

DBMの紹介

2010/12/9
T1076006 塩貝亮宇

資料中での略記

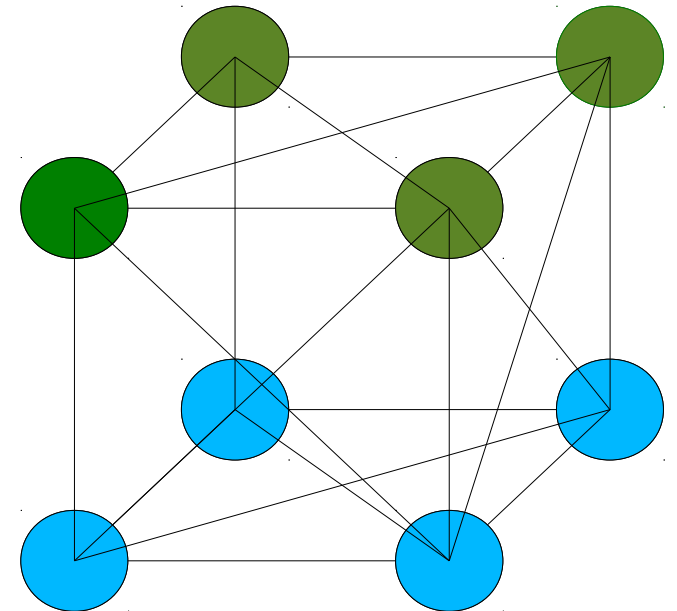
- BM → Boltzmann Machine
- GBM → General Boltzmann Machine
- RBM → Restricted Boltzmann Machine
- DBN → Deep Belief Network
- DBM → Deep Boltzmann Machine

発表の流れ

- 通常のBMについて紹介
 - ギブスサンプリング
 - ボルツマン分布
 - 分配関数
 - 特徴
- 学習手続きに関して
 - BMの学習方法
 -

ボルツマン機械

- Hinton, Sejnowski (1983)らによって提案された確率的に動作するニューラルネットワーク
- 与えられた信号に内在する確率構造の抽出する目的として考えられたモデル
- 全結合された n 個の素子
- 対称結合 $W_{ij} = W_{ji}$
- 自己結合は無し $W_{ii} = 0$
- 可視素子 v , 隠れ素子 h に分けて考えられる
 - $v \in \{0, 1\}^D$
 - $h \in \{0, 1\}^P$
- 状態数 2^n の有限状態のマルコフ連鎖



GBMのアーキテクチャ

- 素子間の結合
 - W: 可視-隠れ
 - L: 可視-可視
 - J: 隠れ-隠れ

- モデルパラメータ $\theta = \{W, L, J\}$

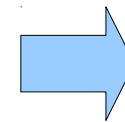
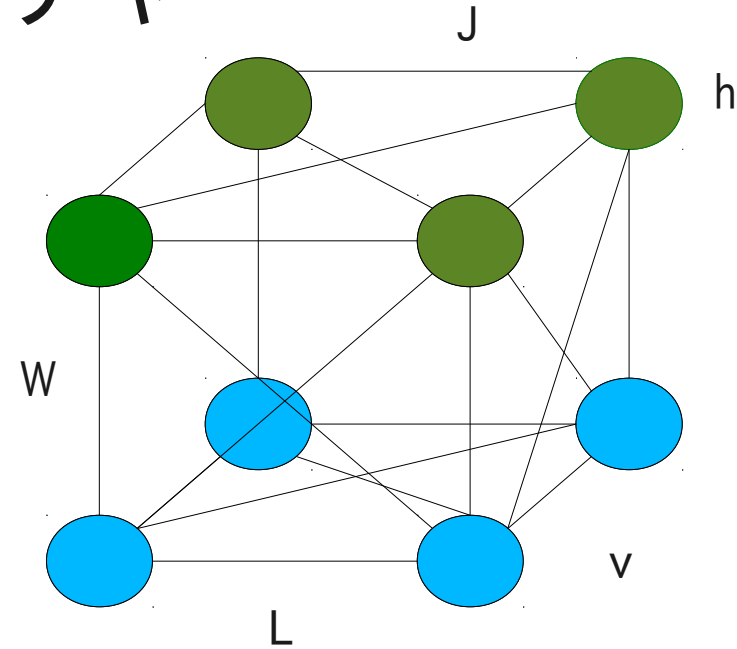
- エネルギー関数 $E(\mathbf{v}, \mathbf{h}; \theta) = -\frac{1}{2} \mathbf{v}^T L \mathbf{v} - \frac{1}{2} \mathbf{h}^T J \mathbf{h} - \mathbf{v}^T W \mathbf{h}$

- モデルに代入される可視素子の確率

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

- 分配関数 $Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$

- 条件付き分布 $p(h_j=1 | \mathbf{v}, \mathbf{h}_{-j}) = \sigma\left(\sum_{i=1}^D W_{ij} v_i + \sum_{m=1 \setminus j}^P J_{jm} h_m\right)$
 $p(v_i=1 | \mathbf{h}, \mathbf{v}_{-i}) = \sigma\left(\sum_{j=1}^P W_{ij} h_j + \sum_{k=1 \setminus i}^D L_{ik} v_k\right)$



Gibbs Sampling
に用いられる

ギブスサンプリング

- 二つ以上の確率変数の結合分布から統計標本の配列を生成するアルゴリズム
- この配列は目標分布の近似として用いられたり積分を計算するのに用いられる
- データ依存の期待値, モデルの期待値の両方を近似させるために使われている

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

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

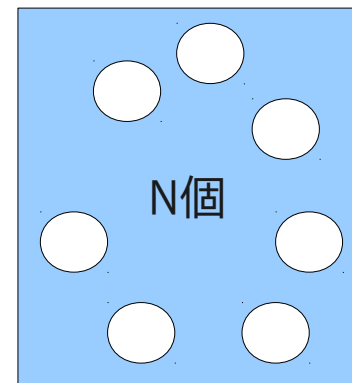
ボルツマン分布

- 統計物理学では、現象を確率的システムとして扱い、個々の要素の振る舞いに対して決定的な予測をせず、要素の振る舞いを代表とする何らかの変数を考えて、時間的平均や集合的平均を予測する。
- 任意に指定されたエネルギー範囲に対して、その範囲に存在する分子の頻度を与え得ることから、エネルギー値に対する「分子の分布」と呼ばれている。
- この式はエネルギーの低いところに多くの分子が分布する傾向にある

$$P(E) \propto \frac{\exp(-\frac{E}{kT})}{\int_{-\infty}^{\infty} \exp(-\frac{E}{kT}) dE}$$



ボルツマン分布



分配関数

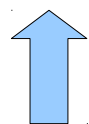
- ある系の物理量の統計集団的平均を計算する際に用いられる規格化定数
 - 単に分配関数と呼ぶとき,カノニカル分布における
 - 分配関数を指し, Z で表す.
- カノニカル分布(正準分布)
 - 統計力学において微視的状态を表現する統計集団の一つである正準集団(カノニカルアンサンブル)に従う統計分布
 - カノニカル分布は熱浴との間でエネルギーを自由にやりとりできる系に対応する統計分布

$$Z(\theta) = \sum_v \sum_h \exp(-E(v, h; \theta))$$

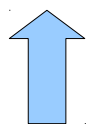
ボルツマン分布の特徴

- 実際に観測される確率分布を
最大にする確率分布

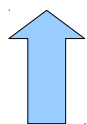
$$F_{\beta}(P_1, P_2, \dots, P_N) = -\sum_{i=1}^n P_i \log P_i - \beta \sum_{i=1}^n P_i \phi(i)$$



確率分布



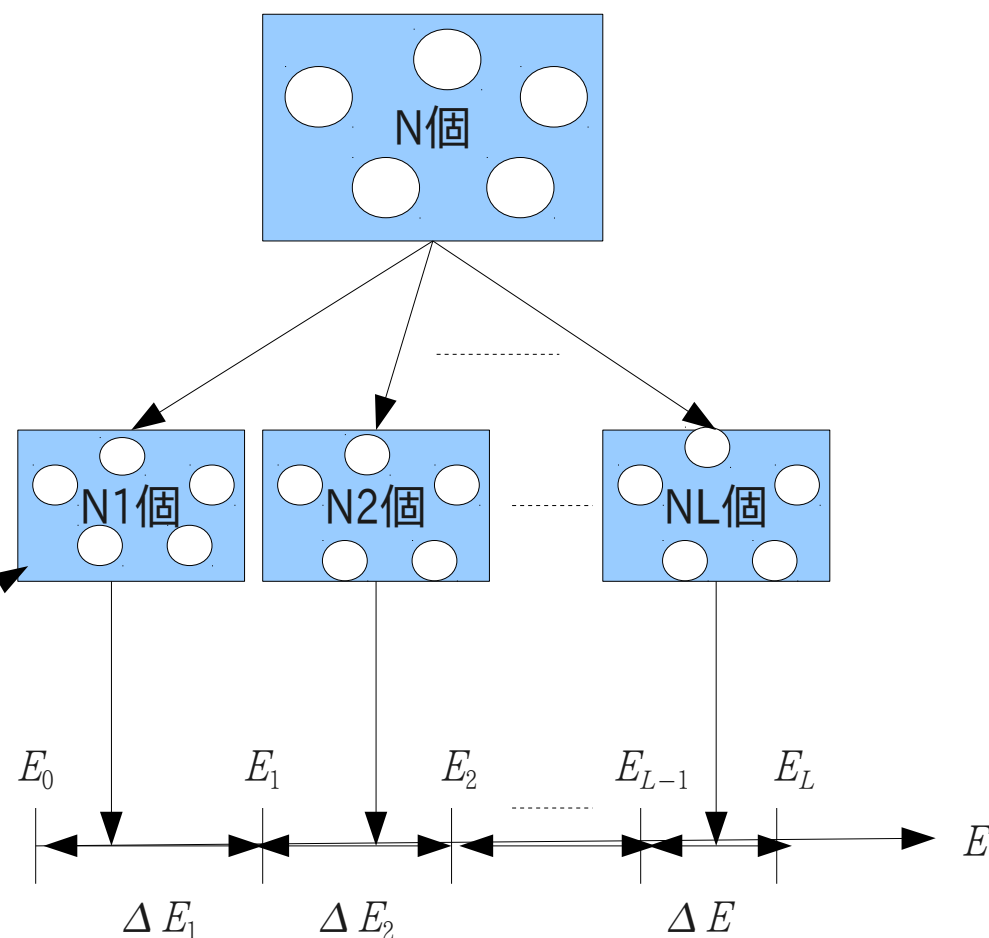
エントロピー



期待値

$$\sup_{P_1, P_2, \dots, P_n} F_{\beta}(P_1, P_2, \dots, P_n) = \sup_{P_1, \dots, P_n} \left(-\sum_{i=1}^n P_i \log P_i - \beta \sum_{i=1}^n P_i \phi(i) \right)$$

正準集団



ボルツマン分布との対応

- 気体のエネルギー総量

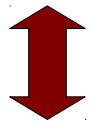
$$E_{total} = \sum_{i=0}^{L-1} N_i E_i$$



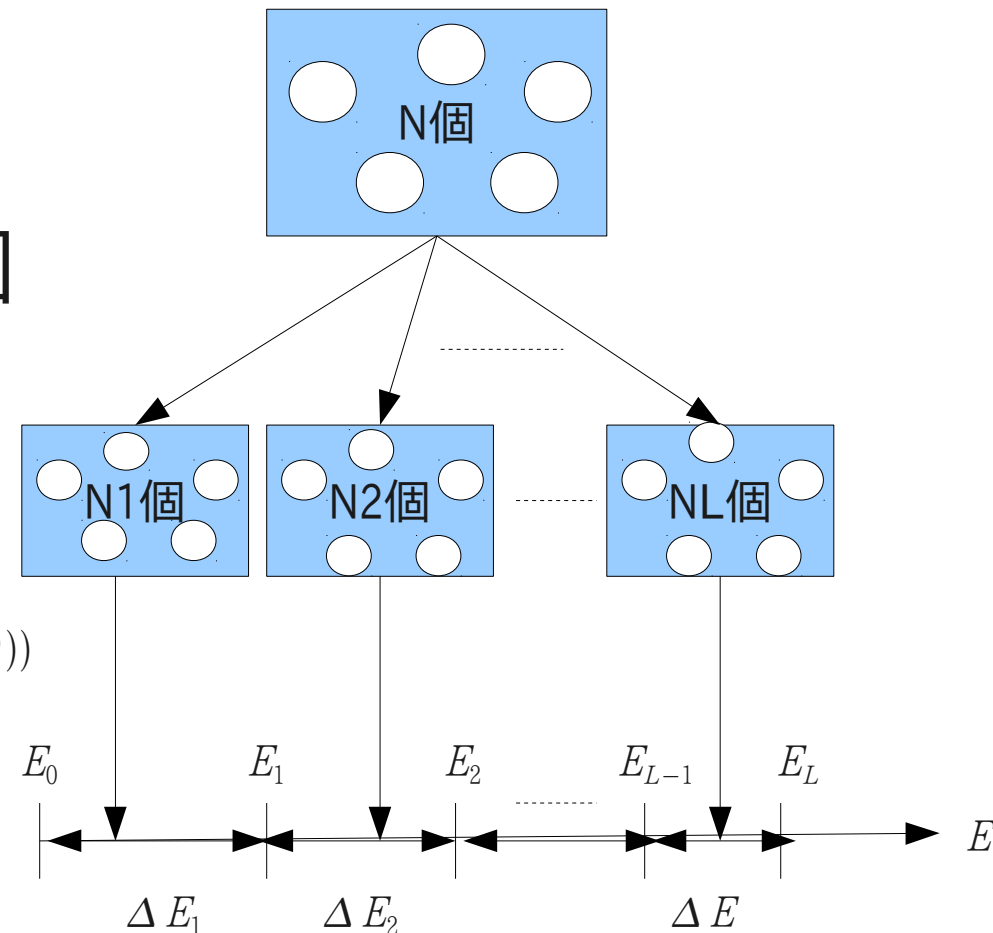
$$Z(\theta) = \sum_v \sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$$

- 容器内の気体分子集団

$$p_i = A \exp\left(-\frac{E_i}{kT}\right)$$



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



学習手続き phase1(記銘)

- 与えられた例題に従い,入力素子を固定してBMを動作させる. 平衡状態において素子 i と j が同時に1となっている

$$p_{ij} = \sum_{\alpha, \beta} q(\alpha, \beta) E(x_i x_j | \alpha, \beta)$$

確率を計算する.

- $q(\alpha, \beta)$ は例題に基づく経験分布で, ϵ を適当な正の定数として, p_{ij} を用いて素子 i, j 間の結合係数 w_{ij} を

$$\Delta w_{ij} = \epsilon p_{ij}$$

- だけ強化する

学習手続き phase2(忘却)

- 入力素子は与えられた例題に従い固定.
- 出力素子は自由に動作させることにし,
平衡状態において

$$p_{ij}' = \sum_{\alpha} q(\alpha) E(x_i x_j | \alpha)$$

- を計算する. ただし, $q(\alpha)$ は
- 例題の入力のみ経験分布である.
- p_{ij}' に基づいて
素子 i, j 間の結合係数 w_{ij} を

$$\Delta w_{ij} = -\epsilon p_{ij}'$$

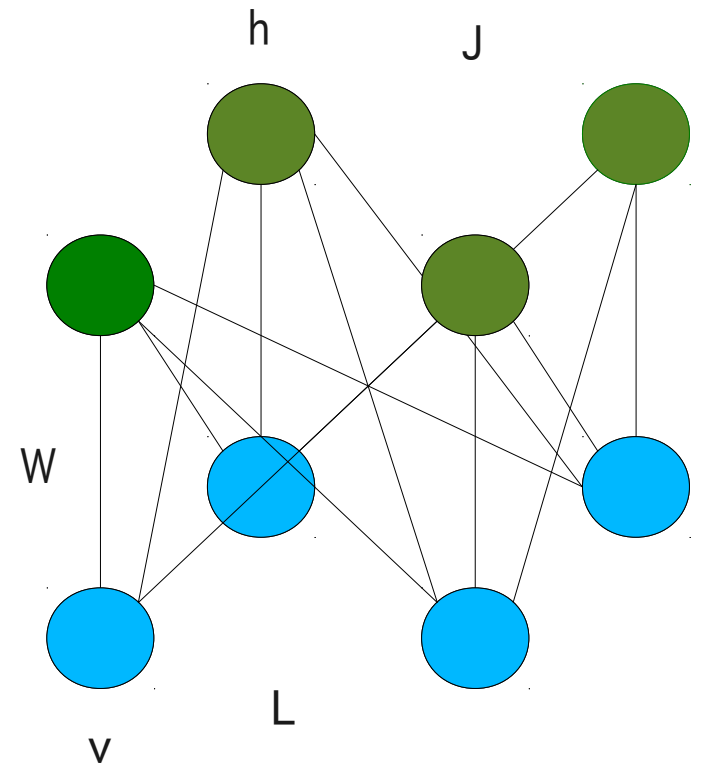
- だけ弱める

学習に関して

- 学習が進むと
 - ボルツマン分布は定常分布になっている
 - $\pi P = \pi$
 - 確率ベクトルに遷移行列を掛けてもが変化しない状態
- アルゴリズムの問題点
 - 定常分布に至るまでに時間がかかる
 - 特にモデルの期待値を推定するとき,
 - ギブス鎖は高次の多数モデルのエネルギー谷を
 - 探索する必要がある

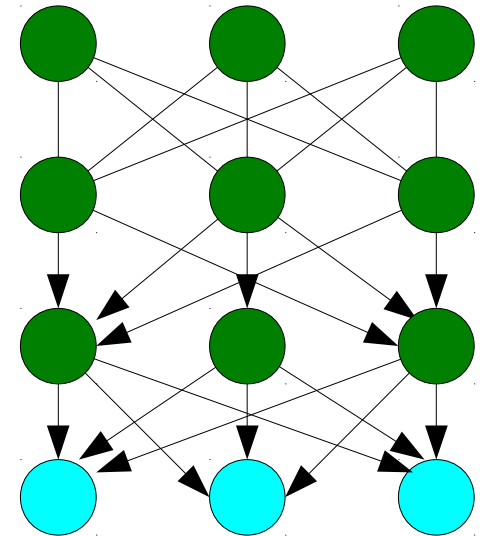
RBMのアーキテクチャ

- 素子間の結合
 - W : 可視-隠れ
 - L : 可視-可視
 - J : 隠れ-隠れ
- 可視層と隠れ層の結合がないモデル
 - $J=0, L=0$
- 一般的なBMより推定が正確
- 可視層の状態に依存する
- 正確な最尤推定学習は困難
- 効率的な学習アルゴリズムとして
Contrastive Divergence (Welling, Hinton, 2002)



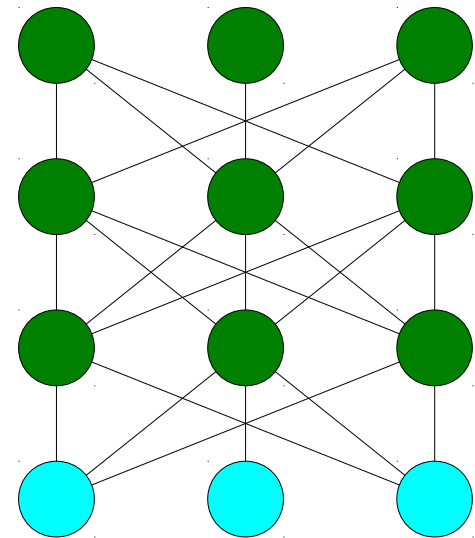
Deep Belief Network

- 最上位は2層の無向グラフ
- 他層は上から下を向いた有向グラフ
- 中間層の表現により物体や音声認識の問題を解ける可能性がある
- 教師無し学習でRBMの持つ層のスタックを同時に学習できる貪欲なアルゴリズムで学習を高速化
 - 事前学習として、回路を初期化する
 - BMのようにランダムに初期化するわけではない



Deep Boltzmann Machine

- 相互結合型 $W_{ij} = W_{ji}$
- ある層より下位の隠れ層の特徴間には高階と相関を持つ
- 中間層の表現により複雑さを増加
- 物体や音声認識の問題を解くための学習手法として考えられている
- 最大推定学習の近似について, GBMの手続きが適応可能
 - しかし, 学習時間は長く,
 - 可視層から離れるに従って増加する



シミュレーション条件

- 2つのボルツマンマシンの比較
 - 2層:h 500, 1000
 - 3層:h 500, 500, 1000
- 分配関数 βk は20,000
- データセットを小さいバッチに再分割
 - ミニバッチには100個のデータ
- fantasy particles
 - 100×100
 - 常に5回のgibbs data
- 初期の学習データは0.005
 - 徐々に0に向かって減少

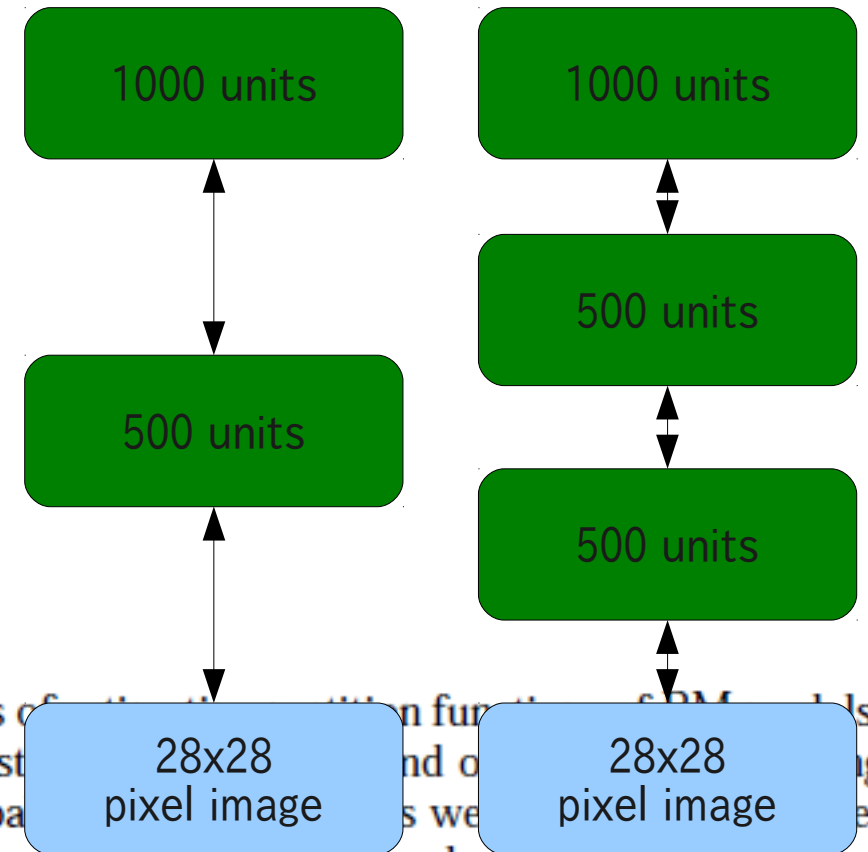


Table 1: Results of the simulation for the 2-layer and 3-layer BMs, along with the estimates of the partition function and the log-probability and test log-probability. The results were averaged over 100 AIS runs.

	Estimates		Avg. log-prob.	
	$\ln \hat{Z}$	$\ln (\hat{Z} \pm \hat{\sigma})$	Test	Train
2-layer BM	356.18	356.06, 356.29	-84.62	-83.61
3-layer BM	456.57	456.34, 456.75	-85.10	-84.49

- 2層DBM
 - 0.9と1.15百万のパラメータ
 - 全てのバイナリ状態を
-
- 標本を生成
-

MNIST データセット

- MNIST デジタル文字セット
 - 28×28画素の0～9の手書き文字
 - 60,000の訓練データ
 - 10,000のテスト画像

Training Samples



シミュレーション結果

(MNISTデータセット)

- 訓練データからランダムにサンプルを与えて初期化
- 100,000ステップのギブスサンプリングを実行した結果
- 貪欲な事前学習なしではMNISTデータセットをDBMの学習で成功させることはできなかった

2-layer BM

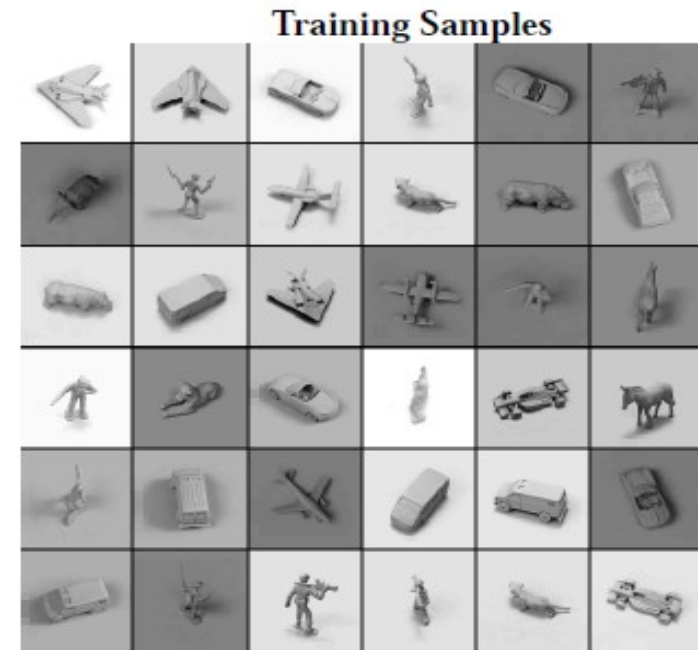


3-layer BM

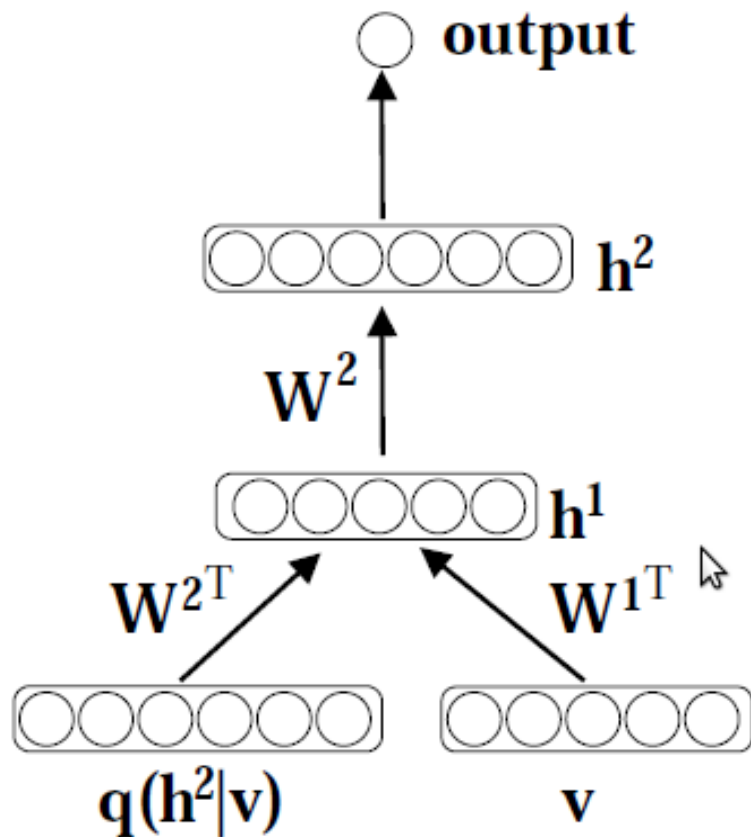


NORB 3D Object

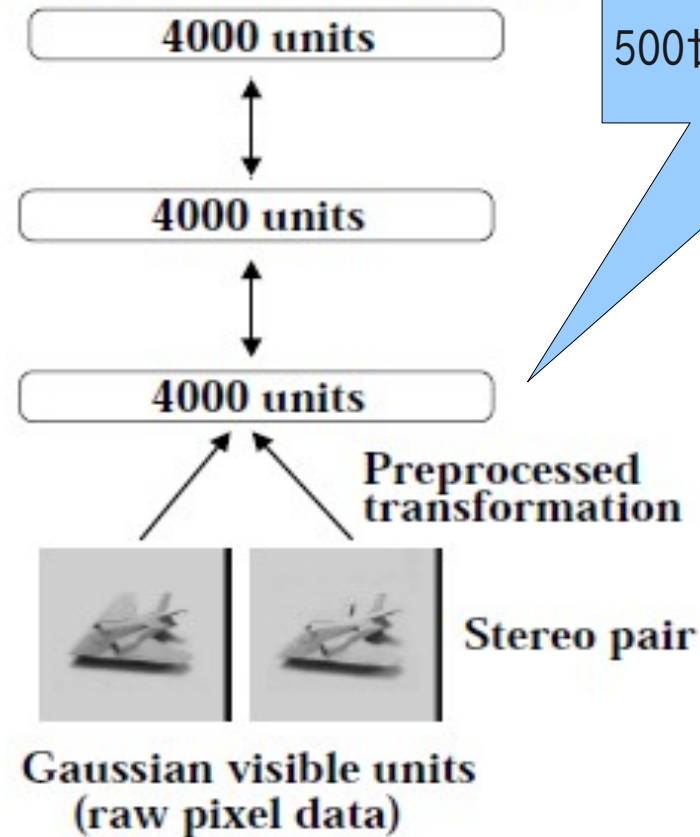
- 3次元の玩具の画像
 - 異なる視点,様々な光源のもとで撮影
 - 24,300ステレオイメージ
 - 5つのクラス
 - 車,トラック,飛行機,動物,人間
 - クラス毎に異なる10個の画像が用意
- 画像データ
 - 96×96 画素
 - $[0,255]$ のグレースケール値



NORBデータセットへの適応



Deep Boltzmann Machine



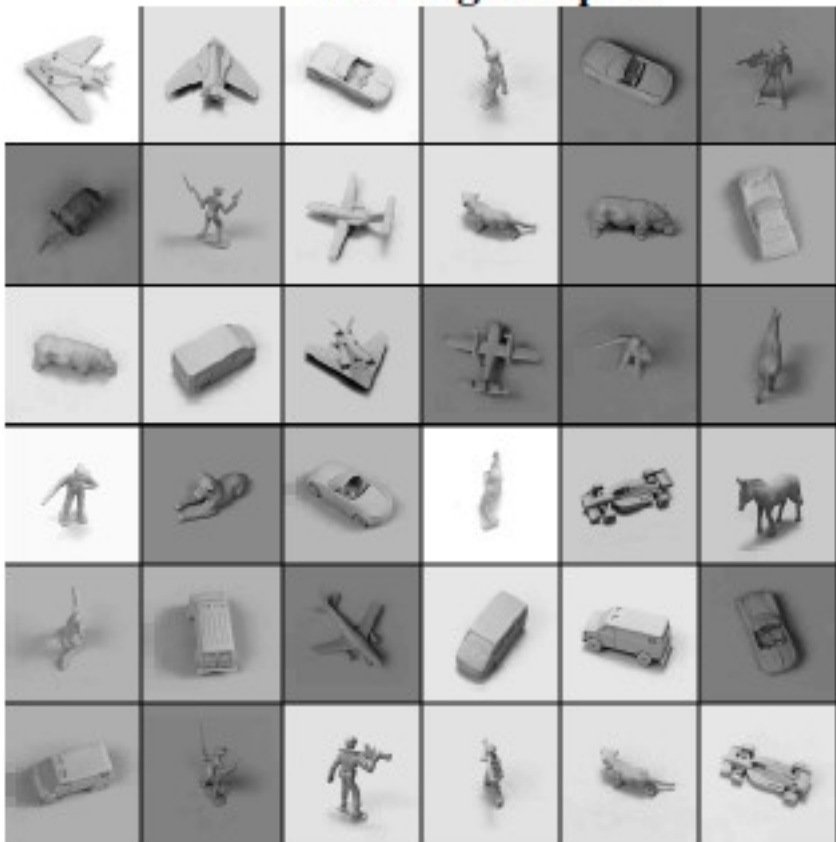
500世代についてCD学習

シミュレーション結果

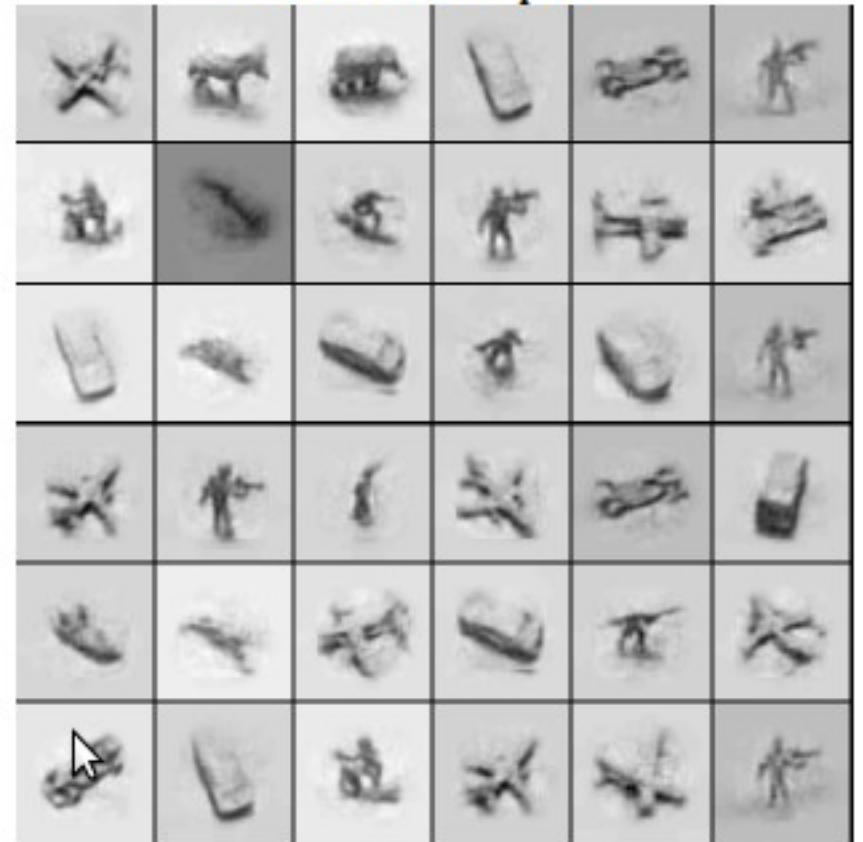
- クラス分けのエラー率
 - DBM(2層,隠れ素子4000) 10.8%
 - SVM 11.6%
 - ロジスティック回帰 22.5%
 - K近傍法 18.4%

生成されたサンプル

Training Samples



Generated Samples



ギブスサンプリングを10,000回実行

まとめ

- BMの訓練アルゴリズムについて簡単に解説
- BMの種類を紹介
- シミュレーション結果はMNISTとNORBの認識タスクで良い結果を示す

参考文献

- 熊沢逸男「学習とニューラルネットワーク」森北出版株式会社 2007
- Ruslan Salakhutdinov, Geoffrey Hinton
「Deep Boltzmann Machine」
- Miguel A. Carreira-Perinan, Geoffrey E. Hinton
On Contrastive Divergence Learning