

競合的な隠れユニットを持つ

神経回路網の学習について

東京大学 工学部 倉田 耕治

The Univ. of Tokyo Koji KURATA

はじめに

ここで論ずるのは，外から与えられた入出力関係を学習することのできる，ある種の神経回路網である．回路は，入力層，中間層（隠れユニット群），出力層の三層からなり，情報は入力層から中間層を経て出力層に伝えられる．中間層のユニットは，たがいに競合するように結合され，全体として最大値検出回路として機能するようになっている．学習は，入力層と中間層の間の結合，および中間層と入力層の間の結合を変化させることによっておこなわれる．このモデル自体は新しいものではないが，理論的な研究は少ない．

この回路のパフォーマンスは，どのような隠れユニットが形成されるかにかかっている．本稿では，与えられた入力分布に対する，隠れユニットの形成に関する理論的な考察を行い，隠れユニットの数が非常に多い場合の，その分布を求める．これにより，出力の平均自乗誤差を最小にするためには，

学習時の入力分布を，実用時の入力分布とは少し異なったものにしたほうがよいことが導かれる．この結果を一言で述べれば，”むずかしいことは何度も練習すべし”となる．

1. 基本競合系

本モデルの中間層は， M 個のユニットからなる基本競合系[1]をなすような内部結合をもっている．まず，これについて述べよう．各ユニットは，ひとつの実数で表される内部状態をもっている． i 番目のユニットの内部状態を u_i と書くことにする．ユニットの出力は， $1[u_i]$ で与えられる．ここで， $1[u]$ は，

$$1[u] = \begin{cases} 1 & (u > 0) \\ 0 & (u \leq 0) \end{cases} \quad (1)$$

なる関数である．各ユニットは，自分を含む中間層内のすべてのユニットに抑制性の信号を送っている．また，自分自身にはこれに加えて興奮性の信号を送っている． u_i のダイナミクスは，つぎの式で表されるものとする．

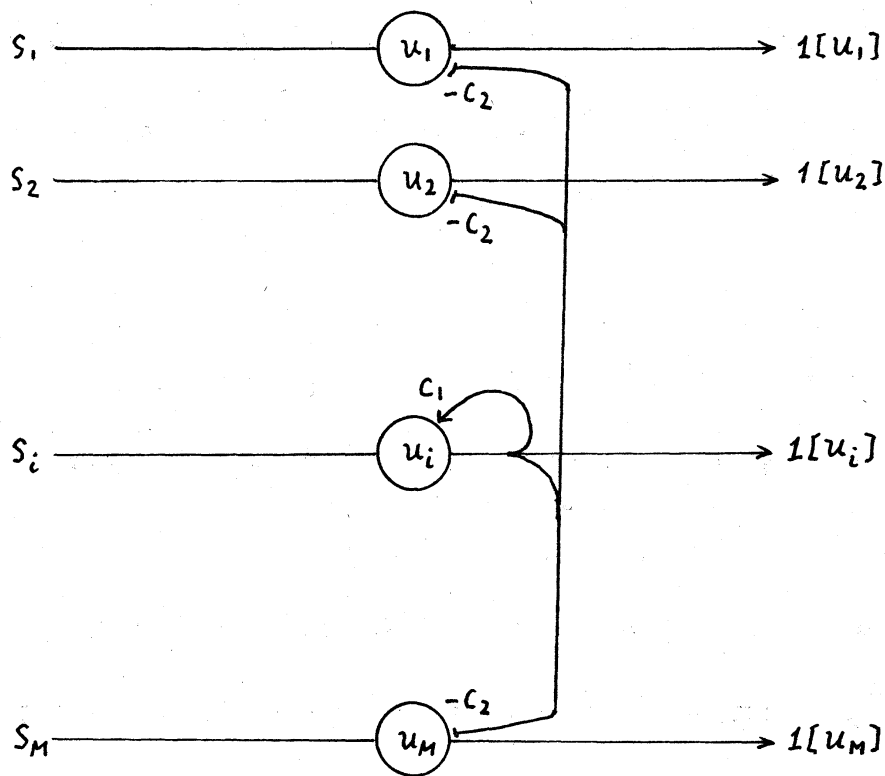


図1 i 番目のユニットから，他のユニットへの結合だけを描いた．

$$\begin{aligned} & du_i(t)/dt \\ &= -u_i(t) + c_1 1[u_i(t)] - c_2 \sum_j 1[u_j(t)] + s_i, \quad (2) \\ & i = 1, 2, \dots, M \end{aligned}$$

ここに c_1 と c_2 はそれぞれ，興奮性と抑制性のシナプス強度を表す正数， s_i は i 番目のユニットが入力層から受け取る信号で $-1 \leq s_i \leq 1$ を満たすものとする．

シナプス強度 c_1, c_2 が， $c_2 + 1 < c_1 < 2c_2 - 1$ を満たしてい

るとき，この回路は，最大値検出器としてはたらく．すなわち，すべての u_i に等しい正の初期値を与えると，十分な時間の後，最大の s_i を受け取るユニットだけが，出力 1 を出し，他の出力は 0 になる．

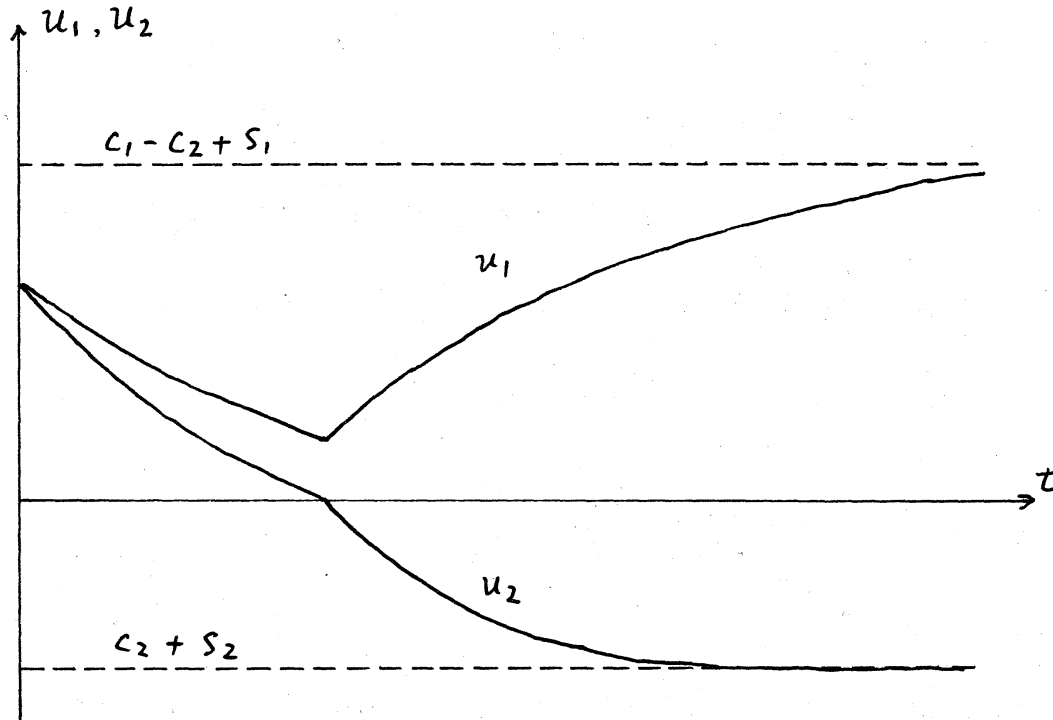


図 2 ユニット 2 個， $s_1 > s_2$ の場合の式 (2) の解

2. 三層回路網と学習則

本稿で扱う三層回路の入力層には， $L+1$ 個のユニットがあり，ここに外部からの入力 $x = (x_1, x_2, \dots, x_L, x_{L+1})'$ が与えられる．ここに ' は転置を表す．入力ベクトルのユークリッドノルムは，常に 1 であると仮定する．つまり， $x \in S^L$ である．入力層の j 番目のユニットから中間層の i 番目のユニ

ットへのシナプスの強度を, w^{1ij} とし, これを j についてまとめて $w^{1i} = (w^{1i1}, w^{1i2}, \dots, w^{1iL}, w^{1iL+1})'$ と表す. すなわち中間層のユニットの入力は,

$$s_i = x' w^{1i}, \quad i = 1, 2, \dots, M \quad (3)$$

で与えられる. 回路への入力同様, シナプスベクトル w^{1i} も S^L 上に制限されると仮定する.

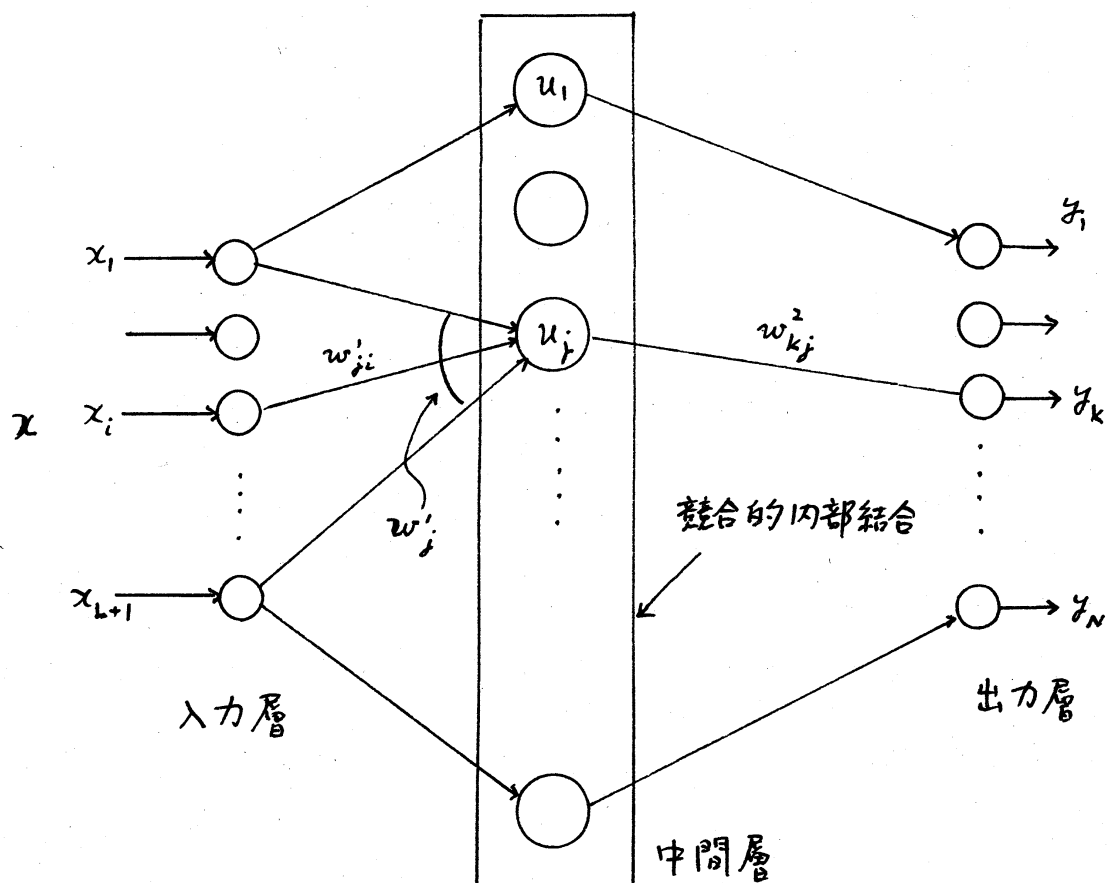


図3 競合的な隠れユニットを持つ三層回路網

出力層は， N 個の線形な特性を持ったユニットから成っており，その出力つまり回路の出力は，つぎのように与えられる．

$$y_i(t) = \sum_j w_{ij}^2 \phi[u_j(t)], \quad i = 1, 2, \dots, N \quad (4)$$

以上，式(1)～(4)で，この回路網の出入力関係が完全に記述された．入力層をある $x \in S^L$ に固定してしばらくすると，中間層のユニットのなかで最大の入力を受け取るものだけが，1を出力するようになる．そのユニットを $k(x)$ 番目と書くこ

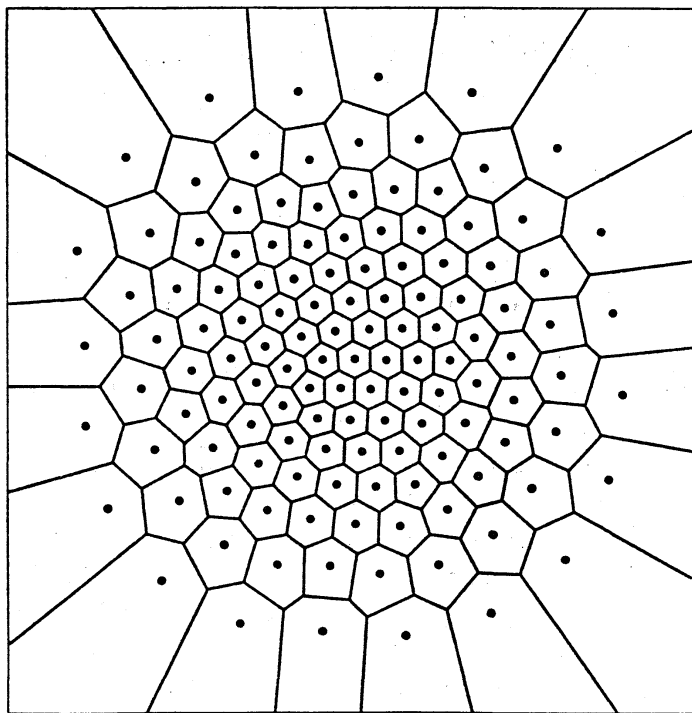


図 4 ボロノイ・ダイヤグラム ($L=2$ の場合)

(鈴木敦夫氏のご好意による)

とにすれば，式(4)より，その時点で回路の出力は $y_i(t) = w_{i, k(x)}$ ， $i = 1, 2, \dots, N$ となっている．これを入力 x に対するこの回路の出力 $y_i(x)$ であると考えことにする．入力の空間 S_L を， $k(x)$ によって，そのなかでは， $k(x)$ が一定であるような部分集合

$$\begin{aligned} V_i &= \{x \mid k(x) = i\} \\ &= \{x \mid x'w_j > w'x_j, \\ &\quad j = 1, 2, \dots, i-1, i+1, \dots, M\} \end{aligned} \quad (5)$$

に (V_i の境界は除いて) 分割すると，この回路の入出力関係は，

$$y_i = \begin{cases} w_{i,1}, & x \in V_1, \\ w_{i,2}, & x \in V_2, \\ w_{i,3}, & x \in V_3, \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ w_{i,n}, & x \in V_n, \end{cases} \quad (6)$$

$$i = 1, 2, \dots, N$$

という階段関数になる。この分割は、 $\{w^{11}, w^{12}, \dots, w^{1n}\}$ によって、 S^1 上に生成されるボロノイ・ダイアグラムである。

学習は、回路網に次々に入力を与え、そのたびにシナプス強度 w^{1ij}, w^{2ij} を少しずつ変化させることによっておこなわれる。中間層内部の結合は固定する。学習の目的は与えられた入出力関数 $f: S^1 \rightarrow R^N$, $f(x) = (f_1(x), f_2(x), \dots, f_N(x))$ をできるだけ正確に近似することである。中間層と出力層の間の結合は誤り訂正学習によって変化させる。

$$\begin{aligned} w^{2ij} &:= w^{2ij} + \epsilon_2 \{f_j(x) - y_i(x)\} 1[u_i] \\ &= w^{2ij} + \epsilon_2 \{f_j(x) - w_{ij}\} 1[u_i], \end{aligned} \quad (7)$$

$$0 < \epsilon_2 < 1,$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M$$

ここで $:=$ は更新を表す。

入力層と中間層の間の結合 w^{1i} の学習はヘブ学習によるが、 w^{1i} は、 S^1 上に制限されているので、学習を与える式は次のようになる。

$$w^{1i} := \{w_{ij} + \epsilon_1 1[u_i] x\} / \|w_{ij} + \epsilon_1 1[u_i] x\|,$$

$$0 < \varepsilon_i < 1, \quad i = 1, 2, \dots, M \quad (8)$$

3. 隠れユニットの分布

回路網への，入力の分布が， $p(x)$ ， $x \in S^L$ であるとき，学習の結果どのような結合が形成されるかを考えよう．バックプロパゲーションと違って，中間層と出力層の間の結合 w^2_{ji} は中間層のユニットの活動に影響を与えないので，先ず入力層と中間層の間の結合 w^1_{ji} の形成だけを考える．これが決まれば， w^2_{ji} の形成は，後で述べるように大変簡単に分析できる．

直観的に言えば， w^1_{ji} は互いに反発しながら $p(x)$ の大きい方へ動くので，もし S^L 上にいくつかの $p(x)$ のピークがあり，隠れユニットの数がそれとほぼ同じならば，椅子取りゲームの様なことがおきる．ここでは隠れユニットの数が非常に大きい場合を考え， w^1_{ji} の S^L 上での分布 $\rho(x)$ を考えることにする．

隠れユニットの数が非常に多く， S^L 上で混み合っているとき，個々のボロノイ領域 $V_i(t)$ は，非常に小さなものになる．このとき式(8)から導かれる平均学習方程式[1]はつぎのように近似できる．

$$dw^{1_i}(t)/dt \propto m_i(t) - w^{1_i}(t), \quad i = 1, 2, \dots, M \quad (9)$$

ここに $m_i(t)$ は, i 番目のユニットが発火しうる入力集合すなわち $V_i(t)$ の $p(x)$ の重み付重心である. 式(9)はつきのかたちのポテンシャルをもつことが分かっている[2].

$$\begin{aligned} E(w^{1_1}, w^{1_2}, \dots, w^{1_M}) \\ = \sum_i \int v_i \|x - w^{1_i}\|^2 p(x) dx \end{aligned} \quad (10)$$

それでは E を最小にする w^{1_i} の分布とはなにか. 隠れユニットの数が非常に多いとき, 局所的には $p(x)$ は一定で, w^{1_i} は, 局所的には一様な最密構造を実現しているものと考えられる. 点 x 付近のボロノイ領域の大きさを $a(x)$ とすると, $a(x) \propto \rho(x)^{-1/L}$. 一つの領域内での $\|x - w^{1_i}\|^2$ の積分 $e(a(x))$ は,

$$\begin{aligned} e(a(x)) &= \int v_i \|x - w^{1_i}\|^2 p(x) dx \\ &\propto p(x) \int_{v_i/a(x)} \|y\|^2 a(x)^L dy \\ &= p(x) a(x)^{L+2} e(1) \\ &\propto p(x) \rho(x)^{-(L+2)/L}. \end{aligned} \quad (11)$$

単位体積あたりのポロノイ領域の数が ρ だから,

$$E \propto \frac{\int_{S_L} p(x) \rho(x) e(a(x)) dx}{\int_{S_L} p(x) \rho(x)^{-2/L} dx} \quad (12)$$

となる. これを最小にする $\rho(x)$ を, $\int_{S_L} \rho(x) dx = 1$ という条件下で変分をつかって求めると

$$\rho(x) \propto p(x)^{L/(L+2)} \quad (13)$$

を得る.

w^2_{ij} の学習の結果は, (7)式から得られる平均学習方程式

$$dw^2_{ij}/dt = \int_{V_j} \{f_i(x) - w^2_{ij}\} p(x) dx \quad (14)$$

の平衡状態として, つぎのように得られる.

$$\begin{aligned} w^2_{ij} &= \int_{V_j} f_i(x) p(x) dx / \int_{V_j} p(x) dx \\ &\propto f_i(m_j) \end{aligned} \quad (15)$$

ここに m_j は, V_j の重心である.

4. 平均自乗誤差を最小にする隠れユニットの分布

前節ではヘブ学習によって形成される．隠れユニットの分布を求めたが，これとは別に，この回路で実現される．入出力関数 $f'(x; w^1, w^2)$ と，与えられた関数 $f(x)$ の平均自乗誤差 (MSE) を最小にするためには隠れユニットは，どのように分布すべきかを考えよう．MSE はつぎのように書ける．

$$\begin{aligned} \text{MSE} &= \int_{S_L} \| f(x) - f'(x; w^1, w) \|^2 p(x) dx \\ &= \sum_{ij} \text{MSE}_{ij}, \end{aligned} \quad (16)$$

$$\begin{aligned} \text{MSE}_{ij} &= \int_{U_j} (f_i(x) - w_{ij}^2)^2 p(x) dx \\ &\propto p(m_j) \int_{U_j} (\text{grad} f_i(m_j))' (x - m_j))^2 dx \\ &= p(m_j) \int_{U_j} v_i' (x - m_j) (x - m_j)' v_i dx \\ &= p(m_j) v_i' C_j v_i \end{aligned} \quad (17)$$

ここに， $v_i = \text{grad} f_i(m_j)$ ， $C_j = \int_{U_j} (x - m_j)(x - m_j)'$ dx である．

本節でも，隠れユニットは，局所的には最密構造になっているという条件下で MSE を最小にする分布を考えることにする．つまり， $\text{grad} f_i(m_j)$ と直交する方向に薄く延びたポロノイ領域を作るようなことは考えない（そうすれば同じユニッ

トの密度で MSE_{ij} を減らすことができる)。なぜなら、その様な配列はヘブ学習によってつくることができないからである。しかし、最密構造のポロノイ領域は球ではないので、これをどの方向に向けるかという問題は残る（つまり、 $L=2$ の場合ならば、六角形の角を $\text{grad} f_i(m_j)$ に関してどちらに向けるかという問題）。これに関してはつぎの補題がある。

補題

ポロノイ領域が最密構造になっているとき、 $v_i^T C_j v_i$ は、 v_i の方向によらない。

略証

C_j は対称行列なので、適当な直交行列 O を用いて、 $C_j = O^T \Lambda O$ 、 $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_L)$ と分解できる。 V_j の対称性から、いくつかの直交行列 O_0 に対して、 $O_0^T C_j O_0 = C_j$ となる。これより、 $Q = O O_0^T$ とすれば、

$$\Lambda = Q^T \Lambda Q \quad (18)$$

ここで、一般性を失うことなく、 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$ と仮定

する．式(18)の(1, 1)要素に着目すると，

$$\lambda_1 = \sum_k \lambda_k q_{k1}^2 \quad (19)$$

Q も直交行列だから， $\sum_k q_{k1}^2 = 1$ ．よって， $q_{k1} \neq 0$ であるような $k = 1, 2, \dots, L$ に対して， $\lambda_k = \lambda_1$ となる．すべての k に対して適当な O がとれて，これが証明できる．

q. e. d.

この補題によりMSEを最小にする隠れユニットの配置問題は， f_j の変化の二次以上のオーダーを無視すれば，密度 $\rho(x)$ のみを問題にすればよいことがわかる．

式(17)から，式(11)と同様に，

$$MSE_{ij} \propto p(m_j) \|v_i(m_j)\|^2 \rho(m_j)^{-(n+2)/n} \quad (20)$$

をえる．これよりMSEは，

$$\begin{aligned} MSE &= \sum_{i,j} MSE_{ij} \\ &\propto \int_{S^L} p(x) \sum_i \|v_i(x)\|^2 \rho(x)^{-(n+2)/n} \rho(x) dx \\ &\propto \int_{S^L} z(x) \rho(x)^{-2/n} p(x) dx \end{aligned} \quad (21)$$

ここに $z(x) = \sum_i \|v_i(x)\|^2 = \sum_i \|\text{grad} f_i(x)\|^2$ であり
これは, x S^L 付近で, 関数 $f(x)$ が, どのくらい速く変化
しているかを表している. これはこの付近での, この関数の
学習のむずかしさと考えていいだろう.

式(21)を最小にする ρ は, 式(13)と同様に

$$\rho(x) \propto \{z(x)p(x)\}^{L/(L+2)} \quad (22)$$

と求めることができる.

5. まとめ

式(12)は, 式(13)の $p(x)$ を $z(x)p(x)$ で置き換えたもの
にほかならない. 従ってつぎのことがいえる. ある入力分布
 $p(x)$ に対して MSE を最小にしたいのなら学習の際は入力分布
として $p(x)$ の代わりに, $z(x)p(x)$ 使うべきである. これを
一言でいえば, "むずかしいことは何度も練習すべし" とい
うことである.

本稿では, 入力 x は S^L 上全体に分布している仮定したが S^L
上の次元の低い部分多様体上に分布していたとしても, 同様
に成り立つ. ただしその場合は L は部分多様体の次元となる.

文献

1. 甘利俊一 神経回路網の数理. (1978) 産業図書.
2. M. Iri, K. Murota, T. Ohya (1984) "A Fast Voronoi-Diagram Algorithm with Applications to Geographical Optimization Problems." , Proc. 11th IFIP Conference on System Modelling and Optimization 1983, Copenhagen, Lecture Notes in Controle and Information Science 59, "System Modelling and Optimization" (P. Thoft-Christensen, ed.), Springer-Verlag, Berlin, 273-288.