近似事後確率q(h;mu)と真の事後確率p(v; θ)を間のKL情報量を最小化
- パラメータは対数尤度上の下限の勾配により更新

variational learning

General Boltzmann Machine



$$W^{t+1} = W^t + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^n (\mu^n)^\top - \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^\top \right)$$

$$\mu^n \longleftarrow \mathbf{v}^n$$

mean field fixed point equation

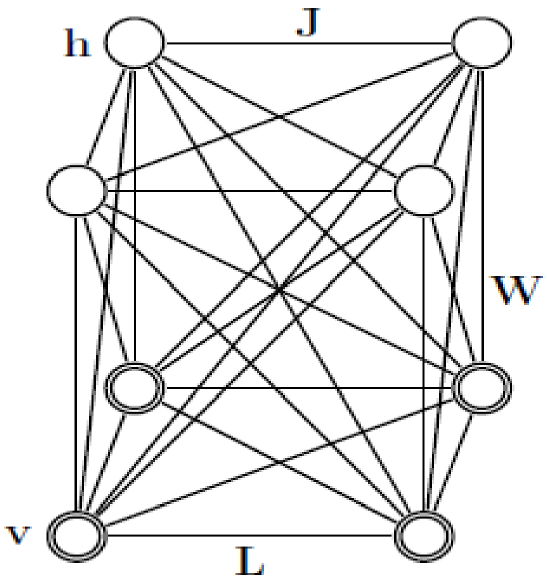
$$\mu_j \leftarrow \sigma\Big(\sum_i W_{ij} v_i + \sum_{m \backslash j} J_{mj} \mu_m \Big).$$

- データに依存する期待値
- 訓練データ $\mathbf{v}^n$ から変分パラメータ μ^n を獲る

variational learning

$$W^{t+1} = W^{t} + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^{n} \boxed{(\mu^{n})^{\top}} - \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^{\top} \right)$$

General Boltzmann Machine



$\boxed{\mu^{n}}$

$\uparrow$

$\boxed{\mathbf{v}^{n}}$

mean field fixed point equation

$$\mu_j \leftarrow \sigma\Big(\sum_i W_{ij} v_i + \sum_{m \backslash j} J_{mj} \mu_m\Big). \qquad (8)$$

$q(\mathbf{h};\mu) = \prod_{j=1}^{P} q(h_i)$, with $q(h_i = 1) = \mu_i$

- 訓練データ $\mathbf{v}^{n}$ から変分パラメータ μ^{n} を獲る

variational learning

$$q(\mathbf{h};\boxed{\mu}) = \prod_{j=1}^{P} q(h_i), \text{ with } q(h_i = 1) = \boxed{\mu_i}$$

$$\boxed{\mu^n} \longleftarrow \boxed{\mathbf{v}^n}$$

$$\mu_j \leftarrow \sigma\Big(\sum_i W_{ij} v_i + \sum_{m \backslash j} J_{mj}\mu_m\Big). \tag{8}$$

- naive mean-field approach
  - 完全に因数分解できるモデル
  - 単純な手続きで収束がはやい
  - 確率分布p(h)の近似値

結合状態の更新則

学習係数　データ依存期待値　モデル依存期待値

$$\Delta \mathbf{W} = \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{h}^\top] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{h}^\top] \right), \tag{6}$$

$$\Delta \mathbf{L} = \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{v}^\top] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{v}^\top] \right),$$

$$\Delta \mathbf{J} = \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{h}\mathbf{h}^\top] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{h}\mathbf{h}^\top] \right),$$

ここをどうやって推定するか?

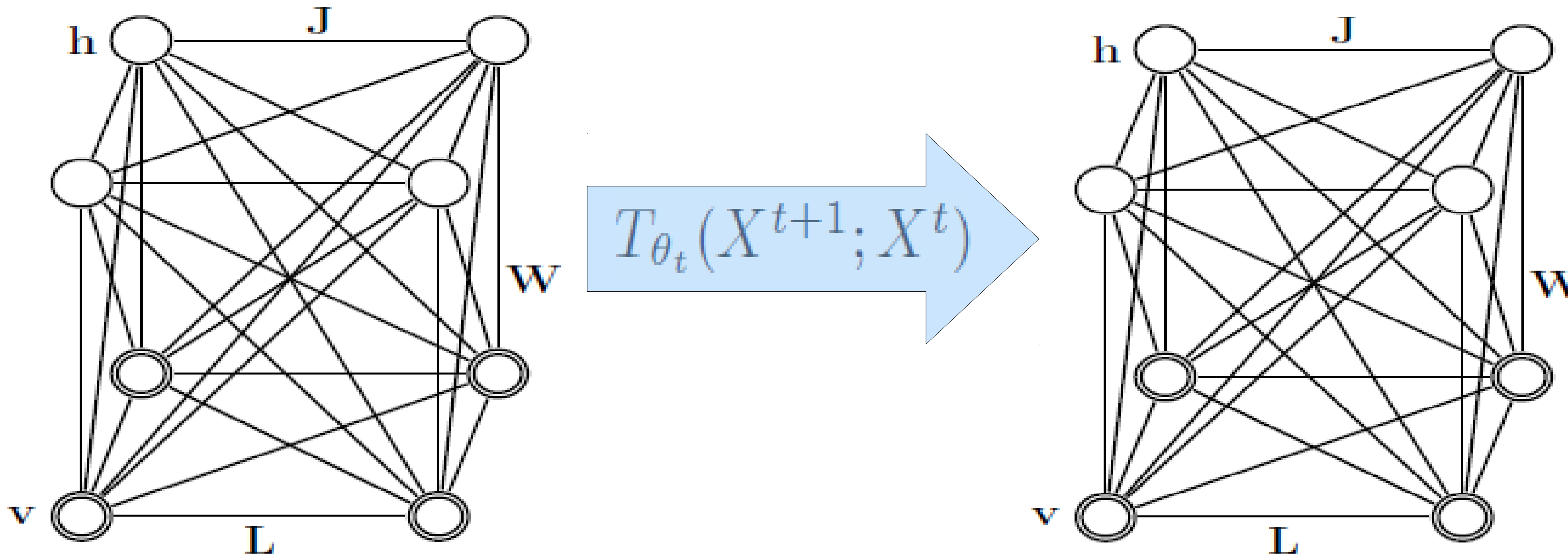
variational approach

stochastic approximation procedure

stochastic approximation procedure

$(\tilde{\mathbf{v}}^{t,m}, \tilde{\mathbf{h}}^{t,m})$.

$(\tilde{\mathbf{v}}^{t+1,m}, \tilde{\mathbf{h}}^{t+1,m})$



- 現在の状態 $(\tilde{\mathrm{v}}^{t,m}, \tilde{\mathrm{h}}^{t,m})$. から次の状態 $(\tilde{\mathrm{v}}^{t+1,m}, \tilde{\mathrm{h}}^{t+1,m})$ を獲る

結合状態の更新則

学習係数　データ依存期待値　モデル依存期待値

$$
\begin{aligned}
\Delta \mathbf{W} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{h}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{h}^{\top}] \right), \qquad (6)\\
\Delta \mathbf{L} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{v}\mathbf{v}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{v}\mathbf{v}^{\top}] \right),\\
\Delta \mathbf{J} &= \alpha \left(\mathrm{E}_{P_{\mathrm{data}}}[\mathbf{h}\mathbf{h}^{\top}] - \mathrm{E}_{P_{\mathrm{model}}}[\mathbf{h}\mathbf{h}^{\top}] \right),
\end{aligned}
$$

ここをどうやって推定するか?

Variational approach

Stochastic approximation procedure

stochastic approximation procedure

transition operator

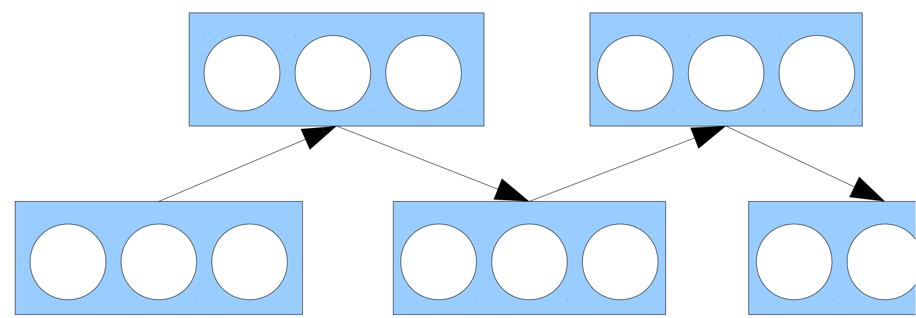
$T_{\theta_t}(X^{t+1}; X^t)$

中身は？

k-step gibbs sampling

$$p(h_j = 1|\mathbf{v}, \mathbf{h}_{-j}) = \sigma\Big(\sum_{i=1}^{D} W_{ij} v_i + \sum_{m=1\backslash j}^{P} J_{jm} h_j\Big), \quad (4)$$

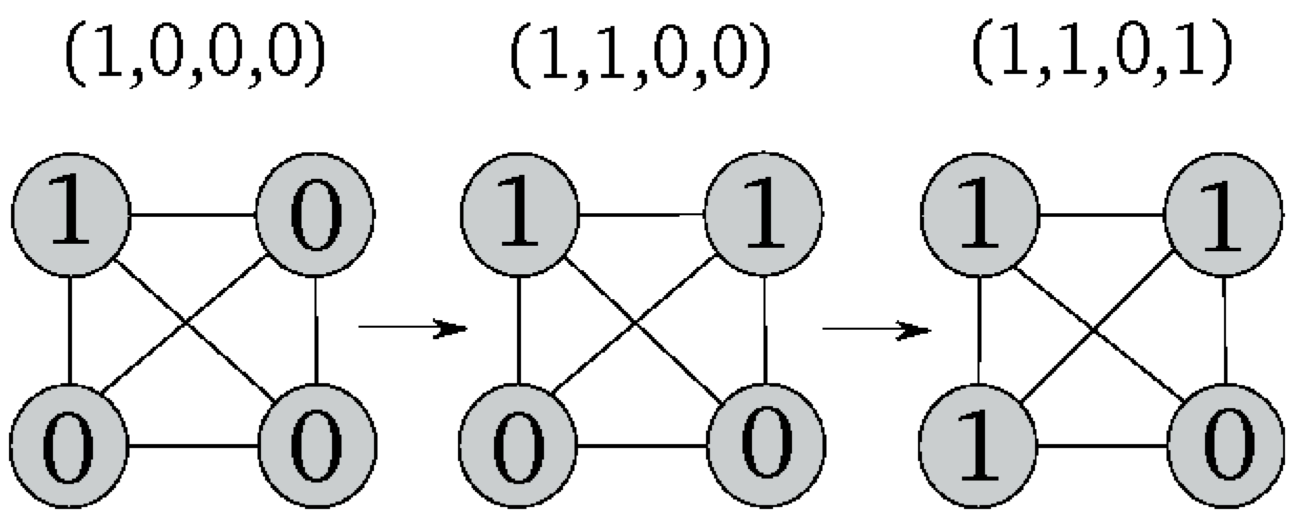
$$p(v_i = 1|\mathbf{h}, \mathbf{v}_{-i}) = \sigma\Big(\sum_{j=1}^{P} W_{ij} h_j + \sum_{k=1\backslash i}^{D} L_{ik} v_j\Big), \quad (5)$$



- 可視層vと隠れ層hを交互に遷移

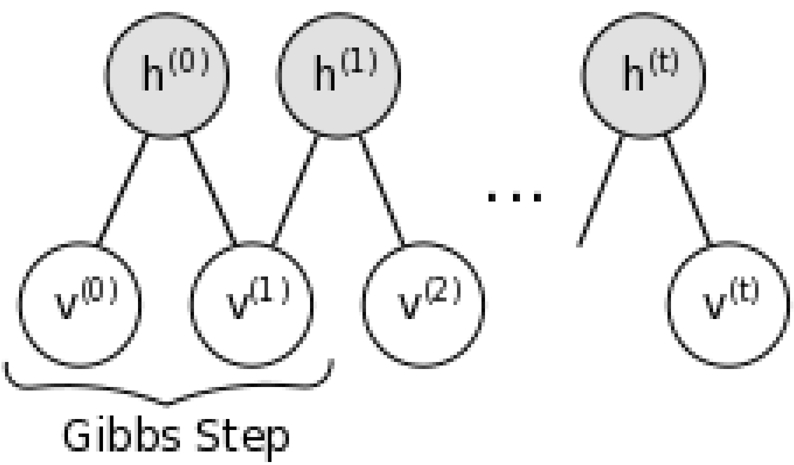
k-step gibbs sampling
(全結合可視素子のみの場合)

$$p(x_i = 1 | x_{i-1}) = \frac{1}{1+\exp(-u_i)}$$

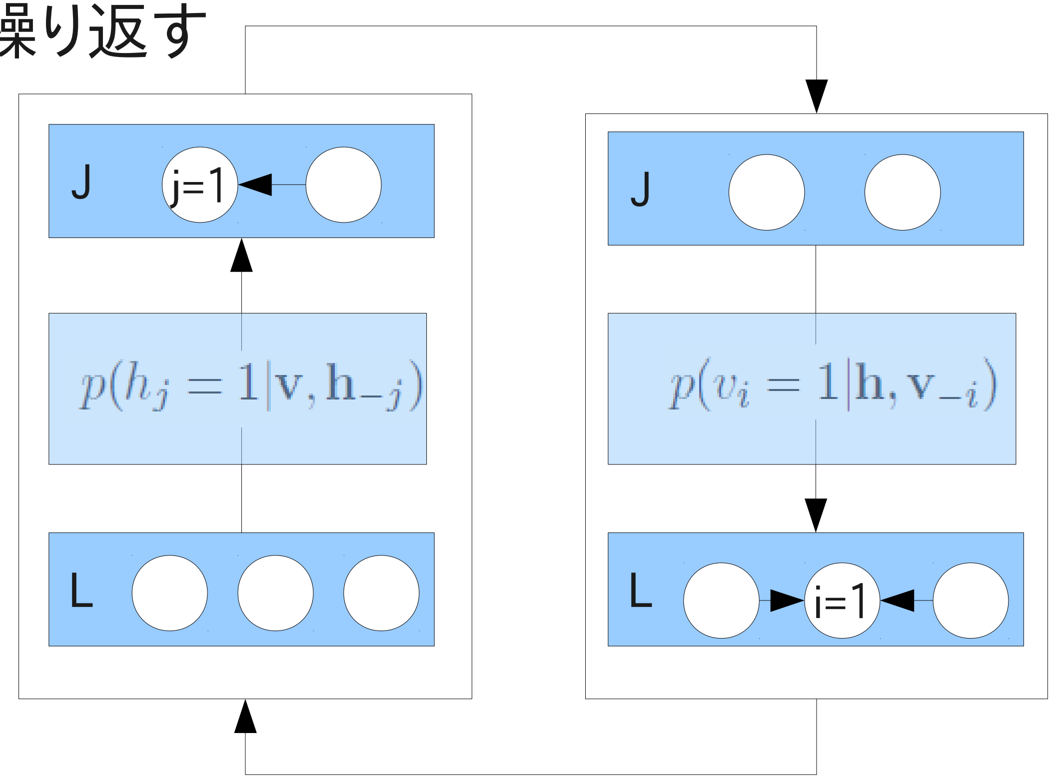


- 1つの神経細胞を選択
- 確率にしたがい発火させる

k-step gibbs sampling
(可視素子v,隠れ素子hを持つとき)



k回繰り返す



- 隠れ層hの素子をランダムに1つ選択. eq(4)の確率で発火
- 可視層vの素子をランダムに1つ選択. eq(5)で確率で発火

$$p(h_j = 1|\mathbf{v}, \mathbf{h}_{-j}) = \sigma(\sum_{i=1}^{D} W_{ij} v_i + \sum_{m=1\backslash j}^{P} J_{jm} h_j), \quad (4)$$

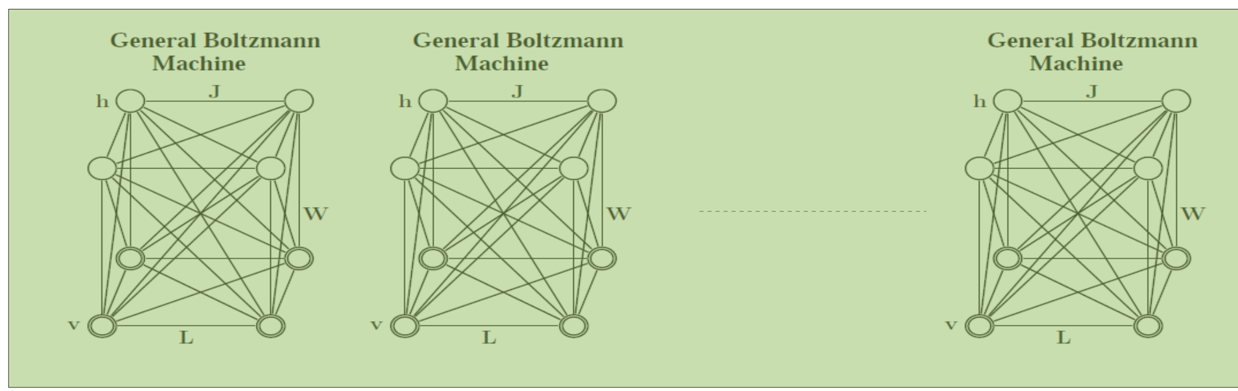
$$p(v_i = 1|\mathbf{h}, \mathbf{v}_{-i}) = \sigma(\sum_{j=1}^{P} W_{ij} h_j + \sum_{k=1\backslash i}^{D} L_{ik} v_j), \quad (5)$$

stochastic approximation procedure

$$W^{t+1} = W^t + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^n (\mu^n)^\top - \boxed{\frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^\top} \right)$$

M fantasy particle

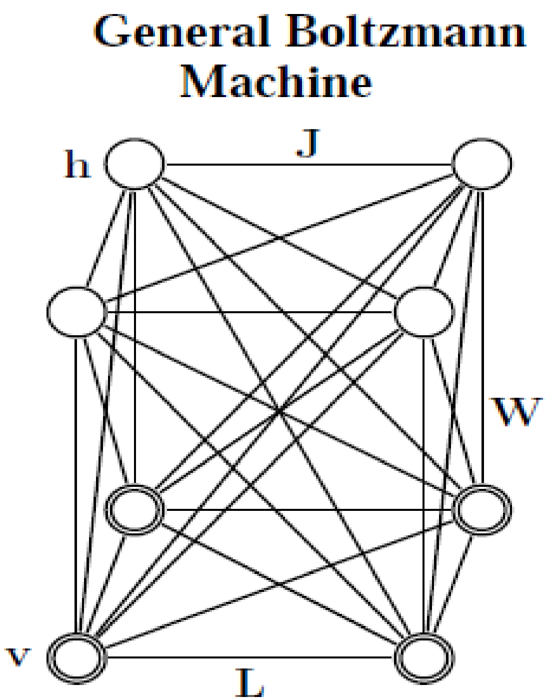
$$(\tilde{\mathbf{v}}^{t,m}, \tilde{\mathbf{h}}^{t,m}).$$



- Robbins-Monro型のアルゴリズム
- モデルパラメータの期待値をとる

Robbins-Monroアルゴリズム

神経回路網の更新



$$W^{t+1} = W^t + \alpha_t \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{v}^n (\mu^n)^\top - \frac{1}{M} \sum_{m=1}^{M} \tilde{\mathbf{v}}^{t+1,m} (\tilde{\mathbf{h}}^{t+1,m})^\top \right)$$