数理脳科学
2016 年 5 月 13 日

課題 2: 自己組織化モデル

提出締切 5 月 27 日（金）

1. 信号 $\boldsymbol{x} = (x_1, x_2)$ の各要素 x_1, x_2 は区間 [0,1) の一様分布にしたがうとする．このような 2 次元の信号 $\boldsymbol{x}$ を 100 個生成し，平面（横軸 x_1，縦軸 x_2）上にプロットせよ．

2. 信号 $\boldsymbol{x} = (x_1, x_2)$ の 各要素 x_1, x_2 がそれぞれ独立に平均 $\mu = 1$, 分散 $\sigma^2 = 1$ の正規分布にしたがうとする．このような信号 $\boldsymbol{x}$ を 100 個生成し，平面上にプロットせよ．

3. **入力信号空間，モデル神経場ともに 1 次元の場合**

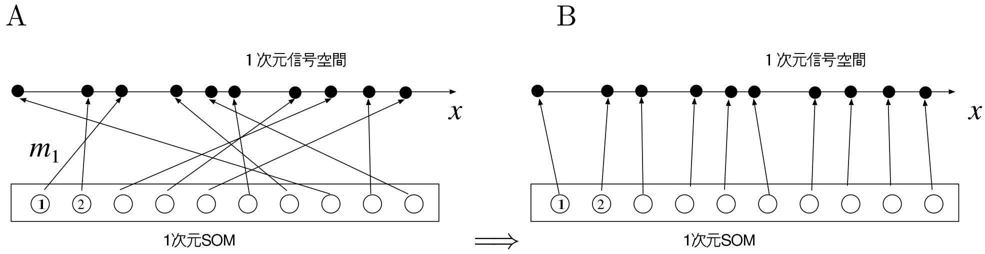


図 1: 1 次元の信号空間に 1 次元 SOM を適用．左：学習前．右：学習後

素子数 $n = 10$ の 1 次元神経場モデル（自己組織化マップ，SOM）を考えよう．各素子は 1 次元の信号 x を受け取り，i 番目の素子は入力空間に対して m_i の重みをもつ（m_i は，一般的には，参照ベクトルとよばれている）．ここで $m_i, i = 1, \cdots, n$ の値は，区間 [0,1) の一様分布にしたがう乱数としておく（図 A）．SOM は，入力 x に対し，以下のルールにしたがい，一つの素子 c だけが興奮する（これを勝者とよぶ）．

$$c = \underset{i}{\operatorname{argmin}} |m_i - x| \tag{1}$$

11 通りの入力 $x = 0.0, 0.10, 0.20, \cdots, 1.0$ に対する勝者（winner, 素子の番号 c）を求めよ（表 1）．

表 1: 入力 x に対する勝者素子の番号

x	0.0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
c											

4. 区間 [0,1) の一様分布にしたがうランダムな入力信号 x を生成し，これに対する勝者を計算せよ．これを 1000 回くりかえし，各素子が勝者になった回数を求めよ．前問 3 で作った回路（前問と同じ $m_i, i = 1, \cdots, n$）を用いて実験をおこなうこと（表 2）．

5. 入力信号 x の構造に応じて学習が進む．その様子をみてみよう．

   (a) $m_i, i = 1, 2, \cdots$ を区間 [0,1) の一様分布にしたがう乱数に設定する（前問で使った値と同じ）．

表 2: 各素子が勝者となった頻度

素子の番号	1	2	3	4	5	6	7	8	9	10
勝者になった回数										

(b) 区間 [0,1) の一様分布にしたがう信号 x を一つ生成し，モデルに入力信号 x を与える．

(c) 以下の式にしたがい係数 $m_i, i = 1, 2, \cdots$ の値を更新する（これを学習という）．

$$m_i \quad := \quad m_i + \alpha \exp\left\{-\frac{|c-i|^2}{2\sigma^2}\right\}(x - m_i) \tag{2}$$

ここで $\alpha > 0$ は学習の強さを表す係数，$\sigma > 0$ は，勝者素子が，どのくらい遠く離れた素子まで影響を与えるかを定める定数である（exp{} の部分は近傍関数とよばれている）．パラメータの値は，たとえば $\underline{\alpha = 0.2, \sigma = 0.8}$ にしておく．

(d) (b) に戻り，2000 回，繰り返す．ここで 100 回ごとに，前問 3,4 で使った統計量を計算せよ（統計量を求めるときは学習しない）．表 3 では，信号 x を 0.1 間隔で入力し，11 通りの入力に対する勝者素子を調べているが，0.05 間隔，21 通りに対する勝者素子を調べる方が分かりやすい（はず）．

表 3: 入力 x に対する勝者素子の番号

t \ 入力 x	0.0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0											
100											
⋮											
2000											

表 4: 各学習段階において，1000 個の入力に対し，各素子が勝者となった回数

t \ 素子の番号	1	2	3	4	5	6	7	8	9	10
0										
100										
⋮										
2000										

6. 図 B は学習後の様子を描いた例である．初期状態から，学習が進む様子を，この図 A,B を参考に作成せよ（本質的な部分が表現できればよい）．

7. 入力信号 x が，平均 $\mu = 0.5$，分散 $\sigma^2 = 1$ の正規分布にしたがうとする．5.(b) だけを変えて実験し，その結果を x が一様分布に従う場合と比較し，考察せよ．

8. 素子数 n, 係数 α, σ の値を適当な値に変えて実験してみよ．

   （コメント） 係数 α，σ は，ともに，$\alpha = 0.3(1.0 - \frac{t}{T}) + 0.01$，$\sigma = 2.0(1.0 - \frac{t}{T}) + 0.1$，のように，時間と共に小さくすれば，きれいな絵が描けるかもしれない．T は $T = 2000$ など，あらかじめ設定しておく最終的な学習回数．T は，もちろん，大きい方が学習が進む．

9. 入力信号空間が 2 次元，モデル神経場が 1 次元の場合：2 次元から 1 次元への写像

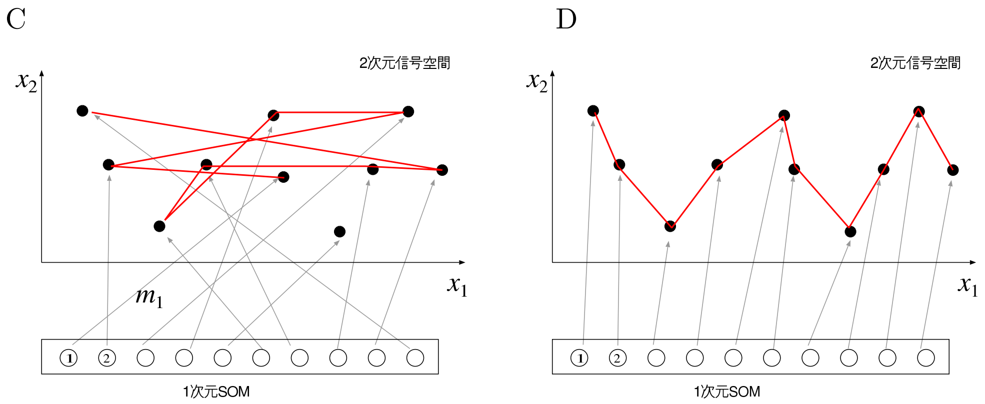


図 2: 2 次元の信号空間に 1 次元 SOM を適用．左：学習前．右：学習後

素子数 $n = 10$ の 1 次元 SOM の各素子が 2 次元の信号 $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$ を受け取る場合を考えよう．i 番目の素子は入力空間に対して $\boldsymbol{m}_i = (m_{i1}, m_{i2})^{\mathrm{T}}$ の重みをもつとする．ここで $m_{i1}, m_{i2}, i = 1, \cdots, n$ の初期値は，区間 [0,1) の一様分布にしたがう乱数とする（図 C）．入力信号 $\boldsymbol{x}$ が $[0,1) \times [0,1)$ 平面上に一様に分布する場合を考えよう．前問の場合と同様に，SOM は，入力 $\boldsymbol{x}$ に対し，一つの素子 c だけが興奮する．

$$c \;=\; \underset{i}{\operatorname{argmin}} ||\boldsymbol{m}_i - \boldsymbol{x}|| \tag{3}$$

入力信号 $\boldsymbol{x}$ を 1 回与えるごとに，以下の式にしたがい係数 $m_i, i = 1, \cdots, n$ の値を更新せよ．

$$m_i \;:=\; m_i + \alpha \exp\left\{-\frac{|c-i|^2}{2\sigma^2}\right\}(\boldsymbol{x} - \boldsymbol{m}_i) \tag{4}$$

図 D は学習後の様子を描いた例である．初期状態から学習が進む様子を，この図 C,D を参考に作成せよ（入力信号 $\boldsymbol{x}$ の平面に，各素子の $\boldsymbol{m}_i$ をプロットし，素子配列上で近傍関係にある素子どうしの $\boldsymbol{m}_i$ を線で結べばよい）．パラメータの値は，たとえば $\alpha = 0.2, \sigma = 0.8$ にしておく．1000 回の学習をおこなえ．素子数を $n = 20, 50$ など何通りかの値に変え，適宜，各パラメータの値，学習回数を調整して計算機実験をおこない，その結果を考察せよ．

10. 入力信号空間，モデル神経場がともに 2 次元の場合

素子数 $n = 10 \times 10$ の 2 次元 SOM を考えよう．各素子は 2 次元の信号 $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$ を受け取り，i 番目の素子は入力空間に対して $\boldsymbol{m}_i = (m_{i1}, m_{i2})^{\mathrm{T}}$ の重みをもつとする．ここで $m_{i1}, m_{i2}, i = 1, \cdots, n$ の初期値は，区間 [0,1) の一様分布にしたがう乱数とする．入力信号 $\boldsymbol{x}$ が $[0,1) \times [0,1)$ 平面上に一様に分布する場合を考えよう．SOM は，入力 $\boldsymbol{x}$ に対し，式 (3) にしたがい，一つの素子 c だけが興奮する．モデルに入力信号 $\boldsymbol{x}$ を 1 回与えるごとに，以下の式にしたがい係数 $\boldsymbol{m}_i, i = 1, \cdots, n$ の値を更新せよ．

$$\boldsymbol{m}_i \quad := \quad \boldsymbol{m}_i + \alpha \exp\left\{ -\frac{||\boldsymbol{r}_c - \boldsymbol{r}_i||^2}{2\sigma^2} \right\} (\boldsymbol{x} - \boldsymbol{m}_i) \tag{5}$$

ここで $\boldsymbol{r}_i$ は，i 番目の素子の配列上（神経場）での位置である．配列上の位置とは，$n = 10 \times 10$ の場合，$i = 1, \cdots, 100$ であり，$\boldsymbol{r}_1 = (0,0)$，$\boldsymbol{r}_2 = (1,0), \boldsymbol{r}_3 = (2,0), \cdots, \boldsymbol{r}_{11} = (0,1), \boldsymbol{r}_{12} = (1,1), \cdots, \cdots, \boldsymbol{r}_{100} = (9,9)$ などと考えればよい．$\boldsymbol{r}_i = (x, y)$ と書くと，$x, y = 0, \cdots, 9$　として $i = x + 1 + 10y$ と表現できる．パラメータの値は，たとえば $\alpha = 0.2, \sigma = 0.8$ にしておく．20,000 回の学習をおこなえ．各 $\boldsymbol{m}_i$ を平面（入力信号空間）に点でプロットし，2 次元神経場上でとなりあう素子の $\boldsymbol{m}_i$ を線で結べば，図 3 のような図が描け，学習が進む様子がよくわかる．

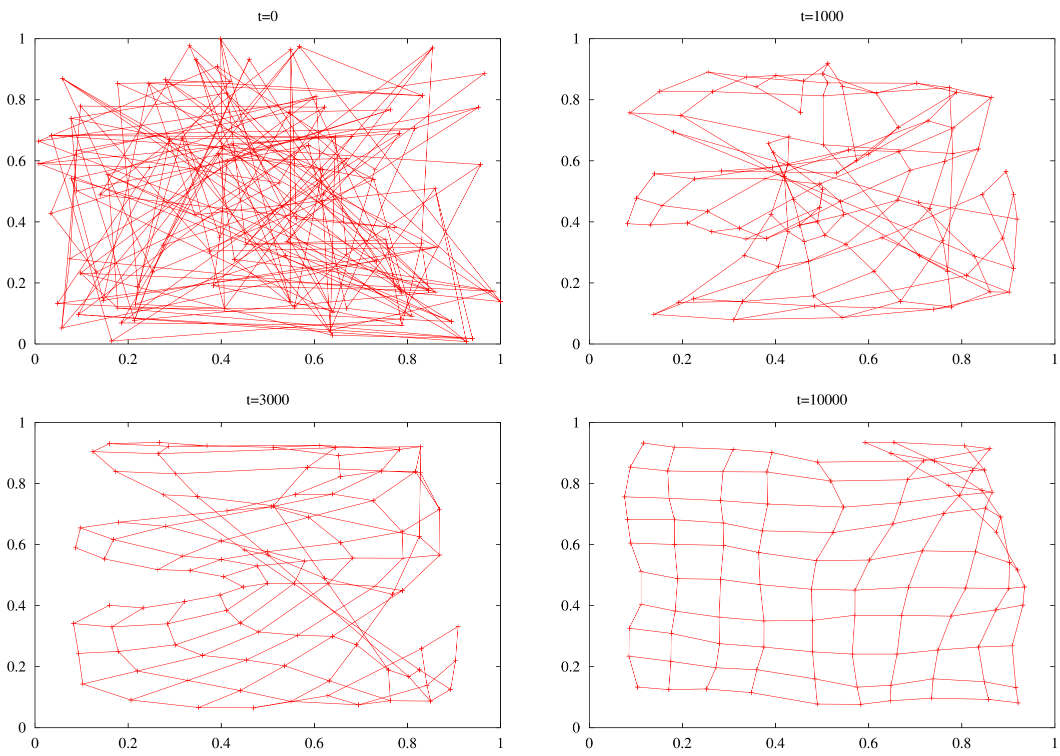


図 3: 自己組織化の過程

11. 前問において，入力信号 $\boldsymbol{x}$ の確率分布を適当に設計し，生成したデータをモデルに与え，どう学習が進むか，いろいろ試して考察せよ．具体的な分布としては，2 次元正規分布，円周の周辺だけにデータが存在する分布，三角形の領域にだけデータが存在

する分布など，各自自由に考えればよい．素子数を増やせば結果は変わる．講義中に例は示すが，面白い振る舞いをする例を独自に探せ．入力信号の次元は 3 次元以上にしてもよい．

注意： 課題が手に負えない場合，はやめに A-334 か A-423 まで，相談にくること．

補足： 「脳はどういう風に動いているのか」，このモデルを通して感じ取れる可能性のある性質をいくつか挙げておく．

(a) トポグラフィックマップの形成：でたらめな結合が，整然とした結合に変化していく（となりの細胞の行き先は，自分の行き先のとなりになること）

(b) 入力空間において，入力頻度の高い領域ほど多くの細胞が割り当てられる．→ 入力信号 $\boldsymbol{x}$ の確率分布を一様分布から偏ったものにしてみる．

(c) 一部の細胞（素子，ニューロン）が壊れても，他の細胞が代わりに働くようになる．→ 学習の途中で一部の細胞を取り除いてみる．

(d) 高次元の情報を低次元でうまく表現できる．
→ 問題 9. がこれに対応している．

課題に曖昧な点があった場合は適当に解釈してよい．その場合は必ず，具体的にどの部分をどう解釈したかをレポートに記述すること．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大変助かる（来年度向けに改善するため）．

本講義においては，単位の認定にあたって，問題設定，計算機実験の手順，計算機実験の結果を，自分の書いたレポートを片手に，ほぼ何もみないで説明できるレベルに到達できていることを要求します．

2011 年 5 月 20 日
A.D.

ノート： gnuplot の使い方

<u>複数の折れ線の描き方</u>

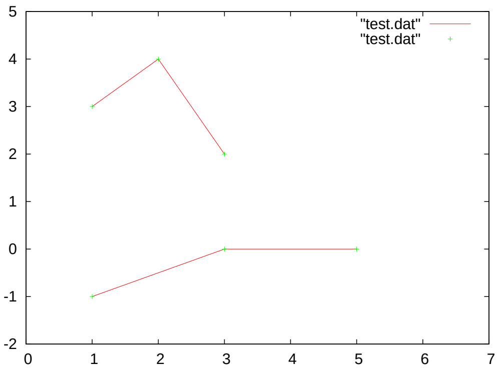
例えばデータファイル test.dat を次のような内容で用意しておく．

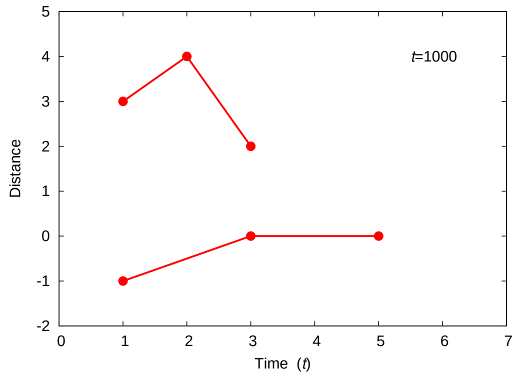
```
1.0 3.0
2.0 4.0
3.0 2.0

1.0 -1.0
3.0 0.0
5.0 0.0
```

（2 次元 (x, y) 座標データ，合計 6 点が記述されている． 間に空行があるのがポイント）

```
gnuplot> plot [0:7][-2:5]"test.dat" with lines, "test.dat" with points
```

とすると





上図（左）が得られる．レポート等には，上図（右）のような美しい図を貼り付けたい． このときは，以下のようなファイルを作り（gp-script としておく），

```
set terminal postscript "Helvetica" 20 color eps enhanced
set xlabel "Time  ({/Italic t})"
set ylabel "Distance"
set nokey  # これを含めると，凡例が表示されない．
set style line 1 lt 1 lw 5 pt 3 ps 1.0    # 線や点の太さの定義
set style line 2 lt 1 lw 3 pt 7 ps 2.0    # 自由に変えて試してみる
set label "{/Italic t}=1000"  at first 5.5, first  4.0
plot [0:7][-2:5]"test.dat" with lines linestyle 1,\
"test.dat" with points linestyle 2
#
# 意味： lt line_type,  lc linecolor,  lw linewidth
#        pt pointtype,  ps pointsize
```

これを

```
$ gnuplot gp-script > test.eps
```

とすれば，LaTeX で利用しやすい eps ファイルが作成できる．