

課題 2: ボルツマン機械

提出締切 5月14日 (火)

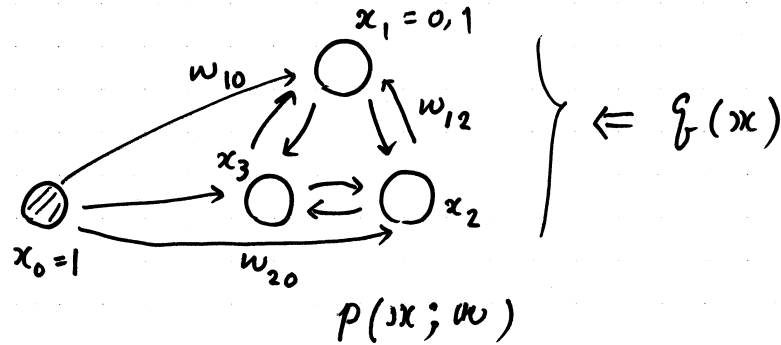


図 1: ボルツマン機械

素子数 $n = 3$ のボルツマン機械を考えよう. この機械の j 番目の素子から i 番目の素子への結合を w_{ij} . 状態を $\mathbf{x} = (x_1, x_2, x_3)$, $x_i \in \{0, 1\}$ と書く. ここで, 結合は対称であり ($w_{ij} = w_{ji}$), 自己結合はないとする ($w_{ii} = 0$). 各素子は非同期的に以下の確率にしたがい自分の状態を変えていく.

$$\Pr\{x_i = 1\} = \frac{1}{1 + \exp\{-u_i/T\}}, \quad \Pr\{x_i = 0\} = 1 - \Pr\{x_i = 1\} \quad (1)$$

$$u_i = \sum_{j=1}^n w_{ij}x_j - h_i = \sum_{j=0}^n w_{ij}x_j \quad (2)$$

ここで, 常に興奮している素子 $x_0 = 1$ を仮想的に考え, しきい値を $h_i = -w_{i0}$ と表現し記述を簡潔にした. パラメータ T の値は, 以下では特別な指示がなければ $T = 1$ とする. 時間 t における回路の状態を $\mathbf{x}(t)$ と書こう. 回路の状態は, $\mathbf{x}(0) \rightarrow \mathbf{x}(1) \rightarrow \mathbf{x}(2) \rightarrow \mathbf{x}(3) \rightarrow \dots$ と, マルコフ連鎖として状態遷移を続けていく. これは次の定常分布

$$p(\mathbf{x}; \mathbf{w}) = c \exp\{-E(\mathbf{x}; \mathbf{w})/T\} \quad (3)$$

をもつことが知られている. ここで

$$E(\mathbf{x}; \mathbf{w}) = - \sum_{i < j} w_{ij}x_i x_j \quad (4)$$

$$= -(w_{01}x_1 + w_{02}x_2 + w_{03}x_3 + w_{12}x_1x_2 + w_{13}x_1x_3 + w_{23}x_2x_3) \quad (5)$$

であり, c は $\sum_{\mathbf{x}} p(\mathbf{x}; \mathbf{w}) = 1$ とするための正規化定数

$$c = c(\mathbf{w}) = \frac{1}{\sum_{\tilde{\mathbf{x}}} \exp\{-E(\tilde{\mathbf{x}}; \mathbf{w})/T\}}, \quad (6)$$

\mathbf{w} はパラメータ $\mathbf{w} = (w_{01}, \dots, w_{23})$ である。回路の性質は \mathbf{w} で決まる。以下、素子数 $n = 3$ の場合を考える。コンピュータプログラムは、素子数が $n = 4, 5, \dots$ となる場合にも簡単に対応できるように書ければよいが、無理はしなくてよい ($n = 16$ とか $n = 32$ の場合のシミュレーションができればよいが、2進数と10進数を変換する関数を書く必要がある)。

1. 初期状態として回路に適当な興奮状態を設定した場合、式 (1) にしたがって興奮状態が変化していく。素子数 $n = 3$ の場合、 $2^3 = 8$ 個の状態をとりうる。それぞれの状態が実際に式 (3) の確率で出現するか、以下の3通りについて確認せよ。

(a) どの素子も互いに結合していないとする ($w_{ij} = 0, i, j = 0, \dots, 3$)。この場合、式 (3) の定常分布を計算すると、どの状態も確率 $\frac{1}{8}$ で出現することがわかる。これを計算機実験で確かめよ。具体的には、

- i. 回路の初期状態を $\mathbf{x}(0) = (0, 0, 0)$ とする。 $t = 0$
- ii. どの素子を更新するか、ランダムに一つの素子を選ぶ。
- iii. 選ばれた素子の状態を更新する。
- iv. $t := t + 1$. (ii) にもどり、100万回繰り返す。
- v. $\mathbf{x}^0 \sim \mathbf{x}^7$ の各状態が出現した回数を調べよ (表1参照)。

表 1: 機械の定常状態 (結合なし)

	状態 \mathbf{x}	出現回数	出現頻度
\mathbf{x}^0	(0, 0, 0)		
\mathbf{x}^1	(0, 0, 1)		
\mathbf{x}^2	(0, 1, 0)		
\mathbf{x}^3	(0, 1, 1)		
\mathbf{x}^4	(1, 0, 0)		
\mathbf{x}^5	(1, 0, 1)		
\mathbf{x}^6	(1, 1, 0)		
\mathbf{x}^7	(1, 1, 1)		

(b) 結合係数 $\mathbf{w} = \{w_{ij}\}$ の値をランダムな値に設定し (たとえば $[-5:5]$ の一様分布にしたがう乱数。ただし $w_{ii} = 0, w_{ij} = w_{ji}$ とする), 式 (3) の定常分布を計算し、各状態の出現確率を求めよ。また、実際に計算機シミュレーションで出現頻度を確かめよ。

(c) 状態 \mathbf{x}^3 と状態 \mathbf{x}^6 だけが出現するような回路を作ることはできるだろうか。相関型連想記憶で記憶パターンを埋め込んだように $w_{ij} = \sum_{\mathbf{x}^3, \mathbf{x}^6} (2x_i - 1)(2x_j - 1)$

表 2: 機械の定常状態 (ランダム結合の場合)

	状態 \boldsymbol{x}	理論値 : $p(\boldsymbol{x}; \boldsymbol{w})$	実験値 : $p(\boldsymbol{x}; \boldsymbol{w})$ $l = 1,000$	実験値 : $p(\boldsymbol{x}; \boldsymbol{w})$ $l = 1,000,000$
\boldsymbol{x}^0	(0, 0, 0)			
\boldsymbol{x}^1	(0, 0, 1)			
\boldsymbol{x}^2	(0, 1, 0)			
\boldsymbol{x}^3	(0, 1, 1)			
\boldsymbol{x}^4	(1, 0, 0)			
\boldsymbol{x}^5	(1, 0, 1)			
\boldsymbol{x}^6	(1, 1, 0)			
\boldsymbol{x}^7	(1, 1, 1)			

としてみよう. 学習前の結合係数が $w_{ij} = 0, i, j = 0, \dots, n$ である場合, 学習後の各係数の具体的な値を求めよ ($w_{ij} = w_{ji}$ であるので, 以下では $i < j$ のみ記述を求めている).

$w_{01} =$, $w_{02} =$, $w_{03} =$, $w_{12} =$, $w_{13} =$, $w_{23} =$

表 3: 機械の定常状態 (連想記憶もどき)

	状態 \boldsymbol{x}	理論値 : $p(\boldsymbol{x})$	実験値 : $p(\boldsymbol{x})$ $l = 1,000$	実験値 : $p(\boldsymbol{x})$ $l = 1,000,000$
\boldsymbol{x}^0	(0, 0, 0)			
\boldsymbol{x}^1	(0, 0, 1)			
\boldsymbol{x}^2	(0, 1, 0)			
$\rightarrow \boldsymbol{x}^3$	(0, 1, 1)			
\boldsymbol{x}^4	(1, 0, 0)			
\boldsymbol{x}^5	(1, 0, 1)			
$\rightarrow \boldsymbol{x}^6$	(1, 1, 0)			
\boldsymbol{x}^7	(1, 1, 1)			

この $\{w_{ij}\}$ と式 (3) を用い, 定常分布における各状態の $p(\boldsymbol{x})$ (理論値. ボルツマン分布) を計算せよ. また, コンピュータシミュレーションをおこない, 各状態が実際に出現した頻度を計算し, それを理論値と比較せよ (繰り返し回数 l を 1 万, 100 万などを変えて, 少なくとも 2 通りは試すこと).

2. 【ボルツマン機械の学習】各状態 \boldsymbol{x} が次の表に示す頻度で出現するボルツマン機械を, 学習により実現しよう. 具体的には,

(a) すべての結合係数を $w_{ij} = 0$ とする (乱数でもよい).

表 4: ボルツマン機械の学習. 信号 \mathbf{x} の出現頻度.

	状態 \mathbf{x}	外界: $q(\mathbf{x})$	確率 $p(\mathbf{x})$
\mathbf{x}^0	(0, 0, 0)	0.1	
\mathbf{x}^1	(0, 0, 1)	0.1	
\mathbf{x}^2	(0, 1, 0)	0.05	
\mathbf{x}^3	(0, 1, 1)	0.05	
\mathbf{x}^4	(1, 0, 0)	0.1	
\mathbf{x}^5	(1, 0, 1)	0.1	
\mathbf{x}^6	(1, 1, 0)	0.4	
\mathbf{x}^7	(1, 1, 1)	0.1	

- (b) 表 4 に示す確率分布 $q(\mathbf{x})$ にしたがって \mathbf{x} を 100 個生成する. x_i と x_j が同時に 1 になった頻度を f_{ij} とせよ. (こんな実験をするのは時間の無駄だと思えば, 期待値を計算すればよい. $f_{01} = 0.7, f_{02} = 0.6, f_{03} = 0.35, f_{12} = 0.5, f_{13} = 0.2, f_{23} = 0.15$. ただし信号の次元が大きくなると, こういう計算はできないので, 実際に信号を生成して統計量を計算する必要がある).
- (c) 回路の初期状態を $\mathbf{x}(0) = (0, 0, 0)$ に設定する.
- (d) ランダムに一つの素子を選び, その素子の状態を更新する. これを 100 回繰り返し, 素子 x_i と x_j が同時に 1 になった頻度を g_{ij} とする (頻度 g_{ij} は同時に 1 になった回数を 100 で割ったもの. $0 \leq g_{ij} \leq 1$).
- (e) 結合係数の値を更新する (学習):

$$w_{ij} := w_{ij} + 0.01(f_{ij} - g_{ij}) \quad (7)$$

- (f) Kullback-Leibler のダイバージェンス

$$D = \sum_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \quad (8)$$

を計算する. ここで $p(\mathbf{x})$ は, 実際に回路を自由に走らせて計算してもよいが, 素子数が少ない場合は, 回路のパラメータ $\{w_{ij}\}$ から計算できる理論値 (式 (3)) を用いればよい.

- (g) (d) にもどり繰り返す (繰り返し回数は任意. 例えば 10000 回).
- (h) 横軸に学習回数 t , 縦軸に Kullback-Leibler のダイバージェンスをプロットした図を描き (曲線が描ける), 学習が進むにしたがって $p(\mathbf{x})$ が $q(\mathbf{x})$ に近づいていく ($D = 0$ に近づく) 様子を確認せよ

(i) 獲得した w_{ij} の値を列挙せよ.

$$w_{01} = \quad , w_{02} = \quad , w_{03} = \quad , w_{12} = \quad , w_{13} = \quad , w_{23} =$$

(j) しきい値がない $n = 3$ の回路を考えよう ($w_{01} = w_{02} = w_{03} = 0$). このとき同様な実験をおこない (もう一度 (c) から実験をする), 結果を比較し考察せよ (2本の曲線の位置関係を比較せよ).

3. しきい値ありの学習済みの回路において1番目の素子の値を $x_1 = 1$ と固定し, x_2, x_3 だけ値の更新を許すとしよう. このとき, 条件付き確率を直接学習したわけではないが, 条件付き確率が学習されていることを確かめよ.

表 5: 問 3. 条件付き確率の獲得. 学習後, $x_1 = 1$ と固定し x_2, x_3 だけ値を更新し (更新回数は 1 万回), 回路の状態を自由に動かし, 各状態が出現する頻度 $p(\mathbf{x}|x_1 = 1)$ を調べよ.

	状態 \mathbf{x}	外界: 同時分布 $q(\mathbf{x})$	外界: 周辺分布 $q(\mathbf{x} x_1 = 1)$	$p(\mathbf{x} x_1 = 1)$
\mathbf{x}^0	(0, 0, 0)	0.1	0	0.0
\mathbf{x}^1	(0, 0, 1)	0.1	0	0.0
\mathbf{x}^2	(0, 1, 0)	0.05	0	0.0
\mathbf{x}^3	(0, 1, 1)	0.05	0	0.0
\mathbf{x}^4	(1, 0, 0)	0.1	0.142	
\mathbf{x}^5	(1, 0, 1)	0.1	0.142	
\mathbf{x}^6	(1, 1, 0)	0.4	0.571	
\mathbf{x}^7	(1, 1, 1)	0.1	0.142	

4. 問 1(c) では, 状態 \mathbf{x}^3 と状態 \mathbf{x}^6 だけがそれぞれ確率 0.5 で出現するような機械を設計することに失敗したはずである. もう一度, 前問の学習則を使い, 機械を学習させてみよう. 獲得した w_{ij} の値を列挙せよ.

$$w_{01} = \quad , w_{02} = \quad , w_{03} = \quad , w_{12} = \quad , w_{13} = \quad , w_{23} =$$

5. (オプション課題) $n = 8, 16, 32, 64$ で, 同様な実験をしてみよ. $n = 8$ の場合, \mathbf{x} の総数は $2^8 = 256$ とおりある. このひとつひとつについて出現確率が計算できる. 一方, 学習させたい確率分布を指定するには, $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$ より, 表 4 のようにして, 255 個の数字を与える必要がある. これは大変であるので, たとえば次のような, マルコフ情報源を考えるとよい. $\mathbf{x} = (x_1, x_2, \dots, x_8), x_i \in \{0, 1\}, i = 1, \dots, 8$ として, 確率分布 $p(\mathbf{x}) = p(x_0)p(x_1|x_0)p(x_2|x_1) \cdots p(x_8|x_7)$ が以下のように与えられていると

表 6: 問 4. 機械の定常状態. 学習後.

	状態 \boldsymbol{x}	外界の信号 : 確率 : $p(\boldsymbol{x})$	実験値 : $p(\boldsymbol{x})$ $l = 1,000$	実験値 : $p(\boldsymbol{x})$ $l = 1,000,000$
	\boldsymbol{x}^0 (0, 0, 0)	0.0		
	\boldsymbol{x}^1 (0, 0, 1)	0.0		
	\boldsymbol{x}^2 (0, 1, 0)	0.0		
	$\rightarrow \boldsymbol{x}^3$ (0, 1, 1)	0.5		
	\boldsymbol{x}^4 (1, 0, 0)	0.0		
	\boldsymbol{x}^5 (1, 0, 1)	0.0		
	$\rightarrow \boldsymbol{x}^6$ (1, 1, 0)	0.5		
	\boldsymbol{x}^7 (1, 1, 1)	0.0		

する.

$$p(x_1) = \text{Prob}(X_1 = x_1) = \begin{cases} 0.5 & \text{if } x_1 = 0 \\ 0.5 & \text{if } x_1 = 1 \end{cases} \quad (9)$$

$$p(x_{i+1}|x_i) = \begin{cases} \theta & \text{if } x_i = 0, x_{i+1} = 0 \\ 1 - \theta & \text{if } x_i = 0, x_{i+1} = 1 \\ \theta & \text{if } x_i = 1, x_{i+1} = 1 \\ 1 - \theta & \text{if } x_i = 1, x_{i+1} = 0 \end{cases} \quad (10)$$

6. (オプション課題) $n = 100$ で実験する場合, とりうる可能性のある状態 \boldsymbol{x} は, $2^{100} = (2^{10})^{10} \approx (10^3)^{10} = 10^{30}$ 通りあり, それぞれの状態の出現頻度を求めることは, もはや現実的には不可能である. 各素子が $x_i \in \{0, 1\}$ の 2 値をとる場合でさえも, 次元数が 100 になると, 各状態 \boldsymbol{x} について個別に確率 $p(\boldsymbol{x})$ を設計していく方法は, 場合の数が多すぎて採用できない. しかし, 音声でも画像でも, 256 次元ベクトル程度の信号をモデル化することはよくおこなわれている. このような高次元データの確率分布を設計する方法を考えてみよう.

(コメント) 一般に, 素子数を n とすると, 回路のパラメータ数は $\frac{n(n+1)}{2}$ であり, とりうる状態の総数は 2^n である. $n = 3$ の場合では, 任意の確率分布をよく近似できる. これはパラメータ数が $\frac{n(n+1)}{2} = 6$ の回路で, 状態数 $2^n = 8$ (確率は足して 1 になるので, 自由度は 7) であったからである. $n = 10$ の場合, パラメータの総数は $\frac{n(n+1)}{2} = 55$ と増えるが, 状態数は素子数の指数のオーダーで $2^n = 1024$ に増えてしまう.

レポートの最後には, 感想, 質問などを記述してほしい. 理解しにくい点があった場合は, このプリント中の, どこの部分が分かりにくかったか具体的に, 指摘してもらえれば助かります (来年度向けに改善するため).