

演習問題解答 (続々・わかりやすいパターン認識)

第 1 章

【演習問題 1.1】

式 (1.38) で示されているパーセプトロンの重み修正法を，識別関数の修正法として表す．そのため，式 (1.38) を転置して右より $\mathbf{x}$ を乗ずると

$$\begin{cases} g(\mathbf{x})' = g(\mathbf{x}) + \rho \cdot b_k \mathbf{x}_k^t \mathbf{x} & (b_k g(\mathbf{x}_k) \leq 0) \\ g(\mathbf{x})' = g(\mathbf{x}) & (\text{otherwise}) \end{cases} \quad (\text{S.1.1})$$

を得る．ここで， $g(\mathbf{x})$ ， $g(\mathbf{x})'$ は，それぞれ修正前，修正後の線形識別関数を表している．

クラス ω_1 に属すパターン $\mathbf{x}_k$ を識別したとき，識別関数の修正が必要となるのは， $g(\mathbf{x}_k) \leq 0$ となって $\mathbf{x}_k$ が正しく識別できなかった場合である．正しい識別結果が得られるようにするには，識別関数値がより大きな値となるような修正が必要である．式 (S.1.1) の第 1 式の修正を施すと， $\mathbf{x}_k \in \omega_1$ より $b_k = 1$ であるから

$$\begin{aligned} g(\mathbf{x}_k)' &= g(\mathbf{x}_k) + \rho \cdot b_k \mathbf{x}_k^t \mathbf{x}_k \\ &= g(\mathbf{x}_k) + \rho \cdot \|\mathbf{x}_k\|^2 \geq g(\mathbf{x}_k) \end{aligned} \quad (\text{S.1.2})$$

が成り立つ．すなわち，修正後の識別関数値 $g(\mathbf{x}_k)'$ は元の識別関数値 $g(\mathbf{x}_k)$ 以上となり，識別関数の改善が図られていることが確かめられる．同様にして，クラス ω_2 に属すパターン $\mathbf{x}_k$ が正しく識別できなかった場合の修正では， $b_k = -1$ とすることにより $g(\mathbf{x}_k)' \leq g(\mathbf{x}_k)$ が成り立つ．以上より，1 回の修正で識別関数は必ず改善される方向に更新されることがわかる．

【演習問題 1.2】

与えられた線形識別関数は

$$g(x) = w_0 + w^t x \quad (\text{S.1.3})$$

である．第 1 章の式 (1.30) で示したように， $x \in \omega_1$ に対しては， $g(x) > 0$ ， $x \in \omega_2$ に対しては， $g(x) < 0$ が成り立たなくてはならない．本問題を 2 次元特徴空間 (x_1, x_2) を用いて示したのが図 S.1.1 である．決定境界 $g(x) = 0$ が超平面（この図では直線）として示されている．この超平面により特徴空間は， $g(x) > 0$ の側（図では直線より上）と， $g(x) < 0$ の側（図では直線より下）の二つの領域に分割される．この超平面上に存在する x に対しては $g(x) = 0$ となる．

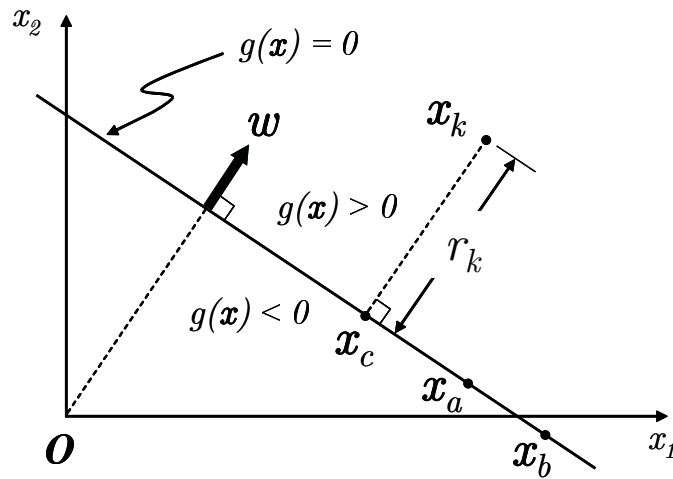


図 S.1.1: 決定境界までの距離 r_k

いま, 2 つの特徴ベクトル x_a および x_b がいずれもこの超平面に存在するなら

$$g(x_a) = w_0 + \mathbf{w}^t \mathbf{x}_a = 0 \quad (\text{S.1.4})$$

$$g(x_b) = w_0 + \mathbf{w}^t \mathbf{x}_b = 0 \quad (\text{S.1.5})$$

が成り立つ．上式を辺々引くと

$$\mathbf{w}^t (\mathbf{x}_a - \mathbf{x}_b) = 0 \quad (\text{S.1.6})$$

が得られる．上式は, 超平面上の任意のベクトルが $\mathbf{w}$ と直交することを示している．したがって, $\mathbf{w}$ はこの超平面の法線ベクトルであることがわかる ([改訂版] の図 2.8 を参照のこと)．図から明らかなように, 法線ベクトルの方向が $g(\mathbf{x}) > 0$ の側であり, 法線ベクトルと反対の方向が $g(\mathbf{x}) < 0$ の側である．ここで, x_k から超平面上へ垂直に下ろした射影点を x_c とすると, $r_k (\geq 0)$ は x_k と x_c の距離であるので, x_k は法線ベクトル $\mathbf{w}$ を用いて

$$\mathbf{x}_k = \mathbf{x}_c \pm r_k \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (\text{S.1.7})$$

と書ける．上式の複号 $\pm$ は, x_k が $g(\mathbf{x}) > 0$ の側にあるときは $+$ を, x_k が $g(\mathbf{x}) < 0$ の側にあるときは $-$ をとることを示している．また, x_c は超平面上にあるので, 下式が成り立つ．

$$g(\mathbf{x}_c) = w_0 + \mathbf{w}^t \mathbf{x}_c = 0 \quad (\text{S.1.8})$$

式 (S.1.7), (S.1.8) より x_c を消去すべく, 式 (S.1.7) の両辺に左から $\mathbf{w}^t$ を乗じ, 続いて式 (S.1.8) を用いると

$$\mathbf{w}^t \mathbf{x}_k = \mathbf{w}^t \mathbf{x}_c \pm r_k \frac{\mathbf{w}^t \mathbf{w}}{\|\mathbf{w}\|} \quad (\text{S.1.9})$$

$$= -w_0 \pm r_k \|\mathbf{w}\| \quad (\text{S.1.10})$$

が得られる．これより

$$\begin{aligned} \pm r_k \|\mathbf{w}\| &= w_0 + \mathbf{w}^t \mathbf{x}_k \\ &= g(\mathbf{x}_k) \end{aligned} \quad (\text{S.1.11})$$

となり, 最終的に下式が得られる．

$$r_k = \frac{|g(\mathbf{x}_k)|}{\|\mathbf{w}\|} \quad (\text{S.1.12})$$

【演習問題 1.3】

問 (1)

学習パターンを 2 次元特徴空間上にプロットした結果を図 S.1.2 に示す．クラス ω_1, ω_2 のパターンを，それぞれ $\circ, \triangle$ で示した．図からわかるように，パターンは線形分離可能である．

指定された条件で学習を行うと，繰り返し数 68 回（エポック数 9）で収束し，寄与ベクトル数は 4 である．得られた重みは下式となる．

$$\mathbf{w} = (w_0, w_1, w_2)^t = (14, -7, 4)^t \quad (\text{S.1.13})$$

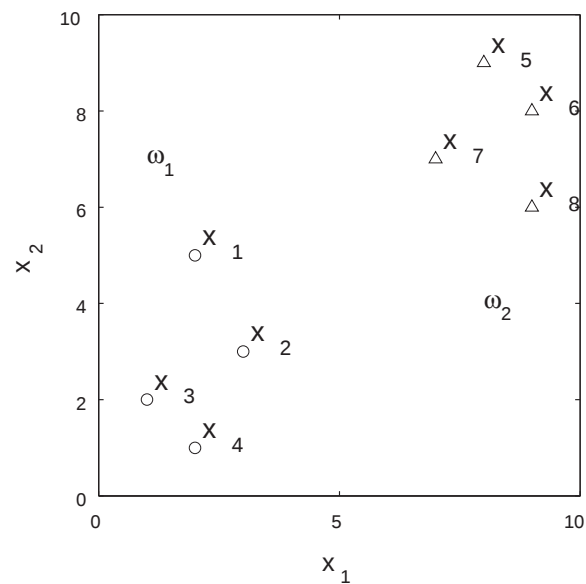


図 S.1.2: 線形分離可能な学習パターン

問 (2)

図 S.1.3 では，学習の過程で寄与ベクトルとなったパターンを黒く塗りつぶして示した．繰り返し数 iter ，寄与ベクトル数 m を図中に示した．

得られた重みによる決定境界 H_0 を図中の太線で示した．この決定境界に最も近接するパターンをクラス ω_1, ω_2 毎に求めると，それぞれ x_4, x_5 であり，いずれも寄与ベクトルである．これらの近接パターンを含み， H_0 に平行な超平面が求めるべき H_1, H_2 であり，それらを図の細線で示した．

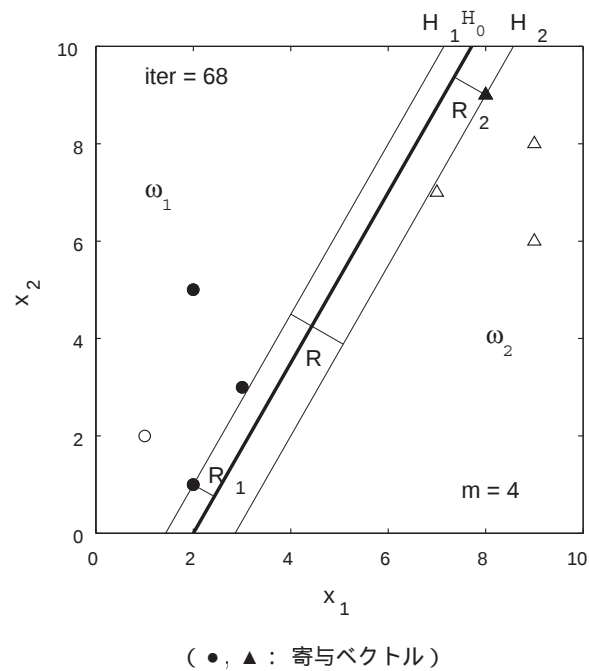


図 S.1.3: パーセプトロンによる決定境界 ($\delta = 0$)

近接パターン x_4, x_5 から H_0 までの距離をそれぞれ R_1, R_2 で表すと

$$\begin{cases} R_1 = 0.496 \\ R_2 = 0.744 \end{cases} \quad (\text{S.1.14})$$

である．求めるべきマージン R は， H_1, H_2 間の距離であるので

$$R = R_1 + R_2 = 0.496 + 0.744 = 1.240 \quad (\text{S.1.15})$$

となる．

【演習問題 1.4】

指定されたとおり $\delta = 1$ に設定し，他の条件は前問と同一にして式 (1.47) により学習を行った結果，繰返し数 466 回（エポック数 59）で収束し，寄与ベクトル数は 4 であり，それらを黒く塗りつぶして示した．得られた重みは

$$\mathbf{w} = (w_0, w_1, w_2)^t = (92, -29, 10)^t \quad (\text{S.1.16})$$

である．前問と同様の要領で決定境界 H_0 と，マージンを決定する超平面 H_1, H_2 を求め，図 S.1.4 に示した．この決定境界に最も近接するパターンを求めると，クラス ω_1 では x_2 であり寄与ベクトルであるが， ω_2 では x_7 であり寄与ベクトルでない（第 1 章の脚注*6 参照）．近接パターン x_2, x_7 から H_0 までの距離 R_1, R_2 はそれぞれ

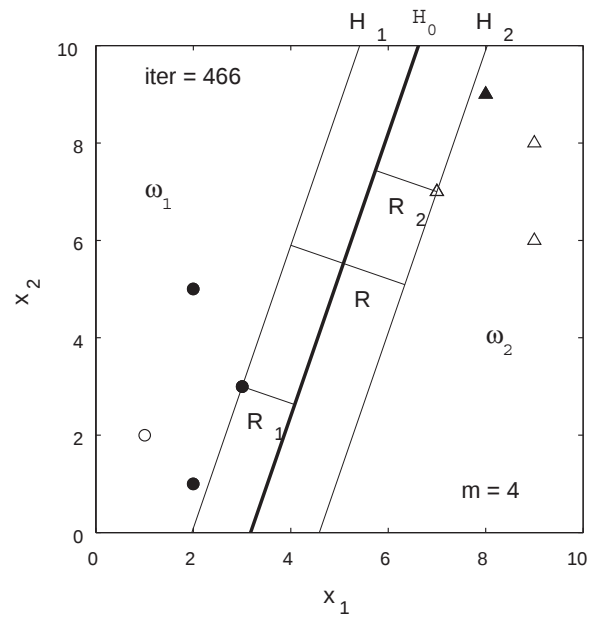
$$\begin{cases} R_1 = 1.141 \\ R_2 = 1.337 \end{cases} \quad (\text{S.1.17})$$

と求められ，マージン R は

$$R = R_1 + R_2 = 1.141 + 1.337 = 2.478 \quad (\text{S.1.18})$$

である．

図 S.1.4 を前問で得られた図 S.1.3 と比較すると，次のことがわかる．すなわち，前問では δ を設定していないので，マージンの大小に関わらず，全パターンを正しく識別できた時点で学習が終了する．そのため，決定境界とパターンとの距離が十分確保できない結果となることも少なくない．実際，図 S.1.3 では，



(●, ▲ : 寄与ベクトル)

図 S.1.4: パーセプトロンによる決定境界 ($\delta = 1$)

x_4 は H_0 に極めて接近しており、未知パターンに対して誤識別を招く可能性が高い。また、本データのように両クラスの分布が十分離れている場合には決定境界設定の自由度が大きく、そのため学習によって得られる決定境界はバラツキが大きい。

一方、図 S.1.4 では、各パターンと H_0 との距離は、少なくとも $\delta = 1$ 以上を確保しているので、図 S.1.3 よりも安定な識別が可能となる。このことは、マージンの値、1.240 と 2.478 の差としても現れている。また、 $\delta = 1$ という制約が課せられているので、収束までの繰り返し数は増大するものの、得られる決定境界のバラツキは抑制される。

第 2 章

【演習問題 2.1】

拡張特徴ベクトル $\mathbf{x}_k$ ($k = 1, \dots, n$) および拡張重みベクトル $\mathbf{w}$ は、ともに $(d+1)$ 次元のベクトルであり、式 (1.16) より

$$\mathbf{x}_k = (x_{k0}, x_{k1}, \dots, x_{kd})^t \quad (\text{ただし } x_{k0} \equiv 1) \quad (\text{S.2.1})$$

$$\mathbf{w} = (w_0, w_1, \dots, w_d)^t \quad (\text{S.2.2})$$

と書ける。ただし、 $\mathbf{x}_k$ の要素を x_{kj} ($j = 0, 1, \dots, d$) と記した。ここで式 (2.25) で示した評価関数 $J_a(\mathbf{w})$ 、および式 (2.26) で定義した $n \times (d+1)$ のパターン行列 $\mathbf{X}$ を以下に再掲する。

$$J_a(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^n (\mathbf{w}^t \mathbf{x}_k - b_k)^2 \quad (\text{S.2.3})$$

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^t \quad (\text{S.2.4})$$

式 (S.2.3) の $J_a(\mathbf{w})$ を w_i で微分すると

$$\frac{\partial J_a(\mathbf{w})}{\partial w_i} = \sum_{k=1}^n (\mathbf{w}^t \mathbf{x}_k - b_k) x_{ki} \quad (\text{S.2.5})$$

が得られる。上式をさらに w_j で微分し、2 回微分の結果を h_{ij} と置くと

$$h_{ij} = \frac{\partial^2 J_a(\mathbf{w})}{\partial w_i \partial w_j} \quad (\text{S.2.6})$$

$$= \sum_{k=1}^n x_{ki} x_{kj} \quad (i, j = 0, 1, \dots, d) \quad (\text{S.2.7})$$

となる。ここで $\mathbf{H} = (h_{ij})$ なる行列を定義する。すなわち $\mathbf{H}$ は、 (i, j) 成分が h_{ij} である $(d+1) \times (d+1)$ の行列である。行列 $\mathbf{H}$ を $J_a(\mathbf{w})$ のヘッセ行列 (Hessian) という。式 (S.2.7) の右辺は、 $(d+1) \times (d+1)$ の行列 $\mathbf{X}^t \mathbf{X}$ の (i, j) 成分であることがわかる。したがって

$$\mathbf{H} = \mathbf{X}^t \mathbf{X} \quad (\text{S.2.8})$$

が成り立つ。

ここで, $y \neq 0$ なる任意の $(d+1)$ 次元列ベクトル y に対し, $y^t H y$ を計算すると, 式 (S.2.8) より

$$\begin{aligned} y^t H y &= y^t X^t X y \\ &= (X y)^t X y \\ &= \|X y\|^2 > 0 \end{aligned} \quad (\text{S.2.9})$$

が成り立つ. したがって, ヘッセ行列 H は正定値 (positive definite) である. ヘッセ行列 H が正定値であるので, 評価関数 $J_a(w)$ は狭義凸関数 (strictly convex function) である.

以上より, 評価関数 $J_a(w)$ を最小化する問題は凸計画問題 (convex programming problem) であり, しかも $J_a(w)$ は狭義凸関数であるので, 式 (2.28) で示された解は大域的最適解であり, 唯一の最小点である.

以上述べた証明には, 凸計画問題と最適化に関する知識が必要となる. 詳細は, [続編] の付録 A.1 を参照されたい.

【演習問題 2.2】

図 S.2.1 に, クラス $\omega_1, \omega_2, \omega_3$ のパターンを, それぞれ $\circ, \triangle, \square$ の印でプロットし, 境界比の大きさを塗りつぶしの濃さで表した. 図の見方は図 2.8 と同様で, 境界比の大きいパターンほど濃く塗り潰されている.

以下では一例として, クラス ω_1 に属するパターン x_1 の境界比を求めてみよう. ここで ω_1 とは異なるクラス, すなわち ω_2 あるいは ω_3 に属するパターンのうち, x_1 の最近傍となるパターンを求めると $x_6 (\in \omega_2)$ である. 次に, x_1 と同一のクラス ω_1 の中で x_6 の最近傍となるパターンを求めると x_4 である. したがって式 (2.34) おいて, $x = x_1, y = x_6, x' = x_4$ と置くことにより, パターン x_1 の境界比 $r(x_1)$ は

$$r(x_1) = \frac{\|x_4 - x_6\|}{\|x_1 - x_6\|} = \frac{\sqrt{8}}{\sqrt{17}} = 0.686 \quad (\text{S.2.10})$$

と求めることができる.

同様にして, パターン x_4 の境界比は, $x = x_4, y = x_6, x' = x_4$ であるので, 下式となる.

$$r(x_4) = \frac{\|x_4 - x_6\|}{\|x_4 - x_6\|} = 1 \quad (\text{S.2.11})$$

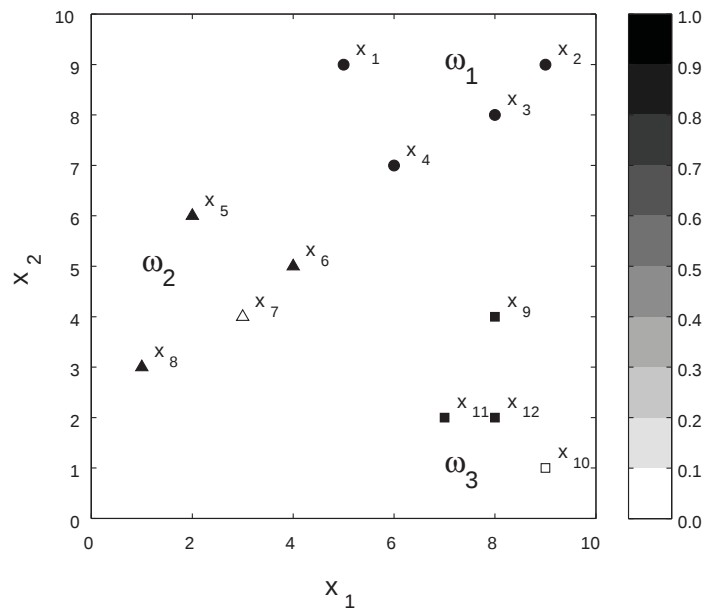


図 S.2.1: 境界比の計算結果 (3 クラス)

他のパターンの境界比も同様にして求めることができ、まとめると以下の結果となる。

$$\left. \begin{aligned} r(x_1) &= 0.686, & r(x_2) &= 0.707, & r(x_3) &= 0.901, & r(x_4) &= 1 \\ r(x_5) &= 0.686, & r(x_6) &= 1, & r(x_7) &= 0.667, & r(x_8) &= 0.697 \\ r(x_9) &= 1, & r(x_{10}) &= 0.644, & r(x_{11}) &= 0.972, & r(x_{12}) &= 0.825 \end{aligned} \right\}$$

以上より明らかなように、境界比が最大値 1 をとるのは、 $x_4 (\in \omega_1)$, $x_6 (\in \omega_2)$, $x_9 (\in \omega_3)$ の 3 パターンで、これらはいずれもクラス間を分離する決定境界に近接しており、妥当な結果である。

【演習問題 2.3】

圧縮型最近傍決定則の処理は，2.6 節 で示した手順に従う．処理の過程を表 S.2.1 に示す．最初に， x_1 をパターンリストから削除し，プロトタイプリストに登録する．それがこの表の繰り返し 0 に示した所度登録である．登録されたプロトタイプは x_1 のみであるので，クラス ω_1 のパターン x_2, x_3, x_4 はすべて正しく識別できる．以上が繰り返し 3 までである．続いてクラス ω_2 のパターン x_5 を識別したときは，唯一のプロトタイプ x_1 が最近傍となり，クラス ω_1 と判

表 S.2.1: 圧縮型最近傍決定則の処理経過

繰り返し	学習 パターン	所属 クラス	正 ○ 誤 ×	プロト タイプ	処理
0	x_1	ω_1	—	x_1	x_1 を初度登録
1	x_2	ω_1	○	x_1	
2	x_3	ω_1	○	x_1	
3	x_4	ω_1	○	x_1	
4	x_5	ω_2	×	x_1	x_5 を追加登録
5	x_6	ω_2	○	x_1, x_5	
6	x_7	ω_2	○	x_1, x_5	
7	x_8	ω_2	○	x_1, x_5	
8	x_9	ω_3	×	x_1, x_5	x_9 を追加登録
9	x_{10}	ω_3	○	x_1, x_5, x_9	
10	x_{11}	ω_3	○	x_1, x_5, x_9	
11	x_{12}	ω_3	○	x_1, x_5, x_9	
12	x_2	ω_1	○	x_1, x_5, x_9	収束
13	x_3	ω_1	○	x_1, x_5, x_9	
14	x_4	ω_1	○	x_1, x_5, x_9	
15	x_6	ω_2	○	x_1, x_5, x_9	
16	x_7	ω_2	○	x_1, x_5, x_9	
17	x_8	ω_2	○	x_1, x_5, x_9	

定され誤識別となる．そこでパターン x_5 は繰り返し 4 の時点でパターンリストからプロトタイプリストに移される．クラス ω_2 のパターン x_6, x_7, x_8 の最近傍は x_5 であるので，すべて正しく識別できる．続いてクラス ω_3 のパターン x_9 を識別したときは， ω_3 に属するプロトタイプが無く，最近傍のプロトタイプは x_1 であるので，クラス ω_1 と判定され誤識別となる．そこでパターン x_9 は繰り返し 8 の時点でパターンリストからプロトタイプリストに移される．その後，繰り返し 9 から 17 まですべて正しく識別され，処理は繰り返し 17 で収束に至る．その結果，プロトタイプとしては x_1, x_5, x_9 が選ばれ，決定境界は図 S.2.2 のようになる．図では，プロトタイプを黒くぬりつぶした．

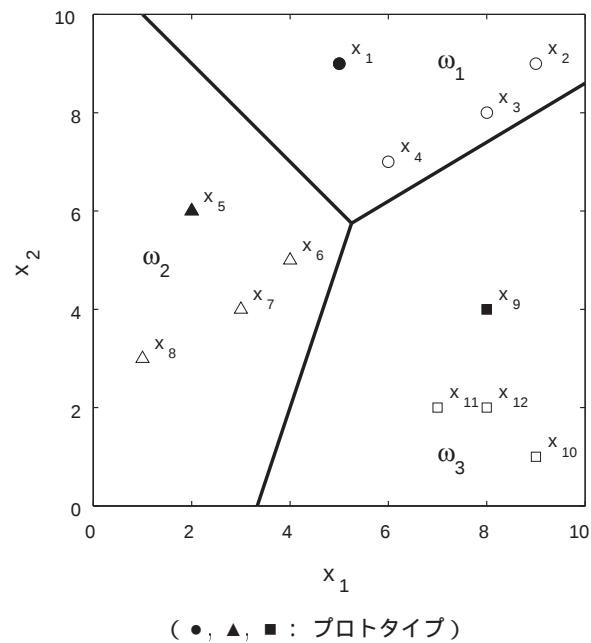


図 S.2.2: 圧縮型最近傍決定則による決定境界とプロトタイプ

第 3 章

【演習問題 3.1】

本問題の ${}_dH_k$ は, k 個の同じ品物を, 異なる d 人に分配するときの分配方法の数の等しい. 具体例として, $k = 5, d = 3$ の場合を取りあげると, 式 (3.40) は

$$\begin{cases} r_1 + r_2 + r_3 = 5 \\ r_1, r_2, r_3 \geq 0 \end{cases} \quad (\text{S.3.1})$$

となり, 求めるべきは ${}_3H_5$ である. すなわち, 5 個の同じ品物を異なる 3 人で分配する方法の数である. そのためには, 5 個の品物を左から右に一列に並べ, その中に 2 個の仕切を設定して 3 分割すればよい. その状況を図 S.3.1 の (a) ~ (d) に示した. 図では, 品物を $\circ$ 印, 仕切を縦棒で示している.

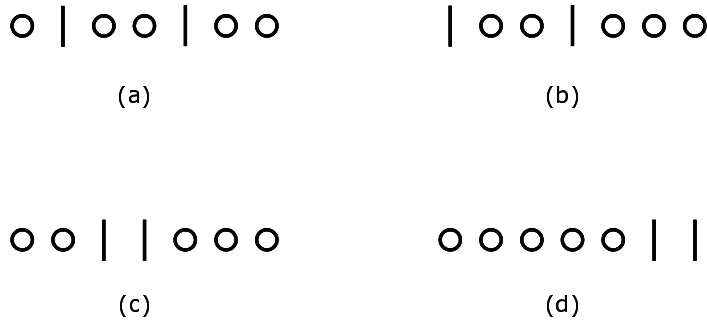


図 S.3.1: k 個の品物を d 人に分配する方法 ($k = 5, d = 3$)

分配数 r_1, r_2, r_3 は, 例えば左から順に r_1, r_2, r_3 と決めておけばよい. 図の分配例 (a) では, $(r_1, r_2, r_3) = (1, 2, 2)$ となる. 分配数 0 を含む例が (b), (c), (d) であり, (r_1, r_2, r_3) は, それぞれ $(0, 2, 3)$, $(2, 0, 3)$, $(5, 0, 0)$ である.

この分配数は, 品物の数 5 と仕切の数 2 の合計 7 個を一列に並べる方法の数である. 一列に並べるには, 7 個の場所を用意し, その中から仕切の場所 2 個を選び, 残った 5 個の場所に品物を割り当てればよい. あるいは, 品物の場所

を先に選んだ後に仕切の場所を割り当ててもよい．したがって，並べ方の数は， ${}_7C_2 = {}_7C_5$ となる．

以上を一般的な場合に拡張すると， k 個の品物と $(d-1)$ 個の仕切を一行に並べる方法の数であるので，その数は ${}_{d+k-1}C_k$ である．すなわち

$${}_dH_k = {}_{d+k-1}C_k \quad (\text{S.3.2})$$

となり，式 (3.41) が証明された．

【演習問題 3.2】

問 (1)

ここで取りあげる p 次の多項式識別関数は，式 (3.6) で示したように，異なる D 個の同次積 (homogeneous product)， $\phi_1(x), \dots, \phi_D(x)$ の線形和として表せる．同時積は， $x_1^{r_1} \cdots x_d^{r_d}$ として表すことができ，下式が成り立つ．

$$\begin{cases} \sum_{j=1}^d r_j \leq p \\ r_j \geq 0 \quad (j = 1, \dots, d) \end{cases} \quad (\text{S.3.3})$$

ただし， r_j ($j = 1, \dots, d$) は整数である．

以下では理解しやすいよう，具体例を用いて説明する．そこで， $d = 2, p = 3$ の場合を取りあげると，式 (3.7) ~ (3.9) で示したように，

$$\begin{cases} 0 \text{ 次の同次積:} & 1 \\ 1 \text{ 次の同次積:} & x_1, \quad x_2 \\ 2 \text{ 次の同次積:} & x_1^2, \quad x_1 x_2, \quad x_2^2 \\ 3 \text{ 次の同次積:} & x_1^2 x_2, \quad x_1 x_2^2, \quad x_1^3, \quad x_2^3 \end{cases} \quad (\text{S.3.4})$$

となり，同次積の数は合計 10 であるで， $D = 10$ である．

同次積の数は，式 (S.3.3) を満たす $(r_1, r_2, \dots, r_d)$ の組み合わせの数である．同次積のうち， k 次の同次積の数は

$$\begin{cases} r_1 + r_2 + \dots + r_d = k \\ r_j \geq 0 \quad (j = 1, \dots, d) \end{cases} \quad (\text{S.3.5})$$

を満たす整数 $(r_1, r_2, \dots, r_d)$ の組み合わせの数である．この数は演習問題 3.1 で明らかにしたように

$${}_dH_k = {}_{d+k-1}C_k \quad (\text{S.3.6})$$

である．たとえば上の具体例で 3 次の同次積の数は，式 (S.3.5) で $d = 2, k = 3$ と置き

$$r_1 + r_2 = 3 \quad (\text{S.3.7})$$

を満たす非負の整数 (r_1, r_2) の組み合わせの数である．これらは

$$(r_1, r_2) = (2, 1), (1, 2), (3, 0), (0, 3) \quad (\text{S.3.8})$$

と 4 個存在し，式 (S.3.4) の

$$x_1^2 x_2, \quad x_1 x_2^2, \quad x_1^3, \quad x_2^3$$

に対応している．実際，式 (S.3.6) において $d = 2, k = 3$ と置くと

$${}_2H_3 = {}_4C_3 = 4 \quad (\text{S.3.9})$$

が得られる．

式 (S.3.3) に示すように，同次積 k としては， $k = 0, 1, 2, \dots, p$ と $(p+1)$ 種存在するので，式 (S.3.6) より $f(d, p)$ は

$$f(d, p) = \sum_{k=0}^p {}_dH_k = \sum_{k=0}^p {}_{d+k-1}C_k \quad (\text{S.3.10})$$

として表せ，式 (3.42) が成り立つことがわかる．

次に，上式から式 (3.10) が導出できることを示す．ここで，組み合わせに関する公式

$${}_nC_r = {}_{n-1}C_r + {}_{n-1}C_{r-1} \quad (\text{S.3.11})$$

を用いる．上式が成り立つことは簡単に確かめられる．本公式で $r = n$ とすると右辺の第 1 項は ${}_{n-1}C_n$ となり， $r = 0$ とすると右辺の第 2 項は ${}_{n-1}C_{-1}$ となるなど数学的に意味のない項が出現する．しかし，これらはすべてゼロとみなして無視することにすれば，本公式はこのような場合にもそのまま成り立つ．

式 (S.3.11) より

$${}_{n-1}C_r = {}_nC_r - {}_{n-1}C_{r-1} \quad (\text{S.3.12})$$

が得られる．本式において $r \rightarrow k$, $n \rightarrow d+k$ と置換すると

$${}_{d+k-1}C_k = {}_{d+k}C_k - {}_{d+k-1}C_{k-1} \quad (\text{S.3.13})$$

が得られる．上式と式 (S.3.10) より

$$f(d, p) = \sum_{k=0}^p {}_{d+k-1}C_k \quad (\text{S.3.14})$$

$$= \sum_{k=0}^p {}_{d+k}C_k - \sum_{k=0}^p {}_{d+k-1}C_{k-1} \quad (\text{S.3.15})$$

$$= dC_0 + {}_{d+1}C_1 + \cdots + {}_{d+p-1}C_{p-1} + {}_{d+p}C_p \\ - dC_0 - {}_{d+1}C_1 - \cdots - {}_{d+p-1}C_{p-1} \quad (\text{S.3.16})$$

$$= {}_{d+p}C_p \quad (\text{S.3.17})$$

が得られ，式 (3.10) が導出できる．なお，式 (S.3.15) の第 2 項からは ${}_{d-1}C_{-1}$ の項が出現するが，この項が無視できることはすでに述べたとおりであるので，式 (S.3.16) では省いてある．

問 (2)

上記の問 (1) と同様に，具体例として， $d = 2$, $p = 3$ を用いて説明する．式 (S.3.4) までは問 (1) と共通である．まず，式 (3.43) を導出する．そのため同次積の集合を， x_d (具体例では x_2) を含まないグループ 1 と， x_d を含むグループ 2 の，二つのグループに分けるものとする．

具体例では，グループ 1 は x_2 を含まない同次積であるので，

$$1, \quad x_1, \quad x_1^2, \quad x_1^3$$

の 4 個がこれに該当する．一般に x_d を含まない同次積の数は， $f(d-1, p)$ に等しいことがわかる．

また，具体例でグループ 2 は x_2 を含む同次積であるので，

$$x_2, \quad x_1x_2, \quad x_2^2, \quad x_1^2x_2, \quad x_1x_2^2, \quad x_2^3$$

の 6 個がこれに該当する．これらはすべて x_d を含むので， x_d で除することができる．この例では x_2 で除することができて，その結果は

$$1, \quad x_1, \quad x_2, \quad x_1^2, \quad x_1x_2, \quad x_2^2$$

となる．これより，一般に x_d を含まない同次積の数は $f(d, p-1)$ に等しいことがわかる．

求めるべき $f(d, p)$ は，上記二つのグループの同次積の個数の和であるから，

$$f(d, p) = f(d-1, p) + f(d, p-1) \quad (\text{S.3.18})$$

が成り立ち，式 (3.43) が導出できた．

次に数学的帰納法を用いて式 (3.10) を証明する．

まず， $d=1$ の場合，異なる同次積としては， $1, x_1, x_1^2, \dots, x_1^p$ があり，その個数は $(p+1)$ であるので

$$f(1, p) = p+1 \quad (\text{S.3.19})$$

である．また，式 (3.10) の右辺で $d=1$ と置くと

$${}_{1+p}C_p = p+1 \quad (\text{S.3.20})$$

であるので

$$f(1, p) = {}_{1+p}C_p \quad (\text{S.3.21})$$

となり， $d=1$ に対して式 (3.10) が成り立つ．

次に $p=1$ の場合，異なる同次積としては， $1, x_1, \dots, x_d$ があり，その個数は $(d+1)$ であるので

$$f(d, 1) = d+1 \quad (\text{S.3.22})$$

である．また，式 (3.10) の右辺で $p=1$ と置くと

$${}_{d+1}C_1 = d+1 \quad (\text{S.3.23})$$

であるので

$$f(d, 1) = {}_{d+1}C_1 \quad (\text{S.3.24})$$

となり, $p = 1$ に対して式 (3.10) が成り立つ.

式 (3.10) が, $(d - 1)$ および $(p - 1)$ に対して成り立つと仮定すると

$$f(d - 1, p) = {}_{d+p-1}C_p \quad (\text{S.3.25})$$

$$f(d, p - 1) = {}_{d+p-1}C_{p-1} \quad (\text{S.3.26})$$

となる. これらを式 (S.3.18) に代入すると

$$f(d, p) = {}_{d+p-1}C_p + {}_{d+p-1}C_{p-1} \quad (\text{S.3.27})$$

$$= \frac{(d + p - 1)!}{(d - 1)! p!} + \frac{(d + p - 1)!}{d! (p - 1)!} \quad (\text{S.3.28})$$

$$= \frac{d(d + p - 1)! + p(d + p - 1)!}{d! p!}$$

$$= \frac{(d + p - 1)! (d + p)}{d! p!}$$

$$= \frac{(d + p)!}{d! p!}$$

$$= {}_{d+p}C_p \quad (\text{S.3.29})$$

が得られ, d, p に対しても式 (3.10) が成り立つ. したがって, 数学的帰納法より, 式 (3.10) が証明された.

【演習問題 3.3】

問 (1)

式 (3.33) を用いると, 2 次元特徴ベクトル $x_1, \dots, x_6$ は非線形変換され, 以下の 10 次元特徴ベクトルとなる.

$$\left. \begin{aligned} \phi(x_1) &= (1, 1, 5, 5, 1, 25, 5, 25, 1, 125)^t \\ \phi(x_2) &= (1, 3, 2, 6, 9, 4, 18, 12, 27, 8)^t \\ \phi(x_3) &= (1, 4, 3, 12, 16, 9, 48, 36, 64, 27)^t \\ \phi(x_4) &= (1, 5, 6, 30, 25, 36, 150, 180, 125, 216)^t \\ \phi(x_5) &= (1, 2, 4, 8, 4, 16, 16, 32, 8, 64)^t \\ \phi(x_6) &= (1, 6, 1, 6, 36, 1, 36, 6, 216, 1)^t \end{aligned} \right\} \quad (\text{S.3.30})$$

これらの特徴ベクトルに対して，指定された初期重みベクトルを用い， $\rho = 1$ として基本パーセプトロンの学習規則を適用する．重みベクトル $\mathbf{w}$ の更新過程を，最初と最後の 5 エポックずつ示したのが表 S.3.1 である．本表よりわかるように，学習はエポック 69 で収束し，得られた重みベクトル $\mathbf{w}$ は下式となる．

$$\begin{aligned}\mathbf{w} &= (w_1, \dots, w_{10})^t \\ &= (81, 198, 196, 342, 410, 370, 416, -686, -174, 28)^t \quad (\text{S.3.31})\end{aligned}$$

得られた識別関数 $g(x)$ は，式 (3.22)，(3.33)，(S.3.31) より

$$g(x) = \mathbf{w}^t \phi(x) \quad (\text{S.3.32})$$

$$\begin{aligned}&= 81 + 198x_1 + 196x_2 + 342x_1x_2 + 410x_1^2 + 370x_2^2 \\ &\quad + 416x_1^2x_2 - 686x_1x_2^2 - 174x_1^3 + 28x_2^3 \quad (\text{S.3.33})\end{aligned}$$

表 S.3.1: 3 次識別関数における重みベクトル $\mathbf{w}$ の更新過程

エポック	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
0	0	0	0	0	0	0	0	0	0	0
1	0	-4	-1	-25	-24	-11	-145	-155	-124	-91
2	2	2	5	-10	-2	11	-90	-114	-40	5
3	4	5	9	-17	-1	13	-169	-221	-73	-51
4	6	11	15	-2	21	35	-114	-180	11	45
5	7	13	14	-14	21	12	-198	-312	-23	-136
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
65	80	192	195	331	380	371	345	-699	-306	33
66	81	197	196	341	401	368	395	-683	-223	4
67	81	196	197	338	398	377	384	-690	-230	65
68	81	198	196	342	410	370	416	-686	-174	28
69	81	198	196	342	410	370	416	-686	-174	28

となる．上式より求めた決定境界 $g(x) = 0$ を，元の 2 次元特徴空間上にプロットしたのが図 S.3.2 である．図では，学習により得られた決定境界を太線で示した．図中， $\circ$ ， $\triangle$ は，それぞれクラス ω_1, ω_2 に所属するパターンを示しており，これらは，線形分離不可能である．ただし，全パターンが寄与ベクトルであるので，図では黒く塗りつぶし， $\bullet$ ， $\blacktriangle$ として示した．決定境界は三次曲線であり，両クラスを正しく分離できていることが確かめられる．

問 (2)

双対パーセプトロンの学習規則を適用すると，誤り計数ベクトル α の更新過程は，表 S.3.2 のようになる．これより，最終的に得られる誤り計数ベクトル α

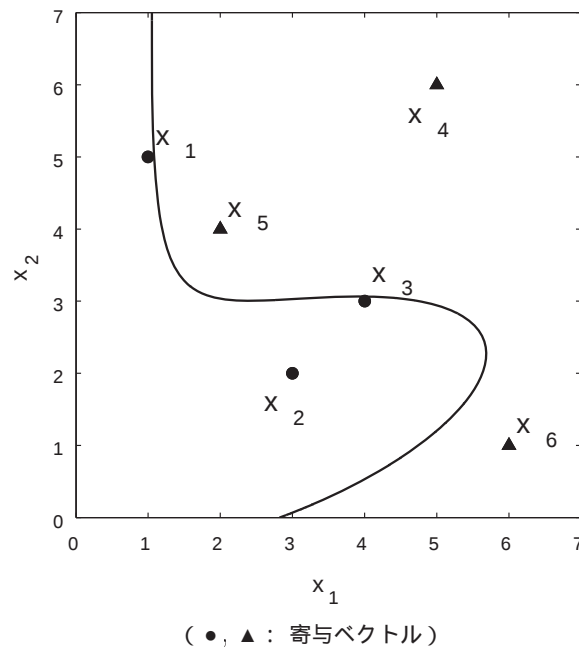


図 S.3.2: 3 次識別関数による決定境界

表 S.3.2: 3 次識別関数における誤り計数ベクトル α の更新過程

エポック	a_1	a_2	a_3	a_4	a_5	a_6
0	0	0	0	0	0	0
1	1	0	0	1	0	0
2	2	1	1	1	1	0
3	3	2	2	2	1	0
4	4	3	3	2	2	0
5	4	4	4	3	2	0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
65	31	51	63	18	32	15
66	31	52	64	18	33	15
67	32	52	64	18	34	15
68	32	52	65	18	35	15
69	32	52	65	18	35	15

は, 下式のようになる .

$$\alpha = (\alpha_1, \dots, \alpha_6)^t = (32, 52, 65, 18, 35, 15)^t \quad (\text{S.3.34})$$

全パターンが寄与ベクトルとなることは, すでに問 (1) で述べたとおりである .
上式で $\alpha_k \neq 0$ ($k = 1, \dots, 6$) であるので, そのことがここでも確かめられる .

教師信号 b_k ($k = 1, \dots, 6$) は

$$(b_1, \dots, b_6)^t = (1, 1, 1, -1, -1, -1)^t \quad (\text{S.3.35})$$

であるので, 式 (3.33), (S.3.30), (S.3.34), (S.3.35) を式 (3.28) に代入するこ

とにより，得られる識別関数 $g(x)$ は

$$g(x) = \sum_{k=1}^n \alpha_k b_k \phi(x_k)^t \phi(x) \quad (\text{S.3.36})$$

$$\begin{aligned} &= 81 + 198x_1 + 196x_2 + 342x_1x_2 + 410x_1^2 + 370x_2^2 \\ &\quad + 416x_1^2x_2 - 686x_1x_2^2 - 174x_1^3 + 28x_2^3 \end{aligned} \quad (\text{S.3.37})$$

となり，式 (S.3.33) と一致することが確かめられる．

第 4 章

【演習問題 4.1】

以下では二つの解法を示す．解法 1 は，多項定理を用いた解法，解法 2 は数学的帰納法を用いた解法である．

[解法 1]

二つの d 次元特徴ベクトル x, y を

$$\begin{cases} x = (x_1, \dots, x_d)^t \\ y = (y_1, \dots, y_d)^t \end{cases} \quad (\text{S.4.1})$$

とする．式 (4.7) において， x_k を y で置き換えた

$$K(x, y) = \exp \left(-\frac{\|x - y\|^2}{2\sigma^2} \right) \quad (\text{S.4.2})$$

が無限次元ベクトルの内積で表されることを示せばよい．関数 $K(x, y)$ は

$$\begin{aligned} K(x, y) &= \exp \left(-\frac{\|x - y\|^2}{2\sigma^2} \right) \\ &= \exp \left(-\frac{\|x\|^2}{2\sigma^2} \right) \exp \left(-\frac{\|y\|^2}{2\sigma^2} \right) \exp \left(\frac{x^t y}{\sigma^2} \right) \end{aligned} \quad (\text{S.4.3})$$

と変形できる．上式の第 3 項を $F(x, y)$ と置き，級数展開の公式

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (\text{S.4.4})$$

および多項定理

$$\begin{aligned} &(x_1 + x_2 + \dots + x_d)^k \\ &= \sum_{\substack{k_1 + k_2 + \dots + k_d = k \\ k_j \geq 0}} \frac{k!}{k_1! k_2! \dots k_d!} x_1^{k_1} x_2^{k_2} \dots x_d^{k_d} \end{aligned} \quad (\text{S.4.5})$$

を用いると

$$F(\boldsymbol{x}, \boldsymbol{y}) \stackrel{\text{def}}{=} \exp\left(\frac{\boldsymbol{x}^t \boldsymbol{y}}{\sigma^2}\right) \quad (\text{S.4.6})$$

$$\begin{aligned} &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\boldsymbol{x}^t \boldsymbol{y}}{\sigma^2}\right)^k \\ &= \sum_{k=0}^{\infty} \left(\frac{1}{\sigma^2}\right)^k \frac{1}{k!} (x_1 y_1 + \dots + x_d y_d)^k \\ &= \sum_{k=0}^{\infty} \sum_{\substack{k_1 + \dots + k_d = k \\ k_j \geq 0}} \left(\frac{1}{\sigma^2}\right)^k \frac{1}{k_1! \dots k_d!} x_1^{k_1} \dots x_d^{k_d} y_1^{k_1} \dots y_d^{k_d} \quad (\text{S.4.7}) \end{aligned}$$

$$\begin{aligned} &= \sum_{k=0}^{\infty} F_k(\boldsymbol{x}, \boldsymbol{y}) \\ &= \lim_{D \rightarrow \infty} \sum_{k=0}^D F_k(\boldsymbol{x}, \boldsymbol{y}) \quad (\text{S.4.8}) \end{aligned}$$

と書ける．ただし $F_k(\boldsymbol{x}, \boldsymbol{y})$ は下式 (S.4.9) で定義され，以下のように変形される．

$$\begin{aligned} F_k(\boldsymbol{x}, \boldsymbol{y}) &\stackrel{\text{def}}{=} \sum_{\substack{k_1 + \dots + k_d = k \\ k_j \geq 0}} \left(\frac{1}{\sigma^2}\right)^k \frac{1}{k_1! \dots k_d!} x_1^{k_1} \dots x_d^{k_d} y_1^{k_1} \dots y_d^{k_d} \quad (\text{S.4.9}) \end{aligned}$$

$$= \sum_{\substack{k_1 + \dots + k_d = k \\ k_j \geq 0}} \frac{1}{k_1! \dots k_d!} \left(\frac{x_1}{\sigma}\right)^{k_1} \dots \left(\frac{x_d}{\sigma}\right)^{k_d} \cdot \left(\frac{y_1}{\sigma}\right)^{k_1} \dots \left(\frac{y_d}{\sigma}\right)^{k_d} \quad (\text{S.4.10})$$

ここで，条件

$$\begin{cases} k_1 + k_2 + \dots + k_d = k \\ k_j \geq 0 \quad (j = 1, \dots, d) \end{cases} \quad (\text{S.4.11})$$

を満たす $k_1, k_2, \dots, k_d$ の組み合わせの数を n_k とすると，演習問題 3.1 の式 (3.41) で示したように

$$n_k = {}_d\text{H}_k = {}_{d+k-1}\text{C}_k \quad (\text{S.4.12})$$

と求められる．その組み合わせの一つ， $k_1, k_2, \dots, k_d$ に対して

$$\begin{cases} u_k(\mathbf{x}; k_1, \dots, k_d) \stackrel{\text{def}}{=} \sqrt{\frac{1}{k_1! \dots k_d!}} \left(\frac{x_1}{\sigma}\right)^{k_1} \dots \left(\frac{x_d}{\sigma}\right)^{k_d} \\ u_k(\mathbf{y}; k_1, \dots, k_d) \stackrel{\text{def}}{=} \sqrt{\frac{1}{k_1! \dots k_d!}} \left(\frac{y_1}{\sigma}\right)^{k_1} \dots \left(\frac{y_d}{\sigma}\right)^{k_d} \end{cases} \quad (\text{S.4.13})$$

を定義する．次にある k に対し， n_k 個の $u_k(\mathbf{x}; k_1, \dots, k_d)$ ， $u_k(\mathbf{y}; k_1, \dots, k_d)$ を要素とする n_k 次元ベクトルをそれぞれ $\mathbf{u}_k(\mathbf{x})$ ， $\mathbf{u}_k(\mathbf{y})$ とすると

$$F_k(\mathbf{x}, \mathbf{y}) = \sum_{\substack{k_1 + \dots + k_d = k \\ k_j \geq 0}} u_k(\mathbf{x}; k_1, \dots, k_d) \cdot u_k(\mathbf{y}; k_1, \dots, k_d) \quad (\text{S.4.14})$$

$$= \mathbf{u}_k(\mathbf{x})^t \mathbf{u}_k(\mathbf{y}) \quad (\text{S.4.15})$$

と書ける．さらに，ベクトル $\mathbf{u}_0(\mathbf{x}), \mathbf{u}_1(\mathbf{x}), \dots, \mathbf{u}_D(\mathbf{x})$ の全要素を並べたベクトルを $\mathbf{v}_D(\mathbf{x})$ とすると，その次元数 $\hat{D}$ は式 (S.4.12) の n_k を用いて

$$\hat{D} = \sum_{k=0}^D n_k \quad (\text{S.4.16})$$

となる．すなわち

$$\mathbf{v}_D(\mathbf{x}) = \underbrace{(\overbrace{\mathbf{u}_0(\mathbf{x})^t}^{n_0}, \dots, \overbrace{\mathbf{u}_k(\mathbf{x})^t}^{n_k}, \dots, \overbrace{\mathbf{u}_D(\mathbf{x})^t}^{n_D})^t}_{\hat{D}} \quad (\text{S.4.17})$$

であり，同様に $\mathbf{v}_k(\mathbf{y})$ が定義できて下式で表される．

$$\mathbf{v}_D(\mathbf{y}) = (\mathbf{u}_0(\mathbf{y})^t, \dots, \mathbf{u}_k(\mathbf{y})^t, \dots, \mathbf{u}_D(\mathbf{y})^t)^t \quad (\text{S.4.18})$$

式 (S.4.3) , (S.4.8) , (S.4.15) より

$$K(\boldsymbol{x}, \boldsymbol{y}) \quad (\text{S.4.19})$$

$$= \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{y}\|^2}{2\sigma^2}\right) F(\boldsymbol{x}, \boldsymbol{y}) \quad (\text{S.4.20})$$

$$= \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\boldsymbol{y}\|^2}{2\sigma^2}\right) \lim_{D \rightarrow \infty} \sum_{k=0}^D F_k(\boldsymbol{x}, \boldsymbol{y}) \quad (\text{S.4.21})$$

$$= \lim_{D \rightarrow \infty} \sum_{k=0}^D \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}\right) \mathbf{u}_k(\boldsymbol{x})^t \cdot \exp\left(-\frac{\|\boldsymbol{y}\|^2}{2\sigma^2}\right) \mathbf{u}_k(\boldsymbol{y}) \quad (\text{S.4.22})$$

$$= \lim_{D \rightarrow \infty} \left(\exp\left(-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}\right) \mathbf{v}_D(\boldsymbol{x}) \right)^t \left(\exp\left(-\frac{\|\boldsymbol{y}\|^2}{2\sigma^2}\right) \mathbf{v}_D(\boldsymbol{y}) \right) \quad (\text{S.4.23})$$

と書ける．そこで

$$\begin{cases} \phi(\boldsymbol{x}) \stackrel{\text{def}}{=} \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}\right) \mathbf{v}_D(\boldsymbol{x}) \\ \phi(\boldsymbol{y}) \stackrel{\text{def}}{=} \exp\left(-\frac{\|\boldsymbol{y}\|^2}{2\sigma^2}\right) \mathbf{v}_D(\boldsymbol{y}) \end{cases} \quad (\text{S.4.24})$$

とすれば

$$K(\boldsymbol{x}, \boldsymbol{y}) = \lim_{D \rightarrow \infty} \phi(\boldsymbol{x})^t \phi(\boldsymbol{y}) \quad (\text{S.4.25})$$

と書ける．ベクトル $\phi(\boldsymbol{x})$, $\phi(\boldsymbol{y})$ は , いずれも $\hat{D}$ 次元であり , $D \rightarrow \infty$ とすると $\hat{D} \rightarrow \infty$ となる．したがって , $K(\boldsymbol{x}, \boldsymbol{y})$ は無限次元ベクトルの内積で表されることが証明できた．

[解法 2]

式 (S.4.1) ~ 式 (S.4.3) は解法 1 と同じである．式 (S.4.3) の第 3 項を $F_d(\boldsymbol{x}, \boldsymbol{y})$

と置くと

$$F_d(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \exp\left(\frac{\mathbf{x}^t \mathbf{y}}{\sigma^2}\right) \quad (\text{S.4.26})$$

$$= \exp\left[(x_1 y_1 + \dots + x_d y_d)/\sigma^2\right] \quad (\text{S.4.27})$$

$$= \prod_{j=1}^d \exp\left(\frac{x_j}{\sigma} \cdot \frac{y_j}{\sigma}\right) \quad (\text{S.4.28})$$

が得られる．ここで級数展開の公式 (S.4.4) の添字 k を，0 からではなく 1 から開始するように変更し^{*1}

$$e^x = \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} \quad (\text{S.4.29})$$

を適用すると

$$\exp\left(\frac{x_j}{\sigma} \cdot \frac{y_j}{\sigma}\right) = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \left(\frac{x_j}{\sigma} \cdot \frac{y_j}{\sigma}\right)^{k-1} \quad (\text{S.4.30})$$

$$= \sum_{k=1}^{\infty} \frac{(x_j/\sigma)^{k-1}}{\sqrt{(k-1)!}} \cdot \frac{(y_j/\sigma)^{k-1}}{\sqrt{(k-1)!}} \quad (\text{S.4.31})$$

と書ける．さらに

$$\begin{cases} u_k(x_j) \stackrel{\text{def}}{=} \frac{(x_j/\sigma)^{k-1}}{\sqrt{(k-1)!}} \\ u_k(y_j) \stackrel{\text{def}}{=} \frac{(y_j/\sigma)^{k-1}}{\sqrt{(k-1)!}} \end{cases} \quad (\text{S.4.32})$$

と定義し，これらを第 k 成分として持つ D 次元ベクトルを，それぞれ $\mathbf{u}_j(\mathbf{x})$ ， $\mathbf{u}_j(\mathbf{y})$ とする．すなわち

$$\begin{cases} \mathbf{u}_j(\mathbf{x}) = (u_1(x_j), \dots, u_k(x_j), \dots, u_D(x_j)) \\ \mathbf{u}_j(\mathbf{y}) = (u_1(y_j), \dots, u_k(y_j), \dots, u_D(y_j)) \end{cases} \quad (\text{S.4.33})$$

^{*1} 本措置により，後出の式表現が簡単になる．

である．これより式 (S.4.31) は

$$\begin{aligned}\exp\left(\frac{x_j}{\sigma} \cdot \frac{y_j}{\sigma}\right) &= \lim_{D \rightarrow \infty} \sum_{k=1}^D u_k(x_j) u_k(y_j) \\ &= \lim_{D \rightarrow \infty} \mathbf{u}_j(\mathbf{x})^t \mathbf{u}_j(\mathbf{y})\end{aligned}\quad (\text{S.4.34})$$

となるので，式 (S.4.28) より

$$F_d(\mathbf{x}, \mathbf{y}) = \lim_{D \rightarrow \infty} \prod_{j=1}^d \mathbf{u}_j(\mathbf{x})^t \mathbf{u}_j(\mathbf{y}) \quad (\text{S.4.35})$$

が得られる．

以下では $F_d(\mathbf{x}, \mathbf{y})$ が二つの無限次元ベクトルの内積として表されることを，数学的帰納法によって証明する．

まず， $d = 1$ のとき，式 (S.4.35) は

$$F_1(\mathbf{x}, \mathbf{y}) = \lim_{D \rightarrow \infty} \mathbf{u}_1(\mathbf{x})^t \mathbf{u}_1(\mathbf{y}) \quad (\text{S.4.36})$$

となる．ベクトル $\mathbf{u}_1(\mathbf{x})$ ， $\mathbf{u}_1(\mathbf{y})$ は， $D \rightarrow \infty$ により無限次元ベクトルになるので， $F_1(\mathbf{x}, \mathbf{y})$ は無限次元ベクトルの内積として表されることが確かめられる．

続いて，次式のような D 次元ベクトル $\alpha(\mathbf{x})$ ， $\alpha(\mathbf{y})$ を考える．

$$\begin{cases} \alpha(\mathbf{x}) = (\alpha_1(\mathbf{x}), \dots, \alpha_i(\mathbf{x}), \dots, \alpha_D(\mathbf{x}))^t \\ \alpha(\mathbf{y}) = (\alpha_1(\mathbf{y}), \dots, \alpha_i(\mathbf{y}), \dots, \alpha_D(\mathbf{y}))^t \end{cases} \quad (\text{S.4.37})$$

上式で $\alpha_i(\mathbf{x})$ ， $\alpha_i(\mathbf{y})$ は，それぞれ d 次元ベクトル $\mathbf{x}$ ， $\mathbf{y}$ のスカラー関数であり，ベクトル $\alpha(\mathbf{x})$ ， $\alpha(\mathbf{y})$ の第 i 成分である．両ベクトルは， $D \rightarrow \infty$ により無限次元ベクトルになる．そこで， $F_d(\mathbf{x}, \mathbf{y})$ が次式のように無限次元ベクトル $\alpha(\mathbf{x})$ ， $\alpha(\mathbf{y})$ の内積として表されると仮定する．

$$\begin{aligned}F_d(\mathbf{x}, \mathbf{y}) &= \lim_{D \rightarrow \infty} \sum_{i=1}^D \alpha_i(\mathbf{x}) \alpha_i(\mathbf{y}) \\ &= \lim_{D \rightarrow \infty} \alpha(\mathbf{x})^t \alpha(\mathbf{y})\end{aligned}\quad (\text{S.4.38})$$

式 (S.4.33) , (S.4.35) , (S.4.38) を用いると , $F_{d+1}(\boldsymbol{x}, \boldsymbol{y})$ は

$$F_{d+1}(\boldsymbol{x}, \boldsymbol{y}) = \lim_{D \rightarrow \infty} F_d(\boldsymbol{x}, \boldsymbol{y}) \cdot \mathbf{u}_{d+1}(\boldsymbol{x})^t \mathbf{u}_{d+1}(\boldsymbol{y}) \quad (\text{S.4.39})$$

$$= \lim_{D \rightarrow \infty} \sum_{i=1}^D \alpha_i(\boldsymbol{x}) \alpha_i(\boldsymbol{y}) \sum_{k=1}^D u_k(x_{d+1}) u_k(y_{d+1}) \quad (\text{S.4.40})$$

$$= \lim_{D \rightarrow \infty} \sum_{i=1}^D \sum_{k=1}^D \alpha_i(\boldsymbol{x}) u_k(x_{d+1}) \alpha_i(\boldsymbol{y}) u_k(y_{d+1}) \quad (\text{S.4.41})$$

となる . ここで

$$\begin{cases} \beta_j(\boldsymbol{x}) = \alpha_i(\boldsymbol{x}) u_k(x_{d+1}) \\ \beta_j(\boldsymbol{y}) = \alpha_i(\boldsymbol{y}) u_k(y_{d+1}) \end{cases} \quad (j = 1, 2, \dots, D^2) \quad (\text{S.4.42})$$

となる $\beta_j(\boldsymbol{x})$, $\beta_j(\boldsymbol{y})$ を定義する . 式 (S.4.41) から明らかなように , 式 (S.4.42) の右辺の添字 i , k は , いずれも $1, \dots, D$ と変化する . したがって , 式 (S.4.42) の左辺の添字 j は , それに対応すべく $1, \dots, D^2$ まで変化させなくてはならない . その対応の実現例を表 S.4.1 に示した . この表で示したように , 添字 i , j , k は次のように関連付けられている . すなわち , 非負の整数 a , b があって , a を b で割った商 (整数) を $\text{div}(a, b)$, 余りを $\text{rem}(a, b)$ で表すと , i , j , k の間には

$$i = \text{div}(j - 1, D) + 1 \quad (\text{S.4.43})$$

$$k = \text{rem}(j - 1, D) + 1 \quad (\text{S.4.44})$$

が成り立つ . さらに次式に示すように , 第 j 成分が $\beta_j(\boldsymbol{x})$, $\beta_j(\boldsymbol{y})$ となる D^2 次元ベクトルを , それぞれ $\beta(\boldsymbol{x})$, $\beta(\boldsymbol{y})$ として定義する .

$$\begin{cases} \beta(\boldsymbol{x}) = (\beta_1(\boldsymbol{x}), \dots, \beta_j(\boldsymbol{x}), \dots, \beta_{D^2}(\boldsymbol{x}))^t \\ \beta(\boldsymbol{y}) = (\beta_1(\boldsymbol{y}), \dots, \beta_j(\boldsymbol{y}), \dots, \beta_{D^2}(\boldsymbol{y}))^t \end{cases} \quad (\text{S.4.45})$$

これらにより , 式 (S.4.41) は

$$F_{d+1}(\boldsymbol{x}, \boldsymbol{y}) = \lim_{D \rightarrow \infty} \sum_{j=1}^{D^2} \beta_j(\boldsymbol{x}) \beta_j(\boldsymbol{y}) \quad (\text{S.4.46})$$

$$= \lim_{D \rightarrow \infty} \beta(\boldsymbol{x})^t \beta(\boldsymbol{y}) \quad (\text{S.4.47})$$

表 S.4.1: 添字 i, j, k の関係

j	i	k
1	1	1
$\vdots$	$\vdots$	$\vdots$
D	1	D
$D + 1$	2	1
$\vdots$	$\vdots$	$\vdots$
$2D$	2	D
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$
$D \cdot (D - 1)$	D	1
$\vdots$	$\vdots$	$\vdots$
D^2	D	D

と書ける．ここで $D \rightarrow \infty$ なら $D^2 \rightarrow \infty$ となり， $\beta(x)$ ， $\beta(y)$ も無限次元ベクトルとなる．したがって，式 (S.4.38) が成り立つと仮定すると，式 (S.4.47) が成り立つことがわかる．すなわち， $F_d(x, y)$ が無限次元ベクトルの内積で表されるなら， $F_{d+1}(x, y)$ も無限次元ベクトルの内積で表される．

以上，数学的帰納法より，すべての d に対し， $F_d(x, y)$ は無限次元ベクトルの内積で表されることが証明された．

式 (S.4.3)，(S.4.26)，(S.4.38) より

$$K(x, y) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|y\|^2}{2\sigma^2}\right) F_d(x, y) \quad (\text{S.4.48})$$

$$= \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|y\|^2}{2\sigma^2}\right) \lim_{D \rightarrow \infty} \alpha(x)^t \alpha(y) \quad (\text{S.4.49})$$

が得られる．上式において，ベクトル $\phi(x)$ ， $\phi(y)$ を

$$\begin{cases} \phi(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \alpha(x) \\ \phi(y) = \exp\left(-\frac{\|y\|^2}{2\sigma^2}\right) \alpha(y) \end{cases} \quad (\text{S.4.50})$$

と定義すると，式 (S.4.49) は

$$K(x, y) = \lim_{D \rightarrow \infty} \phi(x)^t \phi(y) \quad (\text{S.4.51})$$

と書ける．ここで $D \rightarrow \infty$ とすると， $\phi(x)$ ， $\phi(y)$ も無限次元のベクトルとなるので， $K(x, y)$ が無限次元ベクトルの内積で表されることが証明できた．

【演習問題 4.2】

問 (1)

ポテンシャル関数 $K(x, x_k)$ としてガウス関数を用い，ポテンシャル関数法の学習アルゴリズム (4.3 節) を適用する．誤り計数ベクトル α の更新過程を求めると，表 S.4.2 のようになる．本表からわかるように，学習はエポック 5 で収束し，最終的に得られる誤り計数ベクトル α は，下式のようになる．

$$\alpha = (\alpha_1, \dots, \alpha_6)^t = (3, 2, 1, 1, 4, 1)^t \quad (\text{S.4.52})$$

教師信号は，すでに式 (S.3.35) で示したとおりであり，式 (S.4.52) とともに式 (4.9) に代入すると，識別関数 $g(x)$ は

$$g(x) = \sum_{k=1}^6 \alpha_k b_k K(x, x_k) \quad (\text{S.4.53})$$

$$\begin{aligned} &= 3 \cdot K(x, x_1) + 2 \cdot K(x, x_2) + 1 \cdot K(x, x_3) \\ &\quad - 1 \cdot K(x, x_4) - 4 \cdot K(x, x_5) - 1 \cdot K(x, x_6) \end{aligned} \quad (\text{S.4.54})$$

となる．決定境界は，上式において $g(x) = 0$ とすることにより求めることができ，図 S.4.1 に太線で示した．なお，式 (S.4.54) の $K(x, x_k)$ は，式 (4.7) において $\sigma = 2$ とした下式である．

$$K(x, x_k) = \exp\left[-\frac{\|x - x_k\|^2}{8}\right] \quad (k = 1, \dots, 6) \quad (\text{S.4.55})$$

表 S.4.2: ポテンシャル関数法における誤り計数ベクトル α の更新過程 (ガウス)

エポック	a_1	a_2	a_3	a_4	a_5	a_6
0	0	0	0	0	0	0
1	1	0	0	1	1	0
2	1	1	0	1	2	1
3	2	2	0	1	3	1
4	3	2	1	1	4	1
5	3	2	1	1	4	1

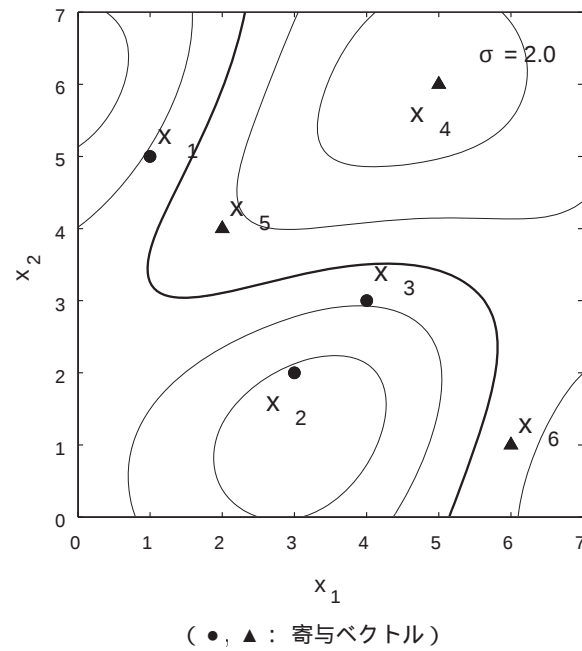


図 S.4.1: ポテンシャル関数法による決定境界 (ガウス)

図で，決定境界の左下側では $g(x) > 0$ となってクラス ω_1 ，右上側では $g(x) < 0$ となってクラス ω_2 と判定される．図より，両クラスのパターンが正しく識別されていることが確認できる．図の細線は $g(x)$ の等高線を示している．

式 (S.4.52) より， $\alpha_1, \dots, \alpha_6 \neq 0$ であるので，全パターンが寄与ベクトルとなり，そのことを示すため図ではそれらを黒く塗り潰している．

問 (2)

ポテンシャル関数として，式 (4.24) の 3 次多項式を用いた場合，誤り計数ベクトル α の更新過程は，表 S.4.3 のようになる．ただし，表では最初の 5 ポックと，収束に至る最後の 5 エポックのみを示した．本表より，学習はエポック 36 で収束し，最終的に得られる誤り計数ベクトル α は，下式のことになること

表 S.4.3: ポテンシャル関数法における誤り計数ベクトル α の更新過程 (多項式)

エポック	a_1	a_2	a_3	a_4	a_5	a_6
0	0	0	0	0	0	0
1	1	0	0	1	0	0
2	2	1	1	1	0	0
3	2	2	2	2	0	0
4	3	3	3	2	0	0
5	4	4	4	3	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
32	15	25	30	9	14	8
33	16	26	31	10	14	8
34	17	27	32	10	14	8
35	17	28	33	10	15	8
36	17	28	33	10	15	8

がわかる .

$$\alpha = (\alpha_1, \dots, \alpha_6)^t = (17, 28, 33, 10, 15, 8)^t \quad (\text{S.4.56})$$

収束に至る過程を図 S.4.2 に示す .

式 (S.3.35) と式 (S.4.56) を式 (4.9) に代入すると , 識別関数 $g(x)$ は

$$g(x) = \sum_{k=1}^6 \alpha_k b_k K(x, x_k) \quad (\text{S.4.57})$$

$$= 17 \cdot K(x, x_1) + 28 \cdot K(x, x_2) + 33 \cdot K(x, x_3) \\ - 10 \cdot K(x, x_4) - 15 \cdot K(x, x_5) - 8 \cdot K(x, x_6) \quad (\text{S.4.58})$$

となる . ただし , 上式の $K(x, x_k)$ は , 式 (4.24) で示したように

$$K(x, x_k) = (1 + x_k^t x)^3 \quad (k = 1, \dots, 6) \quad (\text{S.4.59})$$

である . 式 (S.4.59) を式 (S.4.58) に代入することにより , 識別関数 $g(x)$ として

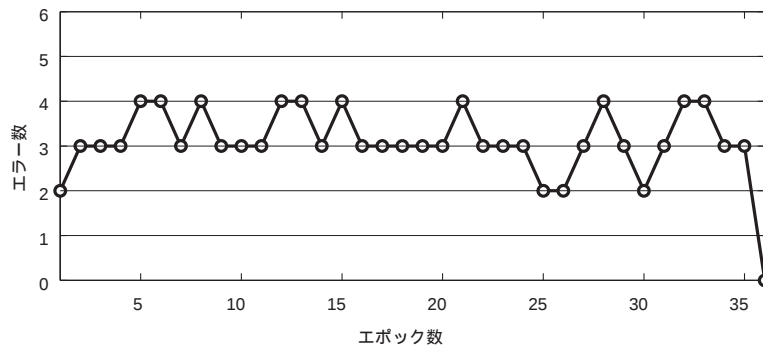


図 S.4.2: ポテンシャル関数法の学習過程 (多項式)

下式が得られる^{*2} .

$$g(x) = 45 + 315x_1 + 336x_2 + 1086x_1x_2 + 597x_1^2 + 678x_2^2 + 435x_1^2x_2 - 1137x_1x_2^2 - 213x_1^3 + 112x_2^3 \quad (\text{S.4.60})$$

決定境界は, 式 (S.4.60) において $g(x) = 0$ とすることにより求めることができ, 図 S.4.3 に太線で示した. 図の見方は図 S.4.1 と同じである. 図より, 両ク

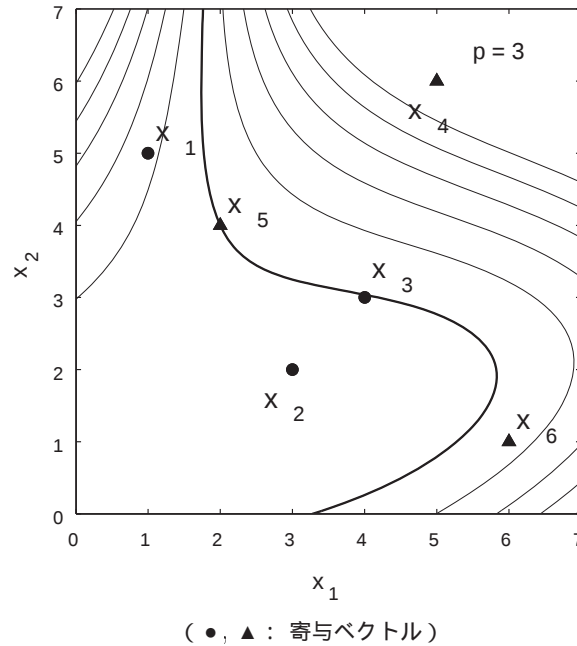


図 S.4.3: ポテンシャル関数法による決定境界 (多項式)

^{*2} 識別関数 $g(x)$ を, 式 (S.4.58) のように $K(x, x_k)$ を用いて表せることがカーネルトリックの魅力である. 式 (S.4.60) は, 式 (S.3.33) と比較するため, $g(x)$ をあえて x_1, x_2 を用いて陽に表したに過ぎない. 実際の $g(x)$ の計算には式 (S.4.60) ではなく, 式 (S.4.58), (S.4.59) を用いる.

ラスのパターンが正しく識別されていることが確認できる^{*3}。式 (S.4.56) より、 $\alpha_1, \dots, \alpha_6 \neq 0$ であるので、ここでも全パターンが寄与ベクトルとなり、それらを黒く塗り潰している。

次に、演習問題 3.3 で得られた結果と比較してみよう。演習問題 3.3 で得られた識別関数は、式 (S.3.33) であり、式 (S.4.60) と同様、3 次識別関数である。前者は $\phi(x)$ として式 (3.33) を用いている。一方、後者はポテンシャル関数法の双対表現であるパーセプトロンに置き換えると、 $\phi(x)$ として式 (4.20) を用いることに相当する。識別関数の各項の係数が異なっているのは、上記の理由による。図 S.4.3 と図 S.3.2 では大きな差は見られない。

^{*3} 図では、 x_3, x_5 が決定境界上にあるように見えるが、余裕が確保できていないだけで、実際は両パターンとも正しく識別できている。

第 5 章

【演習問題 5.1】

式 (5.6) より, 超平面 H_1 と H_2 を表す式は

$$g(x) = b_k = \begin{cases} 1 & (x \in \omega_1) \\ -1 & (x \in \omega_2) \end{cases} \quad (\text{S.5.1})$$

となる. ここで

$$g_0(x) \stackrel{\text{def}}{=} g(x) - b_k \quad (\text{S.5.2})$$

と定義すると, 超平面 H_1, H_2 の式は

$$g_0(x) = 0 \quad (\text{S.5.3})$$

と表せる. したがって, 式 (1.40) より, パターン x_k と, 超平面 H_1 あるいは H_2 との距離 r_k は

$$r_k = \frac{|g_0(x_k)|}{\|w\|} = \frac{|g(x_k) - b_k|}{\|w\|} \quad (\text{S.5.4})$$

となる.

式 (5.57), または式 (5.58) に該当する x_k に対しては $1 - b_k g(x_k) > 0$ であるので, 式 (5.61) より

$$\xi_k = 1 - b_k g(x_k) > 0 \quad (k = 1, \dots, n) \quad (\text{S.5.5})$$

となる. 上式の両辺に b_k を乗じ, $b_k^2 = 1$ であることを用いると

$$\begin{aligned} b_k \cdot \xi_k &= b_k - b_k^2 \cdot g(x_k) \\ &= b_k - g(x_k) \end{aligned} \quad (\text{S.5.6})$$

が得られる. 上式を式 (S.5.4) に代入することにより下式が得られる.

$$r_k = \frac{|-b_k \cdot \xi_k|}{\|w\|} = \frac{\xi_k}{\|w\|} \quad (\text{S.5.7})$$

よって, 式 (5.62) が導出できた.

【演習問題 5.2】

与えられた学習パターンは線形分離可能であるので，5.3 節 [1] の手順に従えばよい．式 (5.36) より行列 $\mathbf{H} = (h_{ij})$ を求めると，下式が得られる．

$$\mathbf{H} = \begin{pmatrix} 29 & 21 & 12 & 9 & -61 & -58 & -49 & -48 \\ 21 & 18 & 9 & 9 & -51 & -51 & -42 & -45 \\ 12 & 9 & 5 & 4 & -26 & -25 & -21 & -21 \\ 9 & 9 & 4 & 5 & -25 & -26 & -21 & -24 \\ -61 & -51 & -26 & -25 & 145 & 144 & 119 & 126 \\ -58 & -51 & -25 & -26 & 144 & 145 & 119 & 129 \\ -49 & -42 & -21 & -21 & 119 & 119 & 98 & 105 \\ -48 & -45 & -21 & -24 & 126 & 129 & 105 & 117 \end{pmatrix} \quad (\text{S.5.8})$$

上式を式 (5.37) に代入し，最適化問題 5.2 の双対問題を解くと，

$$\begin{aligned} \lambda^* &= (\lambda_1^*, \dots, \lambda_8^*) \\ &= (0.0556, 0.0139, 0, 0, 0, 0, 0.0694, 0) \end{aligned} \quad (\text{S.5.9})$$

が得られ，8 パターンのうち x_1, x_2, x_7 の 3 パターン ($m = 3$) がサポートベクトルであることがわかる．上記結果を用いると，式 (5.39) より

$$\begin{aligned} \mathbf{w}^* &= (w_1^*, w_2^*)^t = \sum_{k=1}^8 \lambda_k^* b_k \mathbf{x}_k \\ &= \lambda_1^* b_1 \mathbf{x}_1 + \lambda_2^* b_2 \mathbf{x}_2 + \lambda_7^* b_7 \mathbf{x}_7 \\ &= 0.0556 \cdot (2, 5)^t + 0.0139 \cdot (3, 3)^t - 0.0694 \cdot (7, 7)^t \\ &= (-0.3333, -0.1667)^t \end{aligned} \quad (\text{S.5.10})$$

が得られる．また w_0^* は，式 (5.44) の x_s としてサポートベクトル x_1, x_2, x_7 を用いてそれぞれ計算し，それらの平均値を算出する．その結果， w_0^* は

$$w_0^* = 2.5000 \quad (\text{S.5.11})$$

となる．本計算法については，第 5 章の脚注*6 を参照のこと．

式 (S.5.10)，(S.5.11) より，求めるべき識別関数 $g(x)$ として

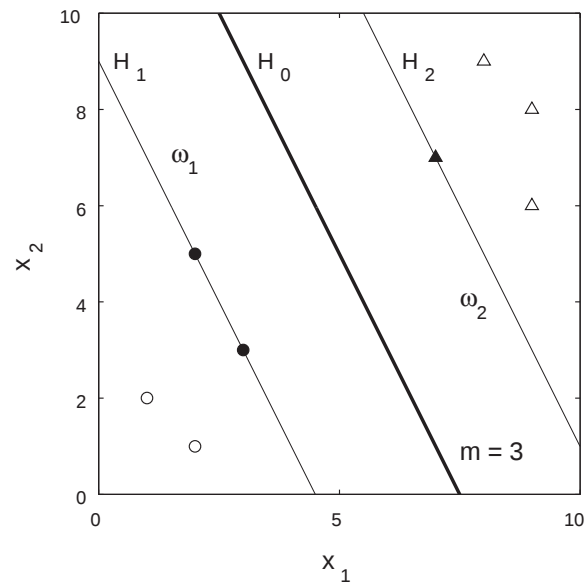
$$\begin{aligned} g(x) &= w_0^* + w_1^* x_1 + w_2^* x_2 \\ &= 2.5000 - 0.3333 x_1 - 0.1667 x_2 \end{aligned} \quad (\text{S.5.12})$$

が得られる．マージン R は，式 (5.8) より

$$R = \frac{2}{\|w^*\|} = 5.3666 \quad (\text{S.5.13})$$

である．決定境界 H_0 は，式 (S.5.12) で $g(x) = 0$ とすることにより求められる．また，マージンを決定する超平面のうち， H_1 は $g(x) = 1$ ， H_2 は $g(x) = -1$ とすることによって求められる．図 S.5.1 にこれらの結果を示した．図で，サポートベクトルは黒く塗りつぶした．

この結果を演習問題 1.3，演習問題 1.4 の結果と比較すると，マージン R が 1.240，2.478 からその最大値 5.367 へと大幅に改善されていることがわかる．



(●, ▲ : サポートベクトル)

図 S.5.1: 線形 SVM による決定境界 (線形分離可能な場合)

【演習問題 5.3】

前問と異なり，学習パターンは線形分離不可能であるので，線形 SVM を適用するには 5.3 節 [2] の手順に従い，最適化問題 5.4 の双対問題を解くことになる．そこでは，ペナルティ c_{pe} を導入する．

まず， $c_{pe} = 0.1$ の場合を示そう．式 (5.89) より行列 $\mathbf{H} = (h_{ij})$ を求めると，下式が得られる．

$$\mathbf{H} = \begin{pmatrix} 26 & 13 & 19 & -35 & -22 & -11 \\ 13 & 13 & 18 & -27 & -14 & -20 \\ 19 & 18 & 25 & -38 & -20 & -27 \\ -35 & -27 & -38 & 61 & 34 & 36 \\ -22 & -14 & -20 & 34 & 20 & 16 \\ -11 & -20 & -27 & 36 & 16 & 37 \end{pmatrix} \quad (\text{S.5.14})$$

上式と， $c_{pe} = 0.1$ の条件を用いて凸二次計画問題を解くと

$$\begin{aligned} \lambda^* &= (\lambda_1^*, \dots, \lambda_6^*) \\ &= (0.0941, 0.1000, 0.1000, 0.0943, 0.1000, 0.0998) \end{aligned} \quad (\text{S.5.15})$$

が得られ，全パターンがサポートベクトル ($m = 6$) であることがわかる．この結果を用いると，前問と同じ要領で

$$\mathbf{w}^* = (w_1^*, w_2^*)^t = \sum_{k=1}^6 \lambda_k^* b_k \mathbf{x}_k = (-0.4762, -0.0952)^t \quad (\text{S.5.16})$$

が得られる．また w_0^* は，式 (5.102) を満たす λ_k^* に対応する $\mathbf{x}_k$ を $\mathbf{x}_s$ と置き，式 (5.106) により計算できる．この条件に該当する λ_k^* は， $\lambda_1^*, \lambda_4^*, \lambda_6^*$ であるので， $\mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_6$ の 3 パターンを $\mathbf{x}_s$ としてそれぞれ式 (5.106) に代入し，得られた値を平均すると下式が得られる．

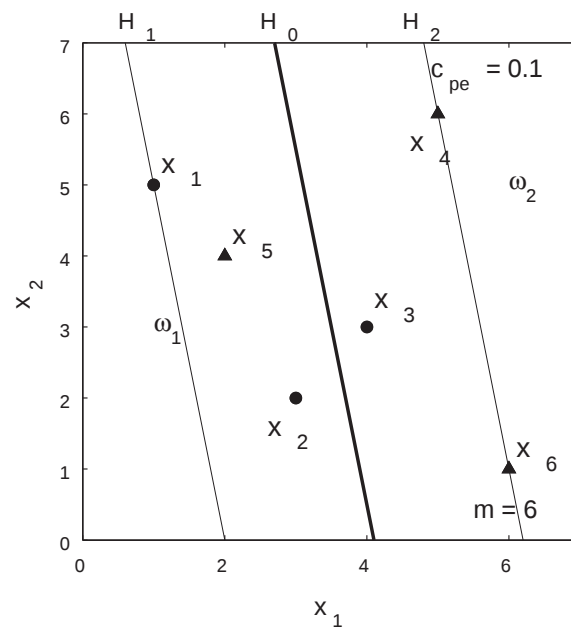
$$w_0^* = 1.9524 \quad (\text{S.5.17})$$

式 (S.5.16)，(S.5.17) より，求めるべき識別関数 $g(\mathbf{x})$ は

$$\begin{aligned} g(\mathbf{x}) &= w_0^* + w_1^* x_1 + w_2^* x_2 \\ &= 1.9524 - 0.4762 x_1 - 0.0952 x_2 \end{aligned} \quad (\text{S.5.18})$$

となる．上式で， $g(x) = 0, 1, -1$ と設定することにより，それぞれ超平面 H_0 ， H_1 ， H_2 が得られる．その結果を図 S.5.2 に示す．超平面 H_1 ， H_2 間の距離は 4.1184 であった．

以上の凸二次計画問題を解くと，全パターンがサポートベクトルとなり，それらを黒く塗り潰した．サポートベクトルのうち，式 (5.102) を満たすのは， x_1 ， x_4 ， x_6 の 3 パターンであり，これらは超平面 H_1 または H_2 上にある．式 (5.103) を満たすのは， x_2 の 1 パターンのみで，これは超平面 H_0 と H_1 の間の自クラス領域にある．以上のパターンは正しく識別されるが， x_3 ， x_5 の 2 パターンは，式 (5.104) を満たし，超平面 H_0 を越えて他クラスの領域に存在するので，誤識別される．



(●, ▲ : サポートベクトル)

図 S.5.2: 線形 SVM による決定境界 (線形分離不可能な場合) ($c_{pe} = 0.1$)

次に, $c_{pe} = 10$ の場合を示す. 行列 H は式 (S.5.14) と同じで, 式 (5.88) において $c_{pe} = 10$ として凸二次計画問題を解くと

$$\begin{aligned}\lambda^* &= (\lambda_1^*, \dots, \lambda_6^*) \\ &= (1.0612, 5.1837, 10.000, 0, 10.000, 6.2499)\end{aligned}\quad (\text{S.5.19})$$

となる. したがって, パターン x_4 以外はサポートベクトルである ($m = 5$). この結果より

$$\begin{aligned}w^* &= (w_1^*, w_2^*)^t = \sum_{k=1}^6 \lambda_k^* b_k x_k \\ &= (-0.8571, -0.5714)^t\end{aligned}\quad (\text{S.5.20})$$

が得られる. また, $0 < \lambda_k^* < c_{pe}$ に対応する x_k は, x_1, x_2, x_6 の 3 パターンであるので, これらを用いて式 (5.106) より得られた値を平均すると

$$w_0^* = 4.7143 \quad (\text{S.5.21})$$

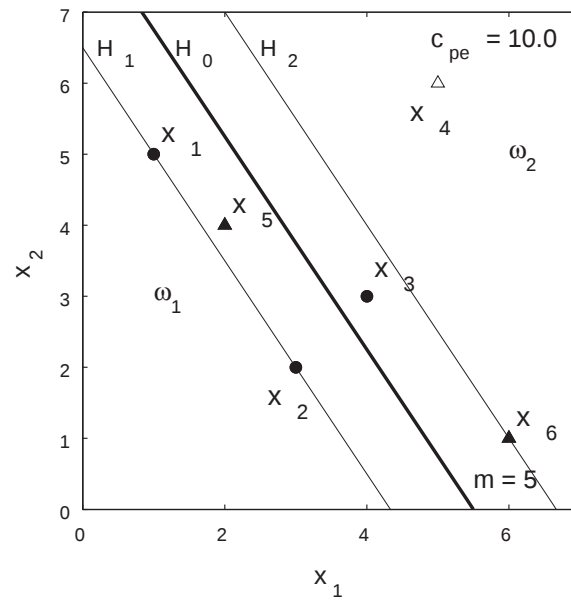
が得られる.

式 (S.5.20), (S.5.21) より, 求めるべき識別関数 $g(x)$ は

$$\begin{aligned}g(x) &= w_0^* + w_1^* x_1 + w_2^* x_2 \\ &= 4.7143 - 0.8571 x_1 - 0.5714 x_2\end{aligned}\quad (\text{S.5.22})$$

となる. 上式で, $g(x) = 0, 1, -1$ と設定することにより, それぞれ超平面 H_0, H_1, H_2 が得られる. その結果を図 S.5.3 に示す. 超平面 H_1, H_2 間の距離は 1.9415 であり, $c_{pe} = 0.1$ の場合に比べて減少していることが確かめられる. 式 (5.101) を満たすのは, 非サポートベクトルの x_4 のみであり, 図で x_4 以外を黒く塗り潰している.

サポートベクトルのうち, 式 (5.102) を満たすのは, x_1, x_2, x_6 の 3 パターンであり, これらは超平面 H_1 または H_2 上にある. 式 (5.103) を満たすパターンは無い. 式 (5.104) を満たす x_3, x_5 の 2 パターンは, 超平面 H_0 を越えて他クラスの領域に存在するので, 誤識別される.



(●, ▲ : サポートベクトル)

図 S.5.3: 線形 SVM による決定境界 (線形分離不可能な場合) ($c_{pe} = 10$)

【演習問題 5.4】

問 (1)

カーネル関数 $K(x_i, x_j)$ として式 (4.7) のガウス関数を用い, $\sigma = 2.0$ とすると

$$K(x_i, x_j) = \exp \left[-\frac{\|x_i - x_j\|^2}{8} \right] \quad (i, j = 1, \dots, 6) \quad (\text{S.5.23})$$

である. 上式を用いて式 (5.116) より行列 $\mathbf{H} = (h_{ij})$ を求めると, 下式が得ら

れる．

$$\mathbf{H} = \begin{pmatrix} 1.000 & 0.197 & 0.197 & -0.119 & -0.779 & -0.006 \\ 0.197 & 1.000 & 0.779 & -0.082 & -0.535 & -0.287 \\ 0.197 & 0.779 & 1.000 & -0.287 & -0.535 & -0.368 \\ -0.119 & -0.082 & -0.287 & 1.000 & 0.197 & 0.039 \\ -0.779 & -0.535 & -0.535 & 0.197 & 1.000 & 0.044 \\ -0.006 & -0.287 & -0.368 & 0.039 & 0.044 & 1.000 \end{pmatrix} \quad (\text{S.5.24})$$

上式を用いて最適化問題 5.2 の双対問題を解くことにより，

$$\begin{aligned} \lambda^* &= (\lambda_1^*, \dots, \lambda_6^*) \\ &= (8.13, 2.38, 4.78, 1.38, 10.85, 3.05) \end{aligned} \quad (\text{S.5.25})$$

が得られる．これより全パターンがサポートベクトルであることがわかる．したがって，全パターンについて式 (5.119) を計算し，それらを平均することにより， w_0^* は下式となる．

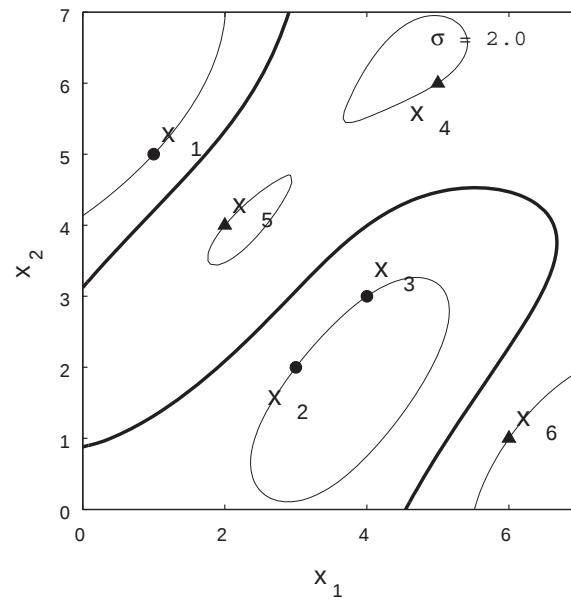
$$w_0^* = 0.0975 \quad (\text{S.5.26})$$

教師信号は，すでに式 (S.3.35) で示したとおりであり，式 (S.5.25)，(S.5.26) を式 (5.118) に代入すると，識別関数 $g(x)$ は

$$\begin{aligned} g(x) &= w_0^* + \sum_{k=1}^6 \lambda_k^* b_k K(x, x_k) \\ &= 0.0975 + 8.13 \cdot K(x, x_1) + 2.38 \cdot K(x, x_2) + 4.78 \cdot K(x, x_3) \\ &\quad - 1.38 \cdot K(x, x_4) - 10.85 \cdot K(x, x_5) - 3.05 \cdot K(x, x_6) \end{aligned} \quad (\text{S.5.27})$$

となる．式 (S.5.27) で $g(x) = 0$ とすることにより決定境界が求められ，図 S.5.4 の太線で示した．また，細線は $g(x) = \pm 1$ に対応しており，サポートベクトルである全パターンがこの細線上にあることが確認できる．

以上，得られた結果を演習問題 4.2(1) の結果と比較すると，図 S.4.1 と図 S.5.4 とで決定境界の設定には違いが見られるものの，クラス間分離の余裕という点では大きな差は認められない．サポートベクトルマシンの手法を採らなくても，図 S.4.1 ではたまたま余裕のある決定境界が得られたことを示している．



(●, ▲ : サポートベクトル)

図 S.5.4: ガウスカーネルを用いた非線形 SVM による決定境界

問 (2)

カーネル関数 $K(x_i, x_j)$ として式 (4.24) の 3 次の多項式を用いると

$$K(x_i, x_j) = (1 + x_i^t x_j)^3 \quad (\text{S.5.28})$$

であるので, 式 (5.116) より行列 $\mathbf{H} = (h_{ij})$ は

$$\mathbf{H} = \begin{pmatrix} 19683 & 2744 & 8000 & -46656 & -12167 & -1728 \\ 2744 & 2744 & 6859 & -21952 & -3375 & -9261 \\ 8000 & 6859 & 17576 & -59319 & -9261 & -21952 \\ -46656 & -21952 & -59319 & 238328 & 42875 & 50653 \\ -12167 & -3375 & -9261 & 42875 & 9261 & 4913 \\ -1728 & -9261 & -21952 & 50653 & 4913 & 54872 \end{pmatrix} \quad (\text{S.5.29})$$

となる．問 (1) と同様にして双対問題を解くと，

$$\begin{aligned}\lambda^* &= (\lambda_1^*, \dots, \lambda_6^*) \\ &= (6.122, 4.596, 4.902, 0, 13.90, 1.718) \times 10^{-3} \quad (\text{S.5.30})\end{aligned}$$

が得られる．上式より， $\lambda_4^* = 0$ であるので， x_4 のみ非サポートベクトルで，残り 5 パターンはサポートベクトルであることがわかる．この結果を用いて，式 (5.119) より

$$w_0^* = 0.7922 \quad (\text{S.5.31})$$

と求めることができる．

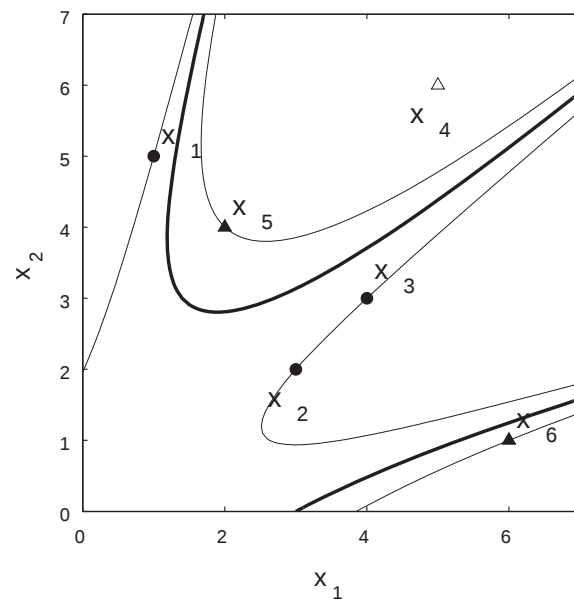
式 (S.3.35)，式 (S.5.30)，(S.5.31) を式 (5.118) に代入すると，識別関数 $g(x)$ は下式となる（式 (S.5.33) については，演習問題解答 35 ページの脚注*2 参照）．

$$\begin{aligned}g(x) &= w_0^* + \sum_{k=1}^6 \lambda_k^* b_k K(x, x_k) \\ &= w_0^* + \sum_{k=1}^6 \lambda_k^* b_k (1 + x_k^t x)^3 \quad (\text{S.5.32})\end{aligned}$$

$$\begin{aligned}&= 0.7922 + 0.0042 \cdot x_1 - 0.0085 \cdot x_2 - 0.0271 \cdot x_1 x_2 \\ &\quad + 0.0254 \cdot x_1^2 - 0.0258 \cdot x_2^2 + 0.1932 \cdot x_1^2 x_2 - 0.2115 \cdot x_1 x_2^2 \\ &\quad - 0.0383 \cdot x_1^3 + 0.0429 \cdot x_2^3 \quad (\text{S.5.33})\end{aligned}$$

上式で $g(x) = 0$ とした決定境界を図 S.5.5 中の太線で示した．また， $g(x) = \pm 1$ に相当する境界を細線で示した．図で， x_4 以外の 5 パターンをサポートベクトルとして黒く塗り潰しており，それらは細線上にあることが確認できる．

以上，得られた結果を演習問題 4.2(2) の結果と比較する．図 S.4.3 と図 S.5.5 とで決定境界を比較すると，前者は演習問題解答 36 ページの脚注*2 でも指摘したように，パターンと決定境界が近接しており，余裕が無い．それに対して後者は十分な余裕が確保できており，SVM によるマージン最大化の効果が発揮されている．



(●, ▲ : サポートベクトル)

図 S.5.5: 多項式カーネルを用いた非線形 SVM による決定境界

第 6 章

【演習問題 6.1】

関数 $k(x, y)$ がカーネル関数であれば、行列 $\mathbf{K}$ の (i, j) 成分である $k(x_i, x_j)$ は

$$k(x_i, x_j) = \phi(x_i)^t \phi(x_j) \quad (i, j = 1, \dots, n) \quad (\text{S.6.1})$$

と書ける。上式で、 $\phi(x_i)$ は、 d 次元列ベクトル x_i を変換して得られる D 次元列ベクトルであり、 $\phi(x_j)$ も同様にして得られる。式 (S.6.1) より、 $k(x_i, x_j) = k(x_j, x_i)$ が成り立つので、 $\mathbf{K}$ は対称行列である。

ここで n 個の列ベクトル $\phi(x_1), \dots, \phi(x_n)$ を用いて次式のような $n \times D$ の行列 $\mathbf{X}$ を作る。

$$\mathbf{X} = (\phi(x_1), \dots, \phi(x_n))^t \quad (\text{S.6.2})$$

次式が成り立つことは明らかである。

$$\mathbf{K} = \mathbf{X} \mathbf{X}^t \quad (\text{S.6.3})$$

ここで、 $y \neq 0$ なる任意の n 次元列ベクトル y に対し、 $y^t \mathbf{K} y$ を計算すると、式 (S.6.3) より

$$\begin{aligned} y^t \mathbf{K} y &= y^t \mathbf{X} \mathbf{X}^t y \\ &= (\mathbf{X}^t y)^t \mathbf{X}^t y \\ &= \|\mathbf{X}^t y\|^2 \geq 0 \end{aligned} \quad (\text{S.6.4})$$

が成り立つ。以上より、 $\mathbf{K}$ は対称行列で、かつ式 (S.6.4) を満たすので、行列 $\mathbf{K}$ は半正定値である（付録 A.3 の式 (A.3.8) も参照のこと）。

【演習問題 6.2】

行列 $\mathbf{K}$ が半正定値なら、 $\mathbf{K}$ は対称行列であるので、付録 A.3 の定理 A.1 より、 $\mathbf{K}$ の n 個の固有ベクトルは正規直交基底を成す。行列 $\mathbf{K}$ の固有値を λ_i 、それに対応する固有ベクトルを $\mathbf{u}_i$ とすると

$$\mathbf{K} \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (i = 1, \dots, n) \quad (\text{S.6.5})$$

である．ここで， $\mathbf{K}$ の n 個の固有列ベクトル $\mathbf{u}_i$ と固有値 λ_i を用いて下式に示す行列 $\mathbf{U}$ と $\mathbf{\Lambda}$ を定義する．

$$\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n) \quad (\text{S.6.6})$$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{pmatrix} \quad (\text{S.6.7})$$

行列 $\mathbf{\Lambda}$ は，固有値 $\lambda_1, \dots, \lambda_n$ を対角成分として持つ対角行列である．式 (S.6.6)，(S.6.7) を用いると，式 (S.6.5) は

$$\mathbf{K}\mathbf{U} = \mathbf{U}\mathbf{\Lambda} \quad (\text{S.6.8})$$

と書ける．固有ベクトル $\mathbf{u}_i$ は正規直交基底を成すので下式が成り立つ．

$$\mathbf{U}^t \mathbf{U} = \mathbf{U} \mathbf{U}^t = \mathbf{I} \quad (\text{S.6.9})$$

ただし， $\mathbf{I}$ は $n \times n$ の単位行列である．

行列 $\mathbf{K}$ が半正定値であるので，付録 A.3 の定理 A.2 より

$$\lambda_i \geq 0 \quad (i = 1, \dots, n) \quad (\text{S.6.10})$$

である．そこで

$$\mathbf{\Lambda}^{1/2} \stackrel{\text{def}}{=} \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \sqrt{\lambda_2} & \\ 0 & & \ddots \\ & & & \sqrt{\lambda_n} \end{pmatrix} \quad (\text{S.6.11})$$

を定義し，式 (S.6.8) の右から $\mathbf{U}^t$ を乗じて式 (S.6.9) を用いると

$$\mathbf{K} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{U}^t = \mathbf{U}\mathbf{\Lambda}^{1/2}(\mathbf{U}\mathbf{\Lambda}^{1/2})^t \quad (\text{S.6.12})$$

$$= \mathbf{Y}\mathbf{Y}^t \quad (\text{S.6.13})$$

を得る．ただし

$$\mathbf{Y} \stackrel{\text{def}}{=} \mathbf{U}\mathbf{\Lambda}^{1/2} \quad (\text{S.6.14})$$

である．ここで， d 次元の列ベクトル x_i を変換して得られる D 次元の列ベクトルを $\phi(x_i)$ とする ($i = 1, \dots, n$)．式 (S.6.13) の $\mathbf{Y}$ を

$$\mathbf{Y} = (\phi(x_1), \dots, \phi(x_n))^t \quad (\text{S.6.15})$$

と置き， $\mathbf{K}$ を求めると，以下のように表せる．

$$\begin{aligned} \mathbf{K} &= \mathbf{Y}\mathbf{Y}^t \\ &= \begin{pmatrix} \phi(x_1)^t\phi(x_1) & \phi(x_1)^t\phi(x_2) & \dots & \phi(x_1)^t\phi(x_n) \\ \phi(x_2)^t\phi(x_1) & \phi(x_2)^t\phi(x_2) & \dots & \phi(x_2)^t\phi(x_n) \\ \vdots & \vdots & \dots & \vdots \\ \phi(x_n)^t\phi(x_1) & \phi(x_n)^t\phi(x_2) & \dots & \phi(x_n)^t\phi(x_n) \end{pmatrix} \end{aligned} \quad (\text{S.6.16})$$

したがって $\mathbf{K}$ は，カーネル関数

$$k(x_i, x_j) = \phi(x_i)^t\phi(x_j) \quad (\text{S.6.17})$$

を用いて式 (6.156) のように書ける．

【演習問題 6.3】

以下に示すように，式 (6.76) の漸化式を表す三つの式に，それぞれ番号を付与して再掲する．

$$K_l^m = 1 \quad (l = 0 \text{ のとき}) \quad (\text{S.6.18})$$

$$K_l^m = 0 \quad (l > m \text{ のとき}) \quad (\text{S.6.19})$$

$$K_l^m = K_l^{m-1} + K_{l-1}^{m-1} x_m y_m \quad (\text{otherwise}) \quad (\text{S.6.20})$$

式 (6.68) で示されているのが， $d = 3$ ， $l = 2$ のときの $\phi(x)$ である．これより，式 (6.69) を用いて ANOVA カーネル $k_a(x, y)$ は

$$k_a(x, y) = x_1 y_1 x_2 y_2 + x_1 y_1 x_3 y_3 + x_2 y_2 x_3 y_3 \quad (\text{S.6.21})$$

であることがわかる．以下では，式 (S.6.18) ~ (S.6.20) の漸化式によって上式が導出できることを示す．

式 (S.6.18) において $(l, m) = (0, 0), (0, 1)$ とすることにより

$$K_0^0 = 1 \quad (\text{S.6.22})$$

$$K_0^1 = 1 \quad (\text{S.6.23})$$

が得られ, 式 (S.6.19) において $(l, m) = (1, 0), (2, 1)$ とすることにより

$$K_1^0 = 0 \quad (\text{S.6.24})$$

$$K_2^1 = 0 \quad (\text{S.6.25})$$

が得られる. 式 (S.6.20) において $(l, m) = (1, 1)$ とし, 式 (S.6.22), (S.6.24) を用いると, 下式が得られる.

$$\begin{aligned} K_1^1 &= K_1^0 + K_0^0 x_1 y_1 \\ &= x_1 y_1 \end{aligned} \quad (\text{S.6.26})$$

式 (S.6.20) において $(l, m) = (1, 2)$ とし, 式 (S.6.23), (S.6.26) を用いると, 下式が得られる.

$$\begin{aligned} K_1^2 &= K_1^1 + K_0^1 x_2 y_2 \\ &= x_1 y_1 + x_2 y_2 \end{aligned} \quad (\text{S.6.27})$$

式 (S.6.20) において $(l, m) = (2, 2)$ とし, 式 (S.6.25), (S.6.26) を用いると, 下式が得られる.

$$\begin{aligned} K_2^2 &= K_2^1 + K_1^1 x_2 y_2 \\ &= x_1 y_1 x_2 y_2 \end{aligned} \quad (\text{S.6.28})$$

式 (S.6.20) において $(l, m) = (2, 3)$ とし, 式 (S.6.27), (S.6.28) を用いると, 下式が得られる.

$$\begin{aligned} K_2^3 &= K_2^2 + K_1^2 x_3 y_3 \\ &= x_1 y_1 x_2 y_2 + (x_1 y_1 + x_2 y_2) x_3 y_3 \\ &= x_1 y_1 x_2 y_2 + x_1 y_1 x_3 y_3 + x_2 y_2 x_3 y_3 \end{aligned} \quad (\text{S.6.29})$$

式 (S.6.29) は式 (S.6.21) と一致し,

$$k_a(x, y) = K_2^3 \quad (\text{S.6.30})$$

となるので, $d = 3, l = 2$ のとき, 漸化式によって, ANOVA カーネル $k_a(x, y)$ が求められることが確認できた.

【演習問題 6.4】

最小化すべき J は, 以下のように書き換えられる.

$$\begin{aligned} J &= \|\mathbf{A}x - \mathbf{b}\|^2 + \lambda\|x\|^2 \\ &= (\mathbf{A}x - \mathbf{b})^t(\mathbf{A}x - \mathbf{b}) + \lambda x^t x \\ &= x^t \mathbf{A}^t \mathbf{A} x - 2\mathbf{b}^t \mathbf{A} x + \mathbf{b}^t \mathbf{b} + \lambda x^t x \end{aligned} \quad (\text{S.6.31})$$

上式を x で偏微分して 0 と置くと

$$\frac{\partial J}{\partial x} = 2\mathbf{A}^t \mathbf{A} x - 2\mathbf{A}^t \mathbf{b} + 2\lambda x = 0 \quad (\text{S.6.32})$$

となる. ベクトル, 行列の微分については, [改訂版] の付録 A.2 を参照されたい. 上式より x^* は

$$x^* = (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t \mathbf{b} \quad (\text{S.6.33})$$

と求まる. このとき

$$\begin{aligned} (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t &= (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I}) (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I})^{-1} \\ &= (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I}) \mathbf{A}^t (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I})^{-1} \\ &= \mathbf{A}^t (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I})^{-1} \end{aligned} \quad (\text{S.6.34})$$

となることから,

$$\begin{aligned} x^* &= (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t \mathbf{b} \\ &= \mathbf{A}^t (\mathbf{A} \mathbf{A}^t + \lambda \mathbf{I})^{-1} \mathbf{b} \end{aligned} \quad (\text{S.6.35})$$

である.

【演習問題 6.5】

式 (6.112) より下式が得られる.

$$\phi_M(\mathbf{z}) = \phi(\mathbf{z}) - \frac{1}{n} \sum_{i=1}^n \phi(x_i) \quad (\text{S.6.36})$$

あるパターン $\mathbf{z}$ に対し, $\phi(\mathbf{z})$ から $\frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i)$ までの距離を $L(\mathbf{z})$ とすると, $L(\mathbf{z})$ は式 (S.6.36) のベクトル $\phi_M(\mathbf{z})$ の長さに等しい. したがって下式が成り立つ.

$$L^2(\mathbf{z}) = \phi_M^t(\mathbf{z}) \cdot \phi_M(\mathbf{z}) \quad (\text{S.6.37})$$

式 (6.125) を再掲すると

$$\mathbf{G}_X(\mathbf{Y}, \mathbf{Z}) = \mathbf{Y}_{\phi_M} \cdot \mathbf{Z}_{\phi_M}^t \quad (\text{S.6.38})$$

であり, 上式の $\mathbf{Y}$ を $\mathbf{Z}$ と置くことにより下式が得られる.

$$\mathbf{G}_X(\mathbf{Z}, \mathbf{Z}) = \mathbf{Z}_{\phi_M} \cdot \mathbf{Z}_{\phi_M}^t \quad (\text{S.6.39})$$

一方, 定義より

$$\mathbf{Z}_{\phi_M} = (\phi_M(\mathbf{z}_1), \phi_M(\mathbf{z}_2), \dots, \phi_M(\mathbf{z}_m))^t \quad (\text{S.6.40})$$

であり, 上式の特別な場合として $m = 1$ とすると

$$\mathbf{Z}_{\phi_M} = \phi_M^t(\mathbf{z}) \quad (\text{S.6.41})$$

となる. 上式を式 (S.6.39) に代入すると

$$\mathbf{G}_X(\mathbf{z}, \mathbf{z}) = \phi_M^t(\mathbf{z}) \cdot \phi_M(\mathbf{z}) \quad (\text{S.6.42})$$

が得られ, 式 (S.6.37) より

$$L^2(\mathbf{z}) = \mathbf{G}_X(\mathbf{z}, \mathbf{z}) \quad (\text{S.6.43})$$

であることがわかる.

正規直交ベクトルの組 $\mathbf{V}_{d'} = (\mathbf{v}_1, \dots, \mathbf{v}_{d'})$ で張られる部分空間への $\phi_M(\mathbf{z})$ の射影ベクトルの長さを $P(\mathbf{z})$ とする. 式 (6.147) で示したように, CLAFIC 法の場合は $\phi(\mathbf{z})$ の射影ベクトルの長さを求めたが, 投影距離法では $\phi_M(\mathbf{z})$ の射影ベクトルの長さを求めることになる. 式 (6.147) と同様にして下式が得られる.

$$P^2(\mathbf{z}) = \sum_{i=1}^{d'} (\mathbf{v}_i^t \phi_M(\mathbf{z}))^2 = \|\mathbf{V}_{d'}^t \phi_M(\mathbf{z})\|^2 \quad (\text{S.6.44})$$

式 (S.6.38) の $\mathbf{Y}$ を $\mathbf{X}$ に置き換えることにより下式を得る .

$$\mathbf{G}_X(\mathbf{X}, \mathbf{Z}) = \mathbf{X}_{\phi_M} \cdot \mathbf{Z}_{\phi_M}^t \quad (\text{S.6.45})$$

パターン行列 $\mathbf{Z}$ は

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m)^t \quad (\text{S.6.46})$$

であるので , $m = 1$ の場合は

$$\mathbf{Z} = \mathbf{z}^t \quad (\text{S.6.47})$$

であり , $m = 1$ の場合の $\mathbf{Z}_{\phi_M}$ は式 (S.6.41) で表される . これらを式 (S.6.45) に代入すると

$$\mathbf{G}_X(\mathbf{X}, \mathbf{z}^t) = \mathbf{X}_{\phi_M} \cdot \phi_M(\mathbf{z}) \quad (\text{S.6.48})$$

となる . 式 (6.111) , (6.146) と同様にして

$$\mathbf{X}_{\phi_M} = \mathbf{U}_{d'} \Lambda_{d'}^{1/2} \mathbf{V}_{d'}^t \quad (\text{S.6.49})$$

が成り立つので , 上式を式 (S.6.48) に代入すると

$$\mathbf{G}_X(\mathbf{X}, \mathbf{z}^t) = \mathbf{U}_{d'} \Lambda_{d'}^{1/2} \mathbf{V}_{d'}^t \phi_M(\mathbf{z}) \quad (\text{S.6.50})$$

を得る . 両辺に左から $\Lambda_{d'}^{-1/2} \mathbf{U}_{d'}^t$ を乗じると

$$\Lambda_{d'}^{-1/2} \mathbf{U}_{d'}^t \mathbf{G}_X(\mathbf{X}, \mathbf{z}^t) = \mathbf{V}_{d'}^t \phi_M(\mathbf{z}) \quad (\text{S.6.51})$$

が得られ , この結果を式 (S.6.44) に代入すると

$$P^2(\mathbf{z}) = \|\Lambda_{d'}^{-1/2} \mathbf{U}_{d'}^t \mathbf{G}_X(\mathbf{X}, \mathbf{z}^t)\|^2 \quad (\text{S.6.52})$$

となる .

投影距離の二乗 $D^2(\mathbf{z})$ は $L^2(\mathbf{z})$ と $P^2(\mathbf{z})$ の差であるから , 式 (S.6.43) , 式 (S.6.52) より , 下式が得られる .

$$\begin{aligned} D^2(\mathbf{z}) &= L^2(\mathbf{z}) - P^2(\mathbf{z}) \\ &= G_X(\mathbf{z}, \mathbf{z}) - \|\Lambda_{d'}^{-1/2} \mathbf{U}_{d'}^t G_X(\mathbf{X}, \mathbf{z}^t)\|^2 \end{aligned} \quad (\text{S.6.53})$$

【演習問題 6.6】

式 (6.103) , (6.137) を以下に再掲する .

$$\mathbf{J}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_{nn} \quad (\in \mathbf{R}^{n \times n}) \quad (\text{S.6.54})$$

$$\mathbf{L} = -\frac{1}{2} \mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n \quad (\in \mathbf{R}^{n \times n}) \quad (\text{S.6.55})$$

以下 , 行列 $\mathbf{A}$ の (i, j) 成分を $(\mathbf{A})_{ij}$ と記すことにする . 式 (S.6.54) より , $\mathbf{J}_n$ の (i, j) 成分 $(\mathbf{J}_n)_{ij}$ は

$$(\mathbf{J}_n)_{ij} = \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_{nn} \right)_{ij} \quad (\text{S.6.56})$$

$$= (\mathbf{I}_n)_{ij} - \frac{1}{n} \quad (\text{S.6.57})$$

$$= \delta_{ij} - \frac{1}{n} \quad (\text{S.6.58})$$

となる . 上式の δ_{ij} は下式のように定義する .

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (\text{S.6.59})$$

ここで , $\mathbf{D}^{(2)}$ の (i, j) 成分 $(\mathbf{D}^{(2)})_{ij}$ は

$$(\mathbf{D}^{(2)})_{ij} = d_{ij}^2 \quad (\text{S.6.60})$$

であるので , $\mathbf{D}^{(2)} \mathbf{J}_n$ の (i, j) 成分 $(\mathbf{D}^{(2)} \mathbf{J}_n)_{ij}$ は

$$(\mathbf{D}^{(2)} \mathbf{J}_n)_{ij} = \sum_{l=1}^n (\mathbf{D}^{(2)})_{il} \cdot (\mathbf{J}_n)_{lj} \quad (\text{S.6.61})$$

$$= \sum_{l=1}^n d_{il}^2 \left(\delta_{lj} - \frac{1}{n} \right) \quad (\text{S.6.62})$$

$$= d_{ij}^2 - \frac{1}{n} \sum_{l=1}^n d_{il}^2 \quad (\text{S.6.63})$$

となる．同様にして $\mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n$ の (i, j) 成分 $(\mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n)_{ij}$ は

$$\begin{aligned} & (\mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n)_{ij} \\ &= \sum_{k=1}^n (\mathbf{J}_n)_{ik} \cdot (\mathbf{D}^{(2)} \mathbf{J}_n)_{kj} \end{aligned} \quad (\text{S.6.64})$$

$$= \sum_{k=1}^n \left(\delta_{ik} - \frac{1}{n} \right) \left(d_{kj}^2 - \frac{1}{n} \sum_{l=1}^n d_{kl}^2 \right) \quad (\text{S.6.65})$$

$$= \sum_{k=1}^n \delta_{ik} \left(d_{kj}^2 - \frac{1}{n} \sum_{l=1}^n d_{kl}^2 \right) - \frac{1}{n} \sum_{k=1}^n \left(d_{kj}^2 - \frac{1}{n} \sum_{l=1}^n d_{kl}^2 \right) \quad (\text{S.6.66})$$

$$= d_{ij}^2 - \frac{1}{n} \sum_{l=1}^n d_{il}^2 - \frac{1}{n} \sum_{k=1}^n d_{kj}^2 + \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n d_{kl}^2 \quad (\text{S.6.67})$$

となる．式 (S.6.67) 第 2 項の添字 l を k に替え，式 (S.6.55) を用いると

$$l_{ij} = -\frac{1}{2} (\mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n)_{ij} \quad (\text{S.6.68})$$

$$= -\frac{1}{2} \left(d_{ij}^2 - \frac{1}{n} \sum_{k=1}^n d_{ik}^2 - \frac{1}{n} \sum_{k=1}^n d_{kj}^2 + \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n d_{kl}^2 \right) \quad (\text{S.6.69})$$

となり，式 (6.138) が証明された．

【演習問題 6.7】

問 (1)

距離行列 $\mathbf{D}$ がユークリッド距離によって定義されているなら，行列 $\mathbf{D}^{(2)}$ の (i, j) 成分 d_{ij}^2 は

$$d_{ij}^2 = \|x_i - x_j\|^2 \quad (\text{S.6.70})$$

$$= x_i^t x_i - 2x_i^t x_j + x_j^t x_j \quad (\text{S.6.71})$$

である．対角行列 $\mathbf{A}$ の対角成分のみを抜き出し，他の成分を 0 とした行列を $\text{diag}(\mathbf{A})$ で表すことにすると

$$\mathbf{D}^{(2)} = \text{diag}(\mathbf{X}\mathbf{X}^t) \mathbf{1}_{nn} - 2\mathbf{X}\mathbf{X}^t + \mathbf{1}_{nn} \text{diag}(\mathbf{X}\mathbf{X}^t) \quad (\text{S.6.72})$$

と書ける．上式を式 (6.137) に代入することにより，下式を得る．

$$\mathbf{L} = -\frac{1}{2}\mathbf{J}_n\mathbf{D}^{(2)}\mathbf{J}_n \quad (\text{S.6.73})$$

$$\begin{aligned} &= -\frac{1}{2}\mathbf{J}_n (\text{diag}(\mathbf{X}\mathbf{X}^t) \mathbf{1}_{nn} - 2\mathbf{X}\mathbf{X}^t + \mathbf{1}_{nn} \text{diag}(\mathbf{X}\mathbf{X}^t)) \mathbf{J}_n \\ &= -\frac{1}{2}\mathbf{J}_n \text{diag}(\mathbf{X}\mathbf{X}^t) \mathbf{1}_{nn} \mathbf{J}_n + \mathbf{J}_n \mathbf{X}\mathbf{X}^t \mathbf{J}_n \\ &\quad - \frac{1}{2}\mathbf{J}_n \mathbf{1}_{nn} \text{diag}(\mathbf{X}\mathbf{X}^t) \mathbf{J}_n \end{aligned} \quad (\text{S.6.74})$$

ここで

$$\begin{aligned} \mathbf{1}_{nn} \mathbf{J}_n &= \mathbf{1}_{nn} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_{nn} \right) \\ &= \mathbf{1}_{nn} - \frac{1}{n} \mathbf{1}_{nn} \mathbf{1}_{nn} \\ &= \mathbf{1}_{nn} - \mathbf{1}_{nn} \\ &= \mathbf{0} \end{aligned} \quad (\text{S.6.75})$$

であり，同様に

$$\mathbf{J}_n \mathbf{1}_{nn} = \mathbf{0} \quad (\text{S.6.76})$$

が導けるので，式 (S.6.74) の第 1，第 3 項はいずれも 0 となる．その結果， $\mathbf{L}$ は式 (S.6.74) の第 2 項のみとなり，さらに式 (A.4.21) で示した $\mathbf{X}_M = \mathbf{J}_n \mathbf{X}$ の関係式を用いることにより

$$\mathbf{L} = \mathbf{J}_n \mathbf{X}\mathbf{X}^t \mathbf{J}_n \quad (\text{S.6.77})$$

$$= \mathbf{X}_M \mathbf{X}_M^t \quad (\text{S.6.78})$$

となる．以上より， $\mathbf{L} = \mathbf{X}_M \mathbf{X}_M^t$ が証明された．また，明らかに $\mathbf{L}^t = \mathbf{L}$ が成り立つので， $\mathbf{L}$ は対称行列である．

ここで， $\mathbf{a} \neq \mathbf{0}$ なる任意の n 次元列ベクトル $\mathbf{a}$ に対し

$$\begin{aligned} \mathbf{a}^t \mathbf{L} \mathbf{a} &= \mathbf{a}^t \mathbf{X}_M \mathbf{X}_M^t \mathbf{a} \\ &= (\mathbf{X}_M^t \mathbf{a})^t (\mathbf{X}_M^t \mathbf{a}) \\ &= \|\mathbf{X}_M^t \mathbf{a}\|^2 \geq 0 \end{aligned} \quad (\text{S.6.79})$$

となるので，式 (6.26) より $\mathbf{L}$ は半正定値の対称行列であることが証明された．

問 (2)

行列 $\mathbf{L}$ の固有値を λ_i ，それに対応する固有ベクトルを $\mathbf{u}_i$ とすると

$$\mathbf{L}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad (i = 1, \dots, n) \quad (\text{S.6.80})$$

が成り立つ．ただし， $\mathbf{u}_i$ は下式のように正規化されている．

$$\mathbf{u}_i^t \mathbf{u}_j = \delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (\text{S.6.81})$$

行列 $\mathbf{L}$ が半正定値の対称行列なら，付録 A.3 の定理 A.2 より， $\mathbf{L}$ の n 個の固有値は非負である．それらを大きい順に

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0 \quad (\text{S.6.82})$$

とする．もし， $\mathbf{L}$ のランクが r ($\leq n$) ならば，付録 A.3 の定理 A.3 より， $\mathbf{L}$ の非零の固有値の数は r であり， $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$ となる．このときスペクトル分解 (spectral decomposition)[金谷 18] の手法を用いると， $\mathbf{L}$ は下式のように表せる．

$$\mathbf{L} = \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{u}_i^t \quad (\text{S.6.83})$$

次に，固有値 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$ を対角成分とする対角行列を $\mathbf{\Lambda}_r$ ($\in \mathbf{R}^{r \times r}$) で表し，行列 $\mathbf{U}_r$ ($\in \mathbf{R}^{n \times r}$) を下式で表す．

$$\mathbf{U}_r = (\mathbf{u}_1, \dots, \mathbf{u}_r) \quad (\text{S.6.84})$$

以上より，式 (S.6.83) は

$$\mathbf{L} = (\mathbf{u}_1, \dots, \mathbf{u}_r) \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_r \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_r^t \end{pmatrix} \quad (\text{S.6.85})$$

$$= \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^t \quad (\text{S.6.86})$$

と書ける．対角行列 Λ_r の対角成分 λ_i を $\sqrt{\lambda_i}$ に置き換えた対角行列を $\Lambda_r^{1/2}$ で表わし，下式の行列 $\mathbf{Y}$ を定義する．

$$\mathbf{Y} = \mathbf{U}_r \Lambda_r^{1/2} \quad (\text{S.6.87})$$

$$= (\sqrt{\lambda_1} \mathbf{u}_1, \dots, \sqrt{\lambda_r} \mathbf{u}_r) \quad (\in \mathbf{R}^{n \times r}) \quad (\text{S.6.88})$$

これにより，式 (S.6.86) は

$$\mathbf{L} = \mathbf{U}_r \Lambda_r^{1/2} \Lambda_r^{1/2} \mathbf{U}_r^t \quad (\text{S.6.89})$$

$$= (\mathbf{U}_r \Lambda_r^{1/2}) (\mathbf{U}_r \Lambda_r^{1/2})^t \quad (\text{S.6.90})$$

$$= \mathbf{Y} \mathbf{Y}^t \quad (\text{S.6.91})$$

と書ける．上式は，式 (S.6.78) と同様，内積行列の形になっている．行列 $\mathbf{Y}$ を n 個の r 次元列ベクトル $\mathbf{y}_i$ ($i = 1, \dots, n$) を用いて下式のように記す．

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)^t \quad (\text{S.6.92})$$

$$= \begin{pmatrix} \mathbf{y}_1^t \\ \vdots \\ \mathbf{y}_n^t \end{pmatrix} \quad (\text{S.6.93})$$

ここで注目すべきは， $\mathbf{Y}$ を式 (S.6.88) と式 (S.6.93) の二通りの方法で表している点である．すなわち， $n \times r$ の行列 $\mathbf{Y}$ を，式 (S.6.88) では r 個の n 次元列ベクトルとして表し，式 (S.6.93) では n 個の r 次元行ベクトルとして表している．式 (6.137) と式 (S.6.91) より

$$\mathbf{L} = \mathbf{Y} \mathbf{Y}^t = -\frac{1}{2} \mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n \quad (\text{S.6.94})$$

であり，このとき $\mathbf{D}$ がユークリッド距離で定義できることを示せばよい．行列 $\mathbf{L}$ の (i, j) 成分を l_{ij} とすると， $\mathbf{L} = \mathbf{Y} \mathbf{Y}^t$ より

$$l_{ij} = \mathbf{y}_i^t \mathbf{y}_j \quad (\text{S.6.95})$$

である．上式を用いて $\mathbf{y}_i$ と $\mathbf{y}_j$ のユークリッド二乗距離を求めると

$$\begin{aligned} (\mathbf{y}_i - \mathbf{y}_j)^t (\mathbf{y}_i - \mathbf{y}_j) &= \mathbf{y}_i^t \mathbf{y}_i - 2 \mathbf{y}_i^t \mathbf{y}_j + \mathbf{y}_j^t \mathbf{y}_j \\ &= l_{ii} - 2l_{ij} + l_{jj} \end{aligned} \quad (\text{S.6.96})$$

となる．以下に式 (6.138) を再掲する．

$$l_{ij} = -\frac{1}{2} \left(d_{ij}^2 - \frac{1}{n} \sum_k d_{ik}^2 - \frac{1}{n} \sum_k d_{kj}^2 + \frac{1}{n^2} \sum_k \sum_l d_{kl}^2 \right) \quad (\text{S.6.97})$$

上式より，

$$\begin{aligned} l_{ii} &= -\frac{1}{2} \left(d_{ii}^2 - \frac{1}{n} \sum_k d_{ik}^2 - \frac{1}{n} \sum_k d_{ki}^2 + \frac{1}{n^2} \sum_k \sum_l d_{kl}^2 \right) \\ &= \frac{1}{n} \sum_k d_{ik}^2 - \frac{1}{2n^2} \sum_k \sum_l d_{kl}^2 \end{aligned} \quad (\text{S.6.98})$$

が得られる．上式の導出に当たっては，6.5 節の coffee break で述べた以下の条件を用いた．

$$d_{ii} = 0 \quad (\text{S.6.99})$$

$$d_{ij} = d_{ji} \quad (i \neq j) \quad (\text{S.6.100})$$

同様にして

$$l_{jj} = \frac{1}{n} \sum_k d_{kj}^2 - \frac{1}{2n^2} \sum_k \sum_l d_{kl}^2 \quad (\text{S.6.101})$$

が得られる．式 (S.6.97) , (S.6.98) , (S.6.101) を式 (S.6.96) に代入することにより

$$(\mathbf{y}_i - \mathbf{y}_j)^t (\mathbf{y}_i - \mathbf{y}_j) = d_{ij}^2 \quad (\text{S.6.102})$$

が得られ， d_{ij}^2 は $\mathbf{y}_i$ と $\mathbf{y}_j$ のユークリッド二乗距離となる．したがって， $\mathbf{L}$ が半正定値の対称行列なら，ユークリッド距離によって定義された $\mathbf{D}$ を求めることができる．

問 (3)

距離行列 $\mathbf{D}$ が n 個の d 次元パターン $x_1, \dots, x_n$ に対するユークリッド距離によって定義されるなら，共分散行列 $\mathbf{C}_X$ と， $\mathbf{X}_M$ との関係は，式 (6.96) より

$$\mathbf{X}_M^t \mathbf{X}_M = n \cdot \mathbf{C}_X \quad (\in \mathbf{R}^{d \times d}) \quad (\text{S.6.103})$$

であり, $\mathbf{L}$ と $\mathbf{X}_M$ との関係は, 式 (S.6.78) より

$$\mathbf{X}_M \mathbf{X}_M^t = \mathbf{L} \quad (\in \mathbf{R}^{n \times n}) \quad (\text{S.6.104})$$

である.

付録 A.3 の特異値分解の項で述べたように, 行列 $\mathbf{X}_M^t \mathbf{X}_M$ と行列 $\mathbf{X}_M \mathbf{X}_M^t$ の非零の固有値は一致する. 非零の固有値の中から大きい順に $d' (\leq d)$ 個選び, それらを $\lambda_1, \dots, \lambda_{d'} (> 0)$ とする. 固有値 λ_i に対応する, 行列 $\mathbf{X}_M \mathbf{X}_M^t$ の固有ベクトルを $\mathbf{u}_i$, 行列 $\mathbf{X}_M^t \mathbf{X}_M$ の固有ベクトルを $\mathbf{v}_i$ とする. これらの固有ベクトルを用いて, 行列 $\mathbf{U}_{d'}$, $\mathbf{V}_{d'}$ を

$$\mathbf{U}_{d'} \stackrel{\text{def}}{=} (\mathbf{u}_1, \dots, \mathbf{u}_{d'}) \quad (\in \mathbf{R}^{n \times d'}) \quad (\text{S.6.105})$$

$$\mathbf{V}_{d'} \stackrel{\text{def}}{=} (\mathbf{v}_1, \dots, \mathbf{v}_{d'}) \quad (\in \mathbf{R}^{d \times d'}) \quad (\text{S.6.106})$$

と定義する. 固有ベクトルのノルムは 1 となるよう正規化されており, 下式が成り立つ.

$$\mathbf{U}_{d'}^t \mathbf{U}_{d'} = \mathbf{I}_{d'} \quad (\in \mathbf{R}^{d' \times d'}) \quad (\text{S.6.107})$$

$$\mathbf{V}_{d'}^t \mathbf{V}_{d'} = \mathbf{I}_{d'} \quad (\in \mathbf{R}^{d' \times d'}) \quad (\text{S.6.108})$$

行列 $\mathbf{X}_M$ は, 特異値分解の式 (A.3.1) より

$$\mathbf{X}_M = \mathbf{U}_{d'} \mathbf{\Lambda}_{d'}^{1/2} \mathbf{V}_{d'}^t \quad (\in \mathbf{R}^{n \times d}) \quad (\text{S.6.109})$$

と書ける. ただし, $\mathbf{\Lambda}_{d'}^{1/2}$ は固有値 $\lambda_1, \dots, \lambda_{d'}$ の平方根を対角成分とする $d' \times d'$ の対角行列である (付録の式 (A.3.4) 参照). 式 (6.140) より

$$\mathbf{Y} = \mathbf{U}_{d'} \mathbf{\Lambda}_{d'}^{1/2} \quad (\in \mathbf{R}^{n \times d'}) \quad (\text{S.6.110})$$

であるので, 上式を式 (S.6.109) に代入することにより

$$\mathbf{X}_M = \mathbf{Y} \mathbf{V}_{d'}^t \quad (\in \mathbf{R}^{n \times d}) \quad (\text{S.6.111})$$

が得られる.

主成分分析を実行し, d 次元空間上のパターンを, よりコンパクトな d' 次元空間上に射影することを考える. そのためには共分散行列 $\mathbf{C}_X$ の固有ベクトルを,

対応する固有値の大きい順に d' 個求める必要がある．式 (S.6.103) から明らかのように，この固有ベクトルは $\mathbf{X}_M^t \mathbf{X}_M$ の固有ベクトルと一致し， $\mathbf{v}_1, \dots, \mathbf{v}_{d'}$ である．主成分分析により， d 次元空間上のパターン x_k を， d' 次元空間上に射影してパターン x'_k が得られたとすると

$$x'_k = \mathbf{V}_{d'}^t x_k \quad (\in \mathbf{R}^{d' \times 1}) \quad (k = 1, \dots, n) \quad (\text{S.6.112})$$

と書ける．パターン $x_1, \dots, x_n$ に対するパターン行列を $\mathbf{X}$ ，パターン $x'_1, \dots, x'_n$ に対するパターン行列を $\mathbf{X}'$ と記すと

$$\mathbf{X} = (x_1, \dots, x_n)^t \quad (\in \mathbf{R}^{n \times d}) \quad (\text{S.6.113})$$

$$\mathbf{X}' = (x'_1, \dots, x'_n)^t \quad (\in \mathbf{R}^{n \times d'}) \quad (\text{S.6.114})$$

であり，式 (S.6.112) より，両者の間には下式が成り立つ．

$$\mathbf{X}' = \mathbf{X} \mathbf{V}_{d'} \quad (\text{S.6.115})$$

パターン x_k, x'_k の平均ベクトルを，それぞれ $\bar{x}, \bar{x}'$ とすると，パターン $x_1, \dots, x_n$ に対する平均偏差行列 $\mathbf{X}_M$ と，パターン $x'_1, \dots, x'_n$ に対する平均偏差行列 $\mathbf{X}'_M$ はそれぞれ

$$\mathbf{X}_M = (x_1 - \bar{x}, \dots, x_n - \bar{x})^t \quad (\in \mathbf{R}^{n \times d}) \quad (\text{S.6.116})$$

$$\mathbf{X}'_M = (x'_1 - \bar{x}', \dots, x'_n - \bar{x}')^t \quad (\in \mathbf{R}^{n \times d'}) \quad (\text{S.6.117})$$

と書け，式 (S.6.112) と同様

$$x'_k - \bar{x}' = \mathbf{V}_{d'}^t (x_k - \bar{x}) \quad (k = 1, \dots, n) \quad (\text{S.6.118})$$

が成り立つので，式 (S.6.115) と同様に下式が成り立つ．

$$\mathbf{X}'_M = \mathbf{X}_M \mathbf{V}_{d'} \quad (\text{S.6.119})$$

式 (S.6.111) の両辺に右より $\mathbf{V}_{d'}$ を乗じ，式 (S.6.108) を適用すると下式が得られる．

$$\mathbf{X}_M \mathbf{V}_{d'} = \mathbf{Y} \quad (\text{S.6.120})$$

上式と式 (S.6.119) より

$$\mathbf{Y} = \mathbf{X}'_M \quad (\text{S.6.121})$$

が得られる．

以上より，多次元尺度法（MDS）による d' 次元空間上でのパターン行列 $\mathbf{Y}$ は，主成分分析（PCA）による d' 次元空間上でのパターン行列 $\mathbf{X}'_M$ と一致することが確かめられた．これにより，距離行列 $\mathbf{D}$ がユークリッド距離で定義されるなら，多次元尺度法は主成分分析と等価であることが証明された．

【演習問題 6.8】

式 (6.140) が成り立つことの確認

前問の解答で述べたように，与えられた距離行列 $\mathbf{D}$ がユークリッド距離によって定義されているなら，求めるべき $\mathbf{Y}$ は式 (S.6.87) に示したとおりである．この $\mathbf{Y}$ によって得られる r 次元空間上のパターン $\mathbf{y}_i$ ($i = 1, \dots, n$) を用いてユークリッド距離によって求めた距離行列は $\mathbf{D}$ と一致する，そのことを示したのが，式 (S.6.102) である．

ここで，式 (S.6.83) の r の代わりに， $d' (< r)$ 個の固有値を大きい順に選ぶと

$$\mathbf{L}' = \sum_{i=1}^{d'} \lambda_i \mathbf{u}_i \mathbf{u}_i^t \quad (\text{S.6.122})$$

と書ける．上式で $d' = r$ とすれば $\mathbf{L}'$ は式 (S.6.83) の $\mathbf{L}$ と一致するが， $d' < r$ ならば， $\mathbf{L}'$ は $\mathbf{L}$ に対する近似式となっており， d' が大きくなるほど近似精度が向上する．

前問と同様に，固有値 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{d'} \geq 0$ を対角成分とする対角行列を $\mathbf{\Lambda}_{d'}$ ($\in \mathbf{R}^{d' \times d'}$)，それらの平方根を対角成分とする対角行列を $\mathbf{\Lambda}_{d'}^{1/2}$ で表す．さらに，行列 $\mathbf{U}_{d'}$ ($\in \mathbf{R}^{n \times d'}$) を

$$\mathbf{U}_{d'} = (\mathbf{u}_1, \dots, \mathbf{u}_{d'}) \quad (\text{S.6.123})$$

で表すと，式 (S.6.122) は，式 (S.6.91) と同様にして

$$\mathbf{L}' = \mathbf{U}_{d'} \mathbf{\Lambda}_{d'} \mathbf{U}_{d'}^t \quad (\text{S.6.124})$$

$$= (\mathbf{U}_{d'} \mathbf{\Lambda}_{d'}^{1/2}) (\mathbf{U}_{d'} \mathbf{\Lambda}_{d'}^{1/2})^t \quad (\text{S.6.125})$$

$$= \mathbf{Y} \mathbf{Y}^t \quad (\text{S.6.126})$$

と書ける．ただし，上式の $\mathbf{Y}$ は，式 (S.6.87) の $\mathbf{Y}$ と異なり

$$\mathbf{Y} = \mathbf{U}_{d'} \mathbf{\Lambda}_{d'}^{1/2} \quad (\in \mathbf{R}^{n \times d'}) \quad (\text{S.6.127})$$

である．式 (S.6.87) で求めた y_i ($i = 1, \dots, n$) は， r 次元空間上のパターンであるのに対し，式 (S.6.127) で求めた y_i は，より低次元の d' ($< r$) 次元空間上のパターンである．そのため，これらの d' 次元のパターンを用いて計算される式 (S.6.126) の L' は L と一致せず，その近似となっている．式 (S.6.122) からわかるように， L' は大きい順に d' 個の固有値を用いて設定しているので，式 (S.6.126) の $L' = YY^t$ は，パターン $y_1, \dots, y_n$ の次元数が d' であるという制約の下で， L の最良近似となっている．

以上は， D がユークリッド距離で定義されている場合である．もし， D がユークリッド距離以外の方法で設定されている場合は，注意が必要である．なぜなら，この場合は L の固有値が負になることがあり，式 (S.6.127) の $\Lambda_{d'}^{1/2}$ が計算できないからである．しかしその場合でも，正の固有値のみを大きい順に $\lambda_1 \geq \dots \geq \lambda_{d'} > 0$ となるよう選択すれば， L を近似できるので，ここで述べた手法はそのまま適用できる．詳細は [MKB79] を参照されたい．

以上より，求めるべき行列 Y は，式 (S.6.127) であることが示された*4．

パターン $y_1, \dots, y_n$ の平均が原点と一致することの確認

すべての要素が 1 の n 次元列ベクトルを $\mathbf{1}_n$ とする．式 (6.137) を以下に再掲する．

$$L = -\frac{1}{2} \mathbf{J}_n \mathbf{D}^{(2)} \mathbf{J}_n \quad (\text{S.6.128})$$

下式が成り立つことは明らかである．

$$\begin{aligned} \mathbf{J}_n \mathbf{1}_n &= \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_{nn} \right) \mathbf{1}_n \\ &= \mathbf{1}_n - \frac{1}{n} \mathbf{1}_{nn} \mathbf{1}_n \\ &= \mathbf{1}_n - \mathbf{1}_n \\ &= \mathbf{0} \end{aligned} \quad (\text{S.6.129})$$

*4 距離行列はパターンの回転や平行移動の影響を受けない．したがって，パターン $y_1, \dots, y_n$ の座標値までは一意に決定できず， Y はユニークには決まらない．式 (S.6.127) は解の一例であり，より一般的な解は， $Y = U_{d'} \Lambda_{d'}^{1/2} W_{d'}$ である．ただし， $W_{d'}$ は任意の直交行列である．

したがって、式 (S.6.128) より

$$\mathbf{L}\mathbf{1}_n = -\frac{1}{2}\mathbf{J}_n\mathbf{D}^{(2)}\mathbf{J}_n\mathbf{1}_n = \mathbf{0} \quad (\text{S.6.130})$$

となるので、 $\mathbf{1}_n$ は $\mathbf{L}$ の固有値 0 に対する固有ベクトルである。ここで、 $y_1, \dots, y_n$ の平均ベクトルを $\bar{y}$ とすると、付録 A.4 の式 (A.4.18) より

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \mathbf{Y}^t \mathbf{1}_n \quad (\text{S.6.131})$$

と書ける。式 (S.6.88) より、 $\mathbf{Y}$ は下式で表される。

$$\mathbf{Y} = (\sqrt{\lambda_1} \mathbf{u}_1, \dots, \sqrt{\lambda_{d'}} \mathbf{u}_{d'}) \quad (\text{S.6.132})$$

固有ベクトル $\mathbf{1}_n$ は、他の固有ベクトル $\mathbf{u}_i$ ($i = 1, \dots, d'$) と直交するので、 $\mathbf{u}_i^t \mathbf{1}_n = 0$ が成り立つ。その結果、式 (S.6.131) より

$$\begin{aligned} \bar{y} &= \frac{1}{n} \mathbf{Y}^t \mathbf{1}_n \\ &= \frac{1}{n} (\sqrt{\lambda_1} \mathbf{u}_1^t \mathbf{1}_n, \dots, \sqrt{\lambda_{d'}} \mathbf{u}_{d'}^t \mathbf{1}_n)^t = \mathbf{0} \end{aligned} \quad (\text{S.6.133})$$

となり、 $y_1, \dots, y_n$ の平均は原点と一致することが示された。

なお、演習問題 6.7(3) の結果を用いても、以下のように同様の結果が得られる。

ベクトル $x_1, \dots, x_n$ の平均ベクトル $\bar{x}$ で表すと、式 (S.6.120) より

$$\mathbf{y}_k = \mathbf{V}_{d'}^t (\mathbf{x}_k - \bar{x}) \quad (k = 1, \dots, n) \quad (\text{S.6.134})$$

である。これより

$$\bar{y} = \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k \quad (\text{S.6.135})$$

$$\begin{aligned} &= \mathbf{V}_{d'}^t \frac{1}{n} \sum_{k=1}^n (\mathbf{x}_k - \bar{x}) \\ &= \mathbf{V}_{d'}^t (\bar{x} - \bar{x}) \\ &= \mathbf{0} \end{aligned} \quad (\text{S.6.136})$$

が得られ、 d' 次元の空間では、射影されたベクトルの平均は 0 となり、原点と一致する。

【演習問題 6.9】

本演習問題の $x_1, \dots, x_6$ は事例を表しており、必ずしもベクトル表現できるとは仮定していない。このような多次元尺度法の適用例は、6.5 節の coffee break でも紹介した。

実は、 $x_1, \dots, x_6$ は、以下に示す次元数 $d = 2$ のベクトルであり、式 (6.161) の $D^{(2)}$ は、これらのパターン間のユークリッド二乗距離を求めて得た結果である。

$$\begin{aligned} x_1 &= (1, 1)^t, & x_2 &= (8, 1)^t, & x_3 &= (8, 5)^t \\ x_4 &= (1, 5)^t, & x_5 &= (3, 4)^t, & x_6 &= (9, 8)^t \end{aligned} \quad (\text{S.6.137})$$

本演習問題を解く前に、以下では上記情報を既知として $x_1, \dots, x_6$ の分布状況を調べてみる。

式 (S.6.137) を $d (= 2)$ 次元空間上にプロットしたのが図 S.6.1 であり、 $x_1, \dots, x_6$ を $\circ$ 印で示した。図の $\bullet$ 印は 6 パターンの平均ベクトル $\bar{x} = (5, 4)^t$ を示している。また、 x_1, x_2, x_3, x_4 が長方形を構成していることを、細線で示している。図の平均ベクトル $\bar{x}$ を通り、互いに直交する二つの直線は、主成分分析の結果得られた主軸と副軸を表している。これらの軸の方向を決定するのが、パターンの共分散行列 C_X の固有ベクトル v_1, v_2 である。これらは次のようにして求めることができる。

平均偏差行列 X_M は式 (A.4.19) より

$$X_M = \begin{pmatrix} -4 & -3 \\ 3 & -3 \\ 3 & 1 \\ -4 & 1 \\ -2 & 0 \\ 4 & 4 \end{pmatrix} \quad (\text{S.6.138})$$

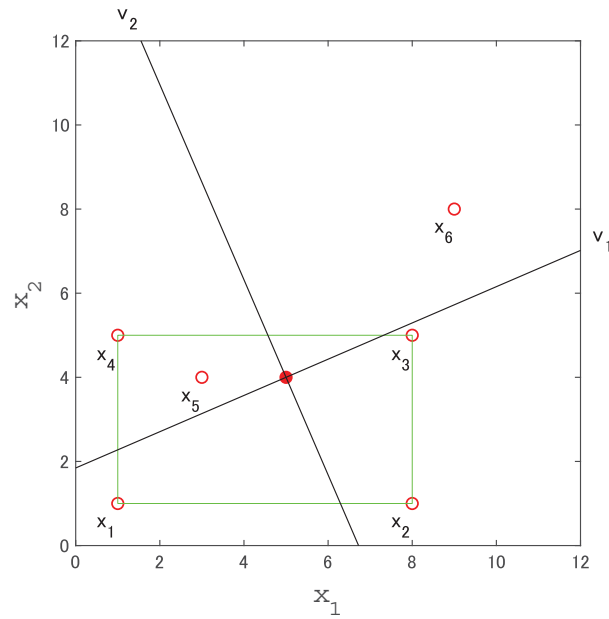


図 S.6.1: 元の d 次元空間上のパターン $x_1, \dots, x_6$

であり，式 (S.6.103) を用いると，下式が得られる．

$$\mathbf{X}_M^t \mathbf{X}_M = 6 \cdot \mathbf{C}_X = \begin{pmatrix} 70 & 18 \\ 18 & 36 \end{pmatrix} \quad (\text{S.6.139})$$

図の $\mathbf{v}_1, \mathbf{v}_2$ は，共分散行列 $\mathbf{C}_X$ の固有ベクトルであり， $\mathbf{X}_M^t \mathbf{X}_M$ の固有ベクトルでもある．そこで， $\mathbf{X}_M^t \mathbf{X}_M$ の固有ベクトルを求め，行列 $\mathbf{V} \stackrel{\text{def}}{=} (\mathbf{v}_1, \mathbf{v}_2)$ とすると，下式が得られる．

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2) = \begin{pmatrix} -0.918 & 0.396 \\ -0.396 & -0.918 \end{pmatrix} \quad (\text{S.6.140})$$

二つの固有ベクトルのうち，最大固有値に対応する固有ベクトルが $\mathbf{v}_1$ であり，他の一つが $\mathbf{v}_2$ である．

次に，MDS の手法を用いた本演習問題の解法を示す．式 (6.161) より，式 (6.137)，または式 (6.138) を用いて行列 L を求めると，下式が得られる．

$$\mathbf{L} = \begin{pmatrix} 25 & -3 & -15 & 13 & 8 & -28 \\ -3 & 18 & 6 & -15 & -6 & 0 \\ -15 & 6 & 10 & -11 & -6 & 16 \\ 13 & -15 & -11 & 17 & 8 & -12 \\ 8 & -6 & -6 & 8 & 4 & -8 \\ -28 & 0 & 16 & -12 & -8 & 32 \end{pmatrix} \quad (\text{S.6.141})$$

次に行列 L の固有値と固有ベクトルを求める．ここでは $d' = 2$ であるので，求めるべきは，二つの固有値 λ_1, λ_2 と，二つの固有ベクトル $\mathbf{u}_1, \mathbf{u}_2$ である．計算の結果

$$\lambda_1 = 77.76 \quad \lambda_2 = 28.24 \quad (\text{S.6.142})$$

であるので， $\Lambda^{1/2}$ は

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} = \begin{pmatrix} 8.818 & 0 \\ 0 & 5.314 \end{pmatrix} \quad (\text{S.6.143})$$

となる．また行列 $\mathbf{U}$ は

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2) = \begin{pmatrix} -0.551 & -0.220 \\ 0.178 & -0.742 \\ 0.357 & -0.051 \\ -0.372 & 0.471 \\ -0.208 & 0.149 \\ 0.596 & 0.393 \end{pmatrix} \quad (\text{S.6.144})$$

である．式 (6.140) に式 (S.6.143)，式 (S.6.144) を代入することにより

$$\mathbf{Y} = \mathbf{U}\Lambda^{1/2} = \begin{pmatrix} \mathbf{y}_1^t \\ \mathbf{y}_2^t \\ \mathbf{y}_3^t \\ \mathbf{y}_4^t \\ \mathbf{y}_5^t \\ \mathbf{y}_6^t \end{pmatrix} = \begin{pmatrix} -4.861 & -1.172 \\ 1.567 & -3.942 \\ 3.151 & -0.269 \\ -3.277 & 2.502 \\ -1.837 & 0.792 \\ 5.257 & 2.090 \end{pmatrix} \quad (\text{S.6.145})$$

を得る．以上で本演習問題の解である $y_1, \dots, y_6$ を求めることができた．これらをプロットした結果が図 S.6.2 である．

すでに述べたように, $x_1, \dots, x_6$ はベクトルで表すことができ, その次元は $d = 2$ であり, かつ $d' = d$ である．しかも与えられた $D^{(2)}$ は, パターン間のユークリッド二乗距離によって算出されている．したがって, 多次元尺度法によって求めた $y_1, \dots, y_6$ は, 元の $x_1, \dots, x_6$ の相対的位置関係を復元しているはずである．ただし, 演習問題解答 64 ページの脚注で述べたように, 与えられているのはパターン間の二乗距離であるので, 回転や平行移動の自由度があり, $y_1, \dots, y_6$ の座標値を一意に決めることはできない．

実際, 図 S.6.2 を図 S.6.1 と比較すると, 両者の座標値は一致しないものの,

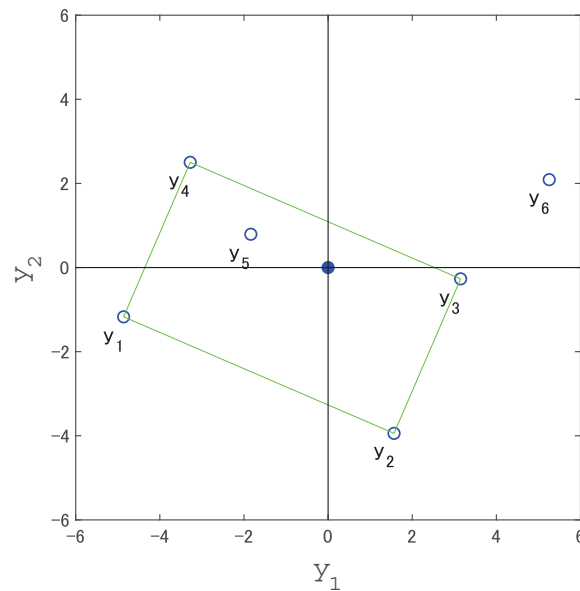


図 S.6.2: 多次元尺度法によって得た d' 次元空間上のパターン $y_1, \dots, y_6$

図 S.6.2 に回転と平行移動を施すことにより，図 S.6.1 と一致させることができる．両図の比較がしやすいよう，図 S.6.2 の y_1, y_2, y_3, y_4 を細線で結び，図 S.6.1 の長方形と対応していることを示した．また，図 S.6.1 の主軸と副軸は図 S.6.2 の横軸と縦軸に対応していることがわかる．これは，式 (S.6.136) で示したように，距離行列がユークリッド距離で定義されているときには，多次元尺度法が主成分分析と等価であることを示している．さらに，図 S.6.2 を見ると，平均が原点となっており，式 (S.6.133)，式 (S.6.136) が成り立つことも確認できる．

【演習問題 6.10】

アルゴリズム（手法）の検証，評価には，様々な条件の下で実験を行い，比較検討を行うことが必要となる．通常，論文等で公開するのは，論文の主張の根拠のうちの一部でしかない．研究開発の過程においては，アルゴリズムが想定どおりに機能しているか，プログラムが正しく実装されているか，などについて丁寧に検証した上で，アルゴリズムの特性の把握，性能の評価を行うことが必要になる．これら一連の実験，分析を経てはじめて，そのアルゴリズムを使いこなすことができたことになると言えよう．

表 6.1 の結果から観察できる事実と考察，およびさらなる議論を進める上で留意すべき点は以下である．

1. 二つの識別法，CLAFIC 法 (CLF) と投影距離法 (PRJ) は，ともに累積寄与率 c_p が 1 に近づくと，学習パターンに対するエラー率が減少する．これは手法の性質から予想される結果であり，プログラムは正しく動いていることの検証にもなっている．
2. 一方，テストパターンに対するエラー率は， c_p に対して，1 つの極小点を持つ滑らかな曲線を描いていると推察される．この表では記載が省かれているが，PRJ については，より小さい c_p での確認が必要である．
3. 識別性能に関しては，CLF と PRJ 間で顕著な差はない．ただし，それが一般的に成り立つ傾向なのか，MNIST データ特有の傾向なのかは，この結果だけから結論づけることはできない．
4. 原空間の次元は $d = 784$ であるが， c_p に対する d' の平均値の推移から，

MNIST データは、各クラスとも、比較的少ない主成分で近似できることがわかる。

5. 累積寄与率 c_p に対する d' の平均値は、CLF と PRJ との間で数倍程度の差があることには注意が必要である。その差が生じている理由を明らかにすることは難しいが、以下に述べる MNIST データの特性に起因している可能性はある。すなわち、実験で使った MNIST データは、784 次元の $[0,1]$ 区間の多値データであるが、データを丁寧に解析してみるとわかるように、実は、ほとんどが二値画像に近く、0 もしくは 1 に近い値を持っている。言い換えれば、784 次元の超立方体の頂点付近に多くのデータが存在しているのである。このことは、MNIST データを用いた実験を行う場合には念頭に置く必要がある。
6. 識別性能を CLF, PRJ 間で比較するときには、 c_p に対するエラー率の差だけでなく、 d' の平均値に対するエラー率の差も観察しておくことが重要である。
7. 本表では、 k -最近傍法については、 $k = 1$ のみが記載されているが、より大きな k に対するエラー率をみておくことは、MNIST におけるクラス間の分布の重なり程度をつかむ上で重要である。

表 6.2 では、カーネル関数として、ガウスカーネル、多項式カーネルを用いたときの特性の一部を比較することができる。ただし、特定のカーネルパラメータに対しての特性である点には注意を払う必要がある。

1. カーネル部分空間法を用いた場合、線形部分空間法 (LSS) と同様に、 c_p の増加とともに、学習データに対するエラー率は減少する。特にガウスカーネルの場合には、小さい c_p で学習パターンに対するエラー率が 0 になる。多項式カーネルと比較して、小さい c_p で、 d' が大きくなることに特徴がある。
2. 一方、テストパターンに対するエラー率は c_p とともに単調減少している点が LSS との大きな違いである。ただし、より大きな c_p においてどのような挙動を示すかは、検証が必要である。また、そのエラー率は、LSS に比べて小さく、カーネル法の有効性が明らかである。
3. ガウスカーネルの場合、多項式カーネルと比較して、小さい c_p で、 d' が

大きくなることに特徴がある．この条件における ϕ 空間の次元が，多項式カーネルでは，式 (6.42) より

$$28 \cdot 28 + 3 C_3 = 787 C_3 \approx 0.8 \text{ 億}$$

であるのに対し，ガウスカーネルでは，無限次元であることは注意したい．また，線形部分空間法（表 6.1）の実験 4, 5 及び実験 8, 9 では， c_p が増えるにしたがって，学習パターンに対するエラー率は減少している一方で，テストパターンに対するエラー率は増加しており，いわゆる，過学習が観測される．しかし，カーネル部分空間法（表 6.2）では，この実験条件の範囲内においては，過学習が観測されていないということは着目してよいだろう．

表 6.3 では， c_p を 0.9 に固定し，カーネルパラメータ（ σ と p ）の影響を調べた結果である．

1. 多項式カーネルでは， p が大きいほど ϕ 空間の次元は大きくなるが，学習パターンに対するエラー率が必ずしも小さくなっていない点に注意が必要である．また， d' の値も p とともに大きくなるわけではない．
2. エラー率， d' の平均値におけるこの事実は，次のことを示唆している．すなわち， p を大きくすることによって得られる ϕ 空間上の新たな次元は，パターンの分布を表す上で，ほとんど意味のない次元である．
3. 一方，ガウスカーネルの場合は， σ とともに d' は小さくなっている．これは，SVM の項においても観察したように，小さい σ では，各パターンが ϕ 空間上における独立した軸を表すのに対し， σ が大きくなるとその独立性が薄れていることを意味している．
4. テストパターンに対するエラー率は， σ ， p のどちらの場合も，一つの極小点をもつ滑らかな変化を示すと推察される．この事実は，最適な σ ， p を探す上で重要である．

第 7 章

【演習問題 7.1】

式 (7.49) の条件のもとで交差エントロピー $H(\mathbf{t}, \mathbf{g})$ が最小となる $\mathbf{t}$ を求めるには、ラグランジュの未定乗数法を用いる。すなわち

$$L = H(\mathbf{t}, \mathbf{g}) - \lambda \left(\sum_{m=1}^c g_m - 1 \right) \quad (\text{S.7.1})$$

$$= - \sum_{m=1}^c b_m \cdot \log g_m - \lambda \left(\sum_{m=1}^c g_m - 1 \right) \quad (\text{S.7.2})$$

と置き、 L を最小化する $\mathbf{g}$ を求めればよい。ただし、 λ は定数である。そこで、上式を g_k で偏微分して 0 とおくことにより

$$\frac{\partial L}{\partial g_k} = -\frac{b_k}{g_k} - \lambda = 0 \quad (\text{S.7.3})$$

が得られ、

$$\lambda g_k = -b_k \quad (k = 1, \dots, c) \quad (\text{S.7.4})$$

となる。上式の両辺に $\sum_{k=1}^c$ を施した後、式 (7.8)、(7.49) を適用すると

$$\lambda = -1 \quad (\text{S.7.5})$$

となり、この結果を式 (S.7.4) に代入することにより

$$g_k = b_k \quad (k = 1, \dots, c) \quad (\text{S.7.6})$$

が得られる。すなわち、交差エントロピー $H(\mathbf{t}, \mathbf{g})$ が最小となるのは、 $\mathbf{g} = \mathbf{t}$ のときであることが示された。

【演習問題 7.2】

出力層にソフトマックス関数を用いたとき、出力層のユニット m からの出力 g_m は、式 (7.48) で表され、再掲すると下式である。

$$\begin{aligned} g_m &= f(h_1, \dots, h_m, \dots, h_c) \\ &= \frac{\exp(h_m)}{\sum_{r=1}^c \exp(h_r)} \quad (m = 1, \dots, c) \end{aligned} \quad (\text{S.7.7})$$

また，ユニット m での ε_m は，式 (7.15) の j を m に変えて

$$\varepsilon_m = \frac{\partial J_p}{\partial h_m} \quad (\text{S.7.8})$$

となる．上式の微分を実行するにあたり，式 (7.55) の J_p を，下式のごとく g_m を含む項と含まない項の二つに分けておく必要がある．

$$J_p = - \sum_{q=1}^c b_q \cdot \log g_q \quad (\text{S.7.9})$$

$$= -b_m \cdot \log g_m - \sum_{q \neq m}^c b_q \cdot \log g_q \quad (\text{S.7.10})$$

上式と式 (S.7.8) より

$$\begin{aligned} \varepsilon_m &= \frac{\partial J_p}{\partial h_m} \\ &= -\frac{b_m}{g_m} \cdot \frac{\partial g_m}{\partial h_m} - \sum_{q \neq m}^c \frac{b_q}{g_q} \cdot \frac{\partial g_q}{\partial h_m} \end{aligned} \quad (\text{S.7.11})$$

が得られる．ここで，式 (S.7.11) の $\partial g_m / \partial h_m$ は，式 (S.7.7) より

$$\frac{\partial g_m}{\partial h_m} = \frac{\exp(h_m) \cdot \sum_{r=1}^c \exp(h_r) - \exp(h_m) \cdot \exp(h_m)}{(\sum_{r=1}^c \exp(h_r))^2} \quad (\text{S.7.12})$$

$$= \frac{\exp(h_m)}{\sum_{r=1}^c \exp(h_r)} - \left(\frac{\exp(h_m)}{\sum_{r=1}^c \exp(h_r)} \right)^2 \quad (\text{S.7.13})$$

$$= g_m - g_m^2 \quad (\text{S.7.14})$$

となり，式 (S.7.11) の $\partial g_q / \partial h_m$ は

$$\frac{\partial g_q}{\partial h_m} = \frac{-\exp(h_q) \cdot \exp(h_m)}{(\sum_{r=1}^c \exp(h_r))^2} \quad (\text{S.7.15})$$

$$= -g_q \cdot g_m \quad (q \neq m) \quad (\text{S.7.16})$$

となる．式 (S.7.14) , (S.7.16) を式 (S.7.11) に代入すると

$$\varepsilon_m = -\frac{b_m}{g_m} \cdot (g_m - g_m^2) + \sum_{q \neq m}^c \frac{b_q}{g_q} (g_q \cdot g_m) \quad (\text{S.7.17})$$

$$= b_m \cdot (g_m - 1) + \sum_{q \neq m}^c b_q \cdot g_m \quad (\text{S.7.18})$$

$$= b_m \cdot g_m + g_m \cdot \sum_{q \neq m}^c b_q - b_m \quad (\text{S.7.19})$$

$$= g_m \cdot \left(b_m + \sum_{q \neq m}^c b_q \right) - b_m \quad (\text{S.7.20})$$

$$= g_m \cdot \sum_{q=1}^c b_q - b_m \quad (\text{S.7.21})$$

$$= g_m - b_m \quad (\text{S.7.22})$$

となり，式 (7.56) が導出できた．なお，式 (S.7.21) から式 (S.7.22) への変形には式 (7.8) を用いた．

第 8 章

【演習問題 8.1】

時間 t を変数とする信号 $f(t)$ のフーリエ変換 $\mathcal{F}(f(t))$ を $F(\omega)$ で表すと

$$F(\omega) = \mathcal{F}(f(t)) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (\text{S.8.1})$$

と書ける．ここで ω は角周波数を表す．同様にして， $g(t)$ ， $h(t)$ のフーリエ変換 $\mathcal{F}(g(t))$ ， $\mathcal{F}(h(t))$ を，それぞれ $G(\omega)$ ， $H(\omega)$ で表すと，式 (8.34) は

$$H(\omega) = F(\omega) \cdot G(\omega) \quad (\text{S.8.2})$$

と書ける．式 (8.33) より

$$\begin{aligned} H(\omega) &= \mathcal{F}(h(t)) \\ &= \mathcal{F}(f(t) * g(t)) \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau \right] e^{-j\omega t} dt \\ &= \int_{-\infty}^{\infty} f(\tau) \left[\int_{-\infty}^{\infty} g(t - \tau) e^{-j\omega t} dt \right] d\tau \end{aligned} \quad (\text{S.8.3})$$

が得られ，ここで $t - \tau = u$ と置くと

$$\begin{aligned} \int_{-\infty}^{\infty} g(t - \tau) e^{-j\omega t} dt &= \int_{-\infty}^{\infty} g(u) e^{-j\omega(u+\tau)} du \\ &= e^{-j\omega\tau} \int_{-\infty}^{\infty} g(u) e^{-j\omega u} du \\ &= e^{-j\omega\tau} G(\omega) \end{aligned} \quad (\text{S.8.4})$$

が成り立つ．式 (S.8.4) を式 (S.8.3) に代入することにより

$$H(\omega) = \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} G(\omega) d\tau \quad (\text{S.8.5})$$

$$= G(\omega) \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} d\tau \quad (\text{S.8.6})$$

$$= F(\omega) \cdot G(\omega) \quad (\text{S.8.7})$$

が得られ，式 (S.8.2)，すなわち式 (8.34) が証明された．