演習問題解答 （わかりやすいパターン認識（第 2 版））

第１章

【演習問題 1.1】

各クラスのプロトタイプを $\mathbf{p}_i$ $(i = 0, 1, 2, \ldots, 9)$ で表す．ここで，$\mathbf{p}_0$ は数字 ‘ 0 ’ のプロトタイプ，$\mathbf{p}_1$ は数字 ‘ 1 ’ のプロトタイプ，以下同様に $\mathbf{p}_i$ は数字 ‘ i ’ のプロトタイプを表すものとする．これらのプロトタイプを用いて，入力パターン $\boldsymbol{x}_k$ $(k = 1, 2, 3, 4)$ を最近傍決定則により識別するには，式 (1.4) より

$$\underset{i}{\operatorname{argmin}}\ D(\boldsymbol{x}_k,\, \mathbf{p}_i) = j \quad \Longrightarrow \quad \boldsymbol{x}_k \text{ は数字 ‘ } j \text{ ’} \tag{S.1.1}$$

とすればよい．ただし，$D(\boldsymbol{x}_k,\, \mathbf{p}_i)$ は，$\boldsymbol{x}_k$ と $\mathbf{p}_i$ とのユークリッド距離を表す．

与えられた入力パターンを見ると，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_3$ は，それぞれ数字 ‘ 2 ’, ‘ 4 ’, ‘ 7 ’ であり，$\boldsymbol{x}_4$ は数字ではないことがわかる．パターン $\boldsymbol{x}_k$ とプロトタイプ $\mathbf{p}_i$ は，ともに 25 次元の 2 値ベクトルである．両者のユークリッド距離 $D(\boldsymbol{x}_k,\, \mathbf{p}_i)$ の計算結果を**表 S.1.1** にまとめた．ただし，表の各行は，入力パターン $\boldsymbol{x}_k$ とプロトタイプ $\mathbf{p}_0$〜$\mathbf{p}_9$ とのユークリッド距離の二乗を示しており，黒線の四角で囲んだ数値はそれらの最小値を示している．また，表の灰色線の四角で囲んだ数値は，距離が 2 番目に小さい数値である．

パターン $\boldsymbol{x}_1$ に対して式 (S.1.1) を適用すると

$$\underset{i}{\operatorname{argmin}}\ D(\boldsymbol{x}_1,\, \mathbf{p}_i) = 2 \quad \Longrightarrow \quad \boldsymbol{x}_1 \text{ は数字 ‘ } 2 \text{ ’} \tag{S.1.2}$$

となり，パターン $\boldsymbol{x}_1$ は数字 ‘ 2 ’ と正しく識別される．このときの最小値は，$D(\boldsymbol{x}_1,\, \mathbf{p}_2)^2 = 3$ である．距離が 2 番目に小さい数値としては，$D(\boldsymbol{x}_1,\, \mathbf{p}_7)^2 = 8$ が該当し，次候補が ‘ 7 ’ であることを示している．ここで，1 位と 2 位との距離の差は十分余裕があるので，高い信頼度を以って識別結果を ‘ 2 ’ とすることができる．

表 S.1.1: 入力パターンとプロトタイプとの距離 (1)

		プロトタイプ									
		$\boldsymbol{p}_0$	$\boldsymbol{p}_1$	$\boldsymbol{p}_2$	$\boldsymbol{p}_3$	$\boldsymbol{p}_4$	$\boldsymbol{p}_5$	$\boldsymbol{p}_6$	$\boldsymbol{p}_7$	$\boldsymbol{p}_8$	$\boldsymbol{p}_9$
入力パターン	$\boldsymbol{x}_1$	12	11	**3**	10	14	11	10	**8**	11	13
	$\boldsymbol{x}_2$	14	13	15	14	**10**	13	**10**	14	13	**9**
	$\boldsymbol{x}_3$	11	10	**8**	9	13	10	13	**7**	10	10
	$\boldsymbol{x}_4$	16	17	**11**	14	14	15	16	14	**13**	**13**

同様にして，パターン $\boldsymbol{x}_2$ に式 (S.1.1) を適用すると，表 S.1.1 より明らかなように ‘ 9 ’ と誤識別される．このときの最小値は，$D(\boldsymbol{x}_2,\, \mathbf{p}_9)^2 = 9$ である．ここで次候補を見てみると，‘ 4 ’ と ‘ 6 ’ であり，両プロトタイプとの距離はいずれも 10 で，最小値 9 に近接していることがわかる．このような場合には，あえて識別結果を出力せず，リジェクトするのがよい．特に数字は金額を記すのに使われることも多く，誤識別の影響は大きい．この例のように，1 位と 2 位との距離差に余裕が無く，識別結果に対する信頼度が低いと判定された場合にはリジェクトし，人間に判断を委ねるのが安全である．

パターン $\boldsymbol{x}_3$ は，最小値が $D(\boldsymbol{x}_3,\, \mathbf{p}_7)^2 = 7$ で，正しい識別結果 ‘ 7 ’ が得られる．しかし，$D(\boldsymbol{x}_3,\, \mathbf{p}_2)^2 = 8$ であり，次候補の ‘ 2 ’ と距離が近接しているので，パターン $\boldsymbol{x}_2$ と同様，リジェクトするのが望ましい．

パターン $\boldsymbol{x}_4$ については，最小値が $D(\boldsymbol{x}_4,\, \mathbf{p}_2)^2 = 11$ であり，単に最小距離識別法を適用した場合には，‘ 2 ’ と識別される．しかし，この最小値 11 ではパターン $\boldsymbol{x}_4$ はいずれのプロトタイプとも距離があり，数字パターンではない可能性が高いことを示唆している．したがって，リジェクトとするのが妥当である．

リジェクトについては 2 種類あることを，1.2 節 ［2］の脚注で述べた．その一つが本問の $\boldsymbol{x}_2$，$\boldsymbol{x}_3$ に対するリジェクトで，1 位と 2 位との距離差が 1 以下の

場合を，複数クラスが候補に挙がったとみなし，リジェクトとした．

他の一つが $\boldsymbol{x}_4$ に対するリジェクトで，距離値が 10 以上の場合を候補無しとみなし，リジェクトとした．

リジェクトの条件に関して，明確な基準があるわけではない．誤識別に対してどの程度のリスクを許容できるか，人間に判断を委ねる際に負担すべき労力をどの程度受け入れられるか等によって決めなくてはならない．

【演習問題 1.2】

各クラスのプロトタイプ $\mathbf{p}_i$ $(i = 0, 1, 2, \ldots, 9)$ と，入力パターン $\boldsymbol{x}_k$ $(k = 1, 2, 3, 4)$ を，図 1.10 のヒストグラムに基づく 10 次元特徴ベクトルで表現する．このようにして得られた特徴ベクトル間で，前問と同様 $D(\boldsymbol{x}_k, \mathbf{p}_i)^2$ を求めた結果が**表 S.1.2** である．表の見方は前問と同様である．

表 S.1.2: 入力パターンとプロトタイプとの距離 (2)

		プロトタイプ									
		$\boldsymbol{p}_0$	$\boldsymbol{p}_1$	$\boldsymbol{p}_2$	$\boldsymbol{p}_3$	$\boldsymbol{p}_4$	$\boldsymbol{p}_5$	$\boldsymbol{p}_6$	$\boldsymbol{p}_7$	$\boldsymbol{p}_8$	$\boldsymbol{p}_9$
入力パターン	$\boldsymbol{x}_1$	20	52	**4**	12	44	26	18	30	12	30
	$\boldsymbol{x}_2$	32	60	48	26	30	42	**14**	42	20	**12**
	$\boldsymbol{x}_3$	16	34	20	**10**	28	32	22	**6**	14	12
	$\boldsymbol{x}_4$	32	98	28	**26**	44	46	32	52	**22**	42

式 (S.1.1) に従い，単に最近傍決定則を適用した場合には，パターン，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_3$，$\boldsymbol{x}_4$ の識別結果はそれぞれ ‘ 2 ’ (正)，‘ 9 ’ (誤)，‘ 7 ’ (正)，‘ 8 ’ (誤) となる．ここで，1 位と 2 位との距離差が 3 以下，あるいは距離値が 20 以上をリジェクトと判定することにすると，上記識別結果は，‘ 2 ’ (正)，‘ リジェクト ’，‘ 7 ’ (正)，‘ リジェクト ’ となる．

第２章

【演習問題 2.1】

問 (1)

各クラスの線形識別関数を求めるには，式 (2.2) を用いればよい．すなわち，クラス ω_1 の線形識別関数 $g_1(\boldsymbol{x})$ は，そのプロトタイプ $\mathbf{p}_1 = (4, 16)^t$ より，

$$\frac{1}{2}\|\mathbf{p}_1\|^2 = (4^2 + 16^2)/2 = 136 \tag{S.2.1}$$

であるので

$$g_1(\boldsymbol{x}) = -136 + 4x_1 + 16x_2 \tag{S.2.2}$$

と求められる．同様にして，$g_2(\boldsymbol{x})$，$g_3(\boldsymbol{x})$ は

$$g_2(\boldsymbol{x}) = -90 + 12x_1 + 6x_2 \tag{S.2.3}$$
$$g_3(\boldsymbol{x}) = -234 + 12x_1 + 18x_2 \tag{S.2.4}$$

となる．

問 (2)

決定境界は $g_{12}(\boldsymbol{x}) = 0$，$g_{13}(\boldsymbol{x}) = 0$，$g_{23}(\boldsymbol{x}) = 0$ の 3 種であり，$g_{ij}(\boldsymbol{x})$ は，$g_{ij}(\boldsymbol{x}) = g_i(\boldsymbol{x}) - g_j(\boldsymbol{x})$ として求められる．たとえば，クラス ω_1，ω_2 間の決定境界は，式 (S.2.2)，(S.2.3) を用いて

$$g_{12}(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) \tag{S.2.5}$$
$$= -46 - 8x_1 + 10x_2 = 0 \tag{S.2.6}$$

として求められる．決定境界 $g_{ij}(\boldsymbol{x})$ は，その符号のみに意味があるので，式 (S.2.6) の係数を正の定数で除すことができる．式 (S.2.6) の係数を 2 で除した結果，決定境界として

$$g_{12}(\boldsymbol{x}) = -23 - 4x_1 + 5x_2 = 0 \tag{S.2.7}$$

が得られる．同様にして

$$g_{13}(\boldsymbol{x}) = 49 - 4x_1 - x_2 = 0 \tag{S.2.8}$$
$$g_{23}(\boldsymbol{x}) = 12 - x_2 = 0 \tag{S.2.9}$$

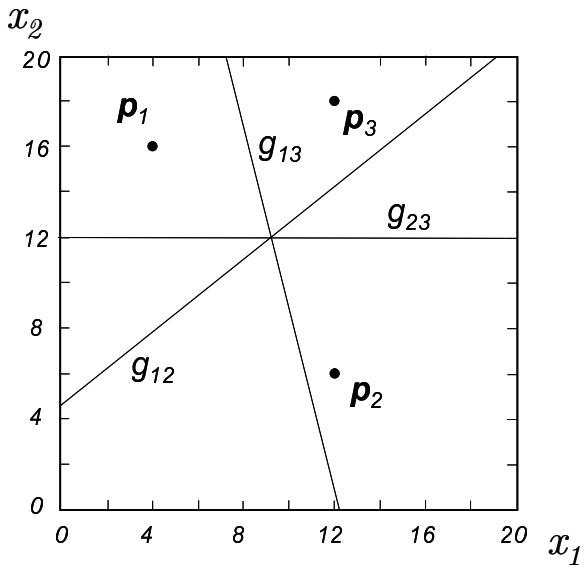


図 S.2.1: 3 種の決定境界

となる．上記 3 種の決定境界を，プロトタイプ $\mathbf{p}_1$，$\mathbf{p}_2$，$\mathbf{p}_3$ とともに**図 S.2.1** に示した．図では決定境界 $g_{ij}(\boldsymbol{x}) = 0$ を単に g_{ij} と記した．

式 (2.3) より，パターン $\boldsymbol{x}$ がクラス ω_1 に属すると判定されるのは

$$g_1(\boldsymbol{x}) > g_2(\boldsymbol{x}), \quad g_1(\boldsymbol{x}) > g_3(\boldsymbol{x}) \tag{S.2.10}$$

が成り立つ場合であるので

$$\boldsymbol{x} \in \omega_1 \text{ の判定領域：} \ g_{12}(\boldsymbol{x}) > 0, \quad g_{13}(\boldsymbol{x}) > 0 \tag{S.2.11}$$

である．同様にして

$$\boldsymbol{x} \in \omega_2 \text{ の判定領域：} \ g_{12}(\boldsymbol{x}) < 0, \quad g_{23}(\boldsymbol{x}) > 0 \tag{S.2.12}$$

$$\boldsymbol{x} \in \omega_3 \text{ の判定領域：} \ g_{13}(\boldsymbol{x}) < 0, \quad g_{23}(\boldsymbol{x}) < 0 \tag{S.2.13}$$

となる．それらを示したのが**図 S.2.2** の太線である．これによって，特徴空間は

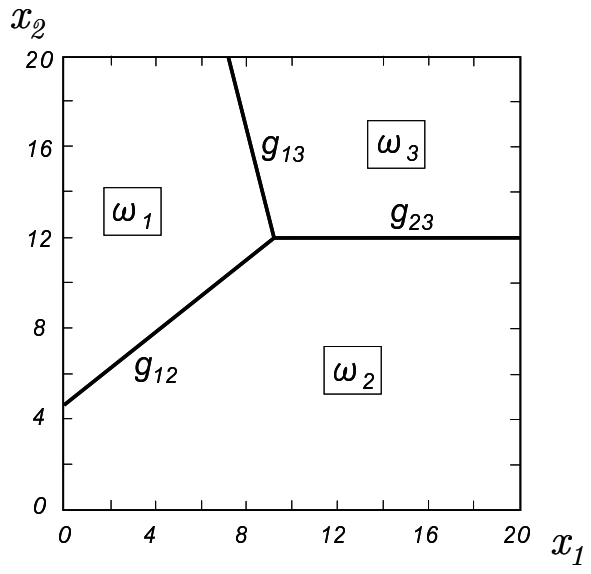


図 S.2.2: 決定境界と 3 クラスの判定領域

3 つの領域に分割され，それぞれがクラス ω_1，ω_2，ω_3 と判定される領域に対応している．各領域のクラス名は□の中に記されている．

問 (3)

式 (S.2.2)〜(S.2.4) と同様にして

$$g_4(\boldsymbol{x}) = -10 + 2x_1 + 4x_2 \tag{S.2.14}$$

と求められる．

問 (4)

式 (S.2.7)〜(S.2.9) と同様にして

$$g_{14}(\boldsymbol{x}) = -63 + x_1 + 6x_2 = 0 \tag{S.2.15}$$
$$g_{24}(\boldsymbol{x}) = -40 + 5x_1 + x_2 = 0 \tag{S.2.16}$$

$$g_{34}(\boldsymbol{x}) = -112 + 5x_1 + 7x_2 = 0 \tag{S.2.17}$$

となる．これらの決定境界を図 S.2.1 に追記すると**図 S.2.3** となる．クラス数は 4 であるので，決定境界の総数は ${}_4C_2 = 6$ となり，本図ではそれらがすべて示されていることになる．クラス数が 3 の場合，決定境界は 2 次元特徴空間上で 1 点

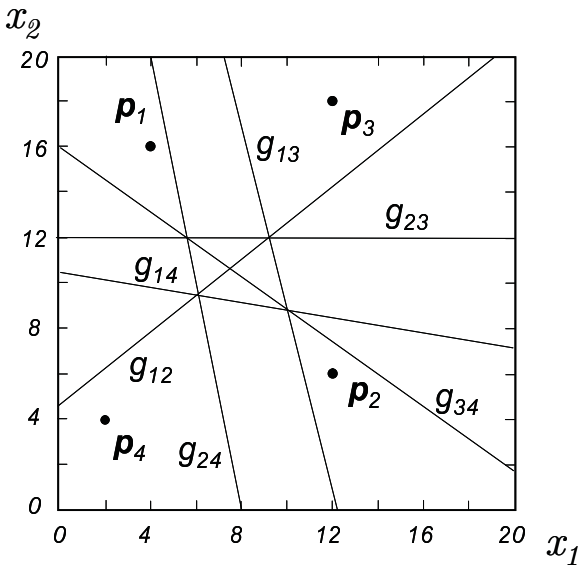


図 S.2.3: 6 種の決定境界

で交わるため，3 領域への分割は容易である．一方，クラス数が 4 になると，各クラスの領域決定はやや面倒になる．

式 (S.2.11)〜(S.2.13) と同様にして，以下が言える．

$$\left.\begin{array}{lll} \boldsymbol{x} \in \omega_1 \text{ の判定領域：} & g_{12}(\boldsymbol{x}) > 0, & g_{13}(\boldsymbol{x}) > 0, \quad g_{14}(\boldsymbol{x}) > 0 \\ \boldsymbol{x} \in \omega_2 \text{ の判定領域：} & g_{12}(\boldsymbol{x}) < 0, & g_{23}(\boldsymbol{x}) > 0, \quad g_{24}(\boldsymbol{x}) > 0 \\ \boldsymbol{x} \in \omega_3 \text{ の判定領域：} & g_{13}(\boldsymbol{x}) < 0, & g_{23}(\boldsymbol{x}) < 0, \quad g_{34}(\boldsymbol{x}) > 0 \\ \boldsymbol{x} \in \omega_4 \text{ の判定領域：} & g_{14}(\boldsymbol{x}) < 0, & g_{24}(\boldsymbol{x}) < 0, \quad g_{34}(\boldsymbol{x}) < 0 \end{array}\right\} \tag{S.2.18}$$

上記によって定まる 4 つの領域を**図 S.2.4** に示した．図の見方は図 S.2.2 と同じである．

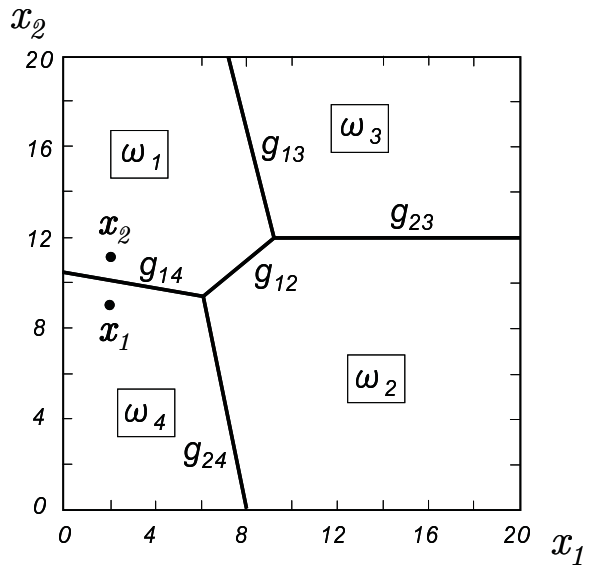


図 S.2.4: 決定境界と 4 クラスの判定領域

ここで注意すべきは，決定境界 $g_{34}(\boldsymbol{x}) = 0$ である．図から明らかなように，$g_{34}(\boldsymbol{x}) = 0$ はクラス ω_3 と ω_4 の境界設定には関与していない．すなわち，式 (S.2.18) の ω_3 の領域を規定する $g_{34}(\boldsymbol{x}) > 0$，および ω_4 の領域を規定する $g_{34}(\boldsymbol{x}) < 0$ は不要である．図 S.2.2 および図 S.2.4 で示した境界は，本文 2.5 節［1］で述べたボロノイ図に他ならない．

問 (5)

パターン $\boldsymbol{x}_1 = (2, 9)^t$ に対しては

$$g_1(\boldsymbol{x}_1) = 16, \quad g_2(\boldsymbol{x}_1) = -12, \quad g_3(\boldsymbol{x}_1) = -48, \quad g_4(\boldsymbol{x}_1) = 30 \qquad \text{(S.2.19)}$$

であるので，$\underset{i}{\text{argmax}}\,\{g_i(\boldsymbol{x})\} = 4$ となり，$\boldsymbol{x}_1 \in \omega_4$ と識別される．またパターン $\boldsymbol{x}_2 = (2, 11)^t$ に対しては

$$g_1(\boldsymbol{x}_2) = 48, \quad g_2(\boldsymbol{x}_2) = 0, \quad g_3(\boldsymbol{x}_2) = -12, \quad g_4(\boldsymbol{x}_2) = 38 \qquad \text{(S.2.20)}$$

であるので，$\underset{i}{\text{argmax}}\{g_i(\boldsymbol{x})\} = 1$ となり，$\boldsymbol{x}_2 \in \omega_1$ と識別される．図 S.2.4 にパターン $\boldsymbol{x}_1$ および $\boldsymbol{x}_2$ がプロットされており，それぞれ ω_4，ω_1 の領域に含まれることから，上記の結果が正しいことが確かめられる．

【演習問題 2.2】

問 (1)

パーセプトロンの学習規則（ 2.3 節 ［2］）に従って重みの学習を行った結果を**表 S.2.1** にまとめた．表の見方は，本文の表 2.1 と同じである．

表 S.2.1: パーセプトロンによる重みの学習過程

エポック	繰り返し	パターン	クラス	拡張特徴ベクトル			拡張重みベクトル			$g(\boldsymbol{x})$	正 ○ 誤 ×	新しい重み		
				x_0	x_1	x_2	w_0	w_1	w_2			w_0'	w_1'	w_2'
1	1	$\boldsymbol{x}_1$	ω_1	1	11	8	−54	13	−15	−31	×	−53	24	−7
	2	$\boldsymbol{x}_2$	ω_1	1	10	10	−53	24	−7	117	○			
	3	$\boldsymbol{x}_3$	ω_1	1	6	3	−53	24	−7	70	○			
	4	$\boldsymbol{x}_4$	ω_2	1	6	5	−53	24	−7	56	×	−54	18	−12
	5	$\boldsymbol{x}_5$	ω_2	1	2	8	−54	18	−12	−114	○			
	6	$\boldsymbol{x}_6$	ω_2	1	1	2	−54	18	−12	−60	○			
2	7	$\boldsymbol{x}_1$	ω_1	1	11	8	−54	18	−12	48	○			
	8	$\boldsymbol{x}_2$	ω_1	1	10	10	−54	18	−12	6	○			
	9	$\boldsymbol{x}_3$	ω_1	1	6	3	−54	18	−12	18	○			
	10	$\boldsymbol{x}_4$	ω_2	1	6	5	−54	18	−12	−6	○			

学習パターンは 6 パターンであるので，6 回の繰り返しが 1 エポックに相当する．本学習では，繰り返し 1 で，クラス ω_1 に属する $\boldsymbol{x}_1$ に対し $g(\boldsymbol{x}_1) = -31 < 0$ となるので誤識別となる．そこで式 (2.25) に従って重み $\mathbf{w}$ は

$$\mathbf{w}' = \mathbf{w} + \rho \cdot \mathbf{x} \tag{S.2.21}$$
$$= (-54,\ 13,\ -15) + 1 \cdot (1,\ 11,\ ,8) \tag{S.2.22}$$
$$= (-53,\ 24,\ -7) \tag{S.2.23}$$

と新しい重み $\mathbf{w}'$ に修正される．繰り返し 2, 3 ではパターンは正しく識別されるので，重みの修正は行われない．

繰り返し 4 で，今度はクラス ω_2 に属する $\boldsymbol{x}_4$ に対し $g(\boldsymbol{x}_4) = 56 > 0$ となるので誤識別となり，式 (2.26) に従って重みは次式のごとく修正される．

$$\begin{aligned} \mathbf{w}' &= \mathbf{w} - \rho \cdot \mathbf{x} && \text{(S.2.24)} \\ &= (-53,\ 24,\ -7) - 1 \cdot (1,\ 6,\ ,5) && \text{(S.2.25)} \\ &= (-54,\ 18,\ -12) && \text{(S.2.26)} \end{aligned}$$

本修正以降，$\boldsymbol{x}_5$, $\boldsymbol{x}_6$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, $\boldsymbol{x}_4$ と 6 パターン連続して正しく識別できるので，この時点で収束と判定できる．すなわち，2 エポック目の途中の，繰り返し 10 で収束する[*4]．最終的に得られた重みは

$$(w_0,\ w_1,\ w_2) = (-54,\ 18,\ -12) \qquad \text{(S.2.27)}$$

である．

問 (2)

重みの初期値

$$(w_0,\ w_1,\ w_2) = (-54,\ 13,\ -15) \qquad \text{(S.2.28)}$$

によって定まる決定境界は

$$g(\boldsymbol{x}) = -54 + 13x_1 - 15x_2 = 0 \qquad \text{(S.2.29)}$$

であり，**図 S.2.5** に鎖線で示されている．同図には，クラス ω_1 のパターンが黒丸で，クラス ω_2 のパターンが白丸でそれぞれ示されている．この決定境界では両クラスを正しく識別できないことがわかる．

問 (3)

同様にして，最終的に得られた重みによって定まる決定境界は

$$g(\boldsymbol{x}) = -54 + 18x_1 - 12x_2 = 0 \qquad \text{(S.2.30)}$$

[*4] 本文でも述べたように，パターン $\boldsymbol{x}_5, \boldsymbol{x}_6$ を識別するための繰り返し 11, 12 は不要である．

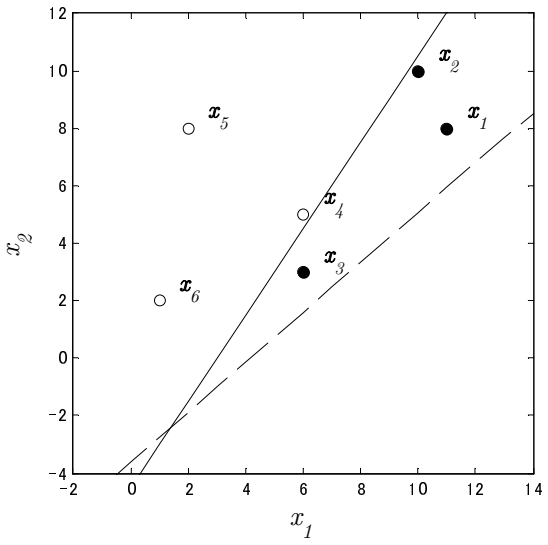


図 S.2.5: 初期設定の重み，および学習後の重みによる決定境界

と求められ，図 S.2.5 に実線で示されている．この決定境界は両クラスを正しく識別できている．

【演習問題 2.3】

問 (1)

図 S.2.6 に，クラス ω_1，ω_2 の学習パターンをそれぞれ黒丸，白丸で示し，最近傍決定則（識別法 1）によって定まる決定境界を実線で示した．この決定境界は，区分的線形識別関数によって設定されている．

問 (2)

クラス ω_1，ω_2 の平均ベクトルをそれぞれ $\mathbf{m}_1$，$\mathbf{m}_2$ とする．最小距離識別法（識別法 2）の決定境界上にある $\boldsymbol{x}$ は，$\mathbf{m}_1$，$\mathbf{m}_2$ から等距離にあるので，$\boldsymbol{x}$ は，$\mathbf{m}_1$ と $\mathbf{m}_2$ を結ぶ線分の中点を通り，$(\mathbf{m}_1 - \mathbf{m}_2)$ と直交する超平面上（2 次元特

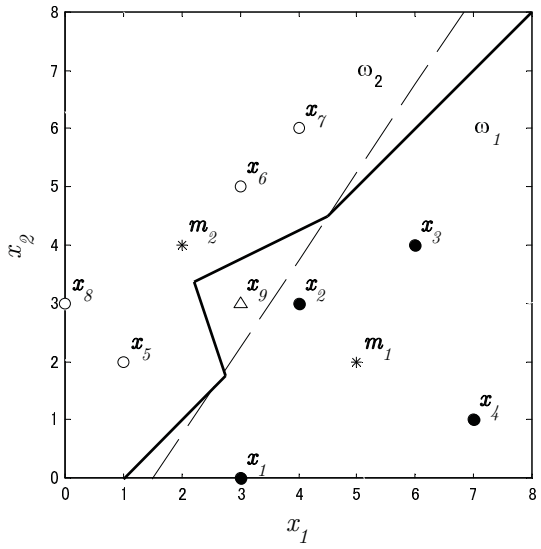


図 S.2.6: 区分的線形識別関数による決定境界

徴空間上では直線）になくてはならない．式で表すと

$$(\mathbf{m}_1 - \mathbf{m}_2)^t\big(\boldsymbol{x} - (\mathbf{m}_1 + \mathbf{m}_2)/2\big) = 0 \tag{S.2.31}$$

と書ける．ここで $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_8$ の座標値より，$\mathbf{m}_1$，$\mathbf{m}_2$ は

$$\mathbf{m}_1 = (5, 2)^t, \qquad \mathbf{m}_2 = (2, 4)^t \tag{S.2.32}$$

と求められ，それらが図中の*印で示されている．式 (S.2.31) に上式を代入することにより，$\boldsymbol{x}(x_1, x_2)$ は

$$6x_1 - 4x_2 - 9 = 0 \tag{S.2.33}$$

を満たす．上式が求めるべき決定境界であり，図の鎖線で示されている．

問 (3)

テストパターン $\boldsymbol{x}_9$ が図中の△印で示されている．これより $\boldsymbol{x}_9$ は，識別法 1 ではクラス ω_1 に，識別法 2 ではクラス ω_2 に識別されることがわかる．

第３章

【演習問題 3.1】

以下述べる証明には，凸計画問題と最適化に関する知識が必要となる．詳細は，文献 [石井 14] の付録 A.1 を参照されたい．

拡張特徴ベクトル $\mathbf{x}_p$ $(p = 1, \ldots, n)$ および拡張重みベクトル $\mathbf{w}_i$ $(i = 1, \ldots, c)$ は，ともに $(d+1)$ 次元のベクトルである．拡張特徴ベクトルは

$$\mathbf{x}_p = (x_{p0}, x_{p1}, \ldots, x_{pd})^t \qquad (\text{ただし } x_{p0} \equiv 1) \tag{S.3.1}$$

と書け，拡張重みベクトルは，式 (2.14) より

$$\mathbf{w}_i = (w_{i0}, w_{i1}, \ldots, w_{id})^t \tag{S.3.2}$$

と書ける．ただし，$\mathbf{x}_p$ の要素を x_{pj} $(j = 0, 1, \ldots, d)$ と記した．ここで式 (3.10) で示した評価関数 $J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c)$，および式 (3.15) 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．

$$J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c) = \frac{1}{2} \sum_{p=1}^{n} \sum_{i=1}^{c} (\mathbf{w}_i^t \mathbf{x}_p - b_{ip})^2 \tag{S.3.3}$$

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^t \tag{S.3.4}$$

ここで $\mathbf{w}_i$ に着目し，式 (S.3.3) の $J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c)$ を $\mathbf{w}_i$ の要素 w_{ik} で微分すると

$$\frac{\partial J}{\partial w_{ik}} = \sum_{p=1}^{n} \left({\mathbf{w}_i}^t \mathbf{x}_p - b_{ip} \right) x_{pk} \tag{S.3.5}$$

が得られる．上式をさらに $\mathbf{w}_i$ の要素 w_{il} で微分し，2 回微分の結果を h_{kl} と置くと

$$h_{kl} = \frac{\partial^2 J}{\partial w_{ik} \partial w_{il}} \tag{S.3.6}$$

$$= \sum_{p=1}^{n} x_{pk} \cdot x_{pl} \qquad (k, l = 0, 1, \ldots, d) \tag{S.3.7}$$

となる．ここで $\mathbf{H} = (h_{kl})$ なる行列を定義する．すなわち $\mathbf{H}$ は，(k, l) 成分が h_{kl} である $(d+1) \times (d+1)$ の行列である．行列 $\mathbf{H}$ を $J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c)$ の**ヘッ**

セ行列 (Hessian) という．式 (S.3.7) の右辺は，$(d+1)\times(d+1)$ の行列 $\mathbf{X}^t\mathbf{X}$ の (k,l) 成分であることがわかる．したがって下式が成り立つ．

$$\mathbf{H}=\mathbf{X}^t\mathbf{X} \tag{S.3.8}$$

ここで，$\mathbf{y}\neq\mathbf{0}$ なる任意の $(d+1)$ 次元列ベクトル $\mathbf{y}$ に対し，$\mathbf{y}^t\mathbf{H}\mathbf{y}$ を計算すると，式 (S.3.8) より

$$\begin{aligned}\mathbf{y}^t\mathbf{H}\mathbf{y}&=\mathbf{y}^t\mathbf{X}^t\mathbf{X}\mathbf{y}\\&=(\mathbf{X}\mathbf{y})^t\mathbf{X}\mathbf{y}\\&=\|\mathbf{X}\mathbf{y}\|^2>0\end{aligned} \tag{S.3.9}$$

が成り立つ．これはすべての $\mathbf{w}_i$ $(i=1,\ldots,c)$ に対して成り立つので，ヘッセ行列 $\mathbf{H}$ は**正定値** (positive definite) である．したがって，評価関数 $J(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_c)$ は**狭義凸関数** (strictly convex function) である．

以上より，評価関数 $J(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_c)$ を最小化する問題は**凸計画問題** (convex programming problem) であり，$J(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_c)$ は狭義凸関数であるので，式 (3.20) 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．

【演習問題 3.2】

本問の識別対象は 2 クラスであるので，式 (3.40) により，重みベクトル $\mathbf{w}$ を一つ求めればよい．しかし，ここではあえて式 (3.20) を用い，二つの重みベクトル $\mathbf{w}_1$，$\mathbf{w}_2$ の算出を試みる．式 (3.15) および式 (3.16) より下式が得られる．

$$\mathbf{X}=\begin{pmatrix}1&0&5\\1&1&1\\1&5&0\\1&6&2\\1&2&6\\1&2&2\end{pmatrix}\qquad\mathbf{X}^t\mathbf{X}=\begin{pmatrix}6&16&16\\16&70&29\\16&29&70\end{pmatrix} \tag{S.3.10}$$

$$\mathbf{b}_1=(1,1,1,0,0,0)^t \tag{S.3.11}$$

$$\mathbf{b}_2=(0,0,0,1,1,1)^t \tag{S.3.12}$$

逆行列を求めると

$$(\mathbf{X}^t\mathbf{X})^{-1}=\begin{pmatrix}1.207&-0.195&-0.195\\-0.195&0.049&0.024\\-0.195&0.024&0.049\end{pmatrix} \tag{S.3.13}$$

となるので

$$
\begin{aligned}
&(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \\
&= \begin{pmatrix} 0.232 & 0.817 & 0.232 & -0.354 & -0.354 & 0.427 \\ -0.073 & -0.122 & 0.049 & 0.146 & 0.049 & -0.049 \\ 0.049 & -0.122 & -0.073 & 0.049 & 0.146 & -0.049 \end{pmatrix}
\end{aligned} \tag{S.3.14}
$$

と計算できる．したがって，式 (3.20) より

$$
\begin{aligned}
\mathbf{w}_1 &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{b}_1 = (\ \ 1.281,\ \ -0.146,\ \ -0.146)^t \\
\mathbf{w}_2 &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{b}_2 = (\ -0.281,\ \ 0.146,\ \ 0.146)^t
\end{aligned} \tag{S.3.15}
$$

と求められ，これより下式が得られる．

$$
g_1(\boldsymbol{x}) = \mathbf{w}_1^t\mathbf{x} = \ \ 1.281 - 0.146\,x_1 - 0.146\,x_2 \tag{S.3.16}
$$

$$
g_2(\boldsymbol{x}) = \mathbf{w}_2^t\mathbf{x} = -0.281 + 0.146\,x_1 + 0.146\,x_2 \tag{S.3.17}
$$

したがって，決定境界は

$$
g_{12}(\boldsymbol{x}) = g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) \tag{S.3.18}
$$

$$
= 1.562 - 0.292\,x_1 - 0.292\,x_2 = 0 \tag{S.3.19}
$$

と求められ，これは

$$
x_1 + x_2 = 5.35 \tag{S.3.20}
$$

なる直線であり，図 3.11 の点線で示された決定境界を表している．図より，与えられた 6 個のパターンは線形分離不可能であり，線形識別関数では 2 クラスを完全に分離することはできない．この決定境界を適用した場合には，クラス ω_2 に属するパターン $\boldsymbol{x}_6$ がクラス ω_1 と誤識別されることがわかる．

【演習問題 3.3】

ここでは前問とは異なり，2 クラス問題として，式 (3.40) により重みベクトル $\mathbf{w}$ を求めることにする．閉じた形の解を求めるために必要なパターン行列 $\mathbf{X}$ は式 (3.43) と同じであり，$(\mathbf{X}^t\mathbf{X})^{-1}\,\mathbf{X}^t$ も式 (3.45) と同じになる．一方 $\mathbf{b}$ は

$$
\begin{aligned}
\mathbf{b} &= (b_1,\, b_2,\, \ldots,\, b_n)^t \\
&= (1,\, 1,\, 1,\, -1,\, -1,\, -1)^t
\end{aligned} \tag{S.3.21}
$$

である．これらより

$$\begin{aligned}\mathbf{w} = (w_0,\, w_1) &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{b} \\ &= (0.281,\, 0.938)^t \end{aligned} \tag{S.3.22}$$

が得られる．したがって，決定境界 x は

$$x = -w_0/w_1 = -0.281/0.938 = -0.300 \tag{S.3.23}$$

と求められる．この決定境界は式 (3.49) と全く同じであり，二つのクラスを正しく分離できていることがわかる．

次に，式 (3.37) によって求めた $J(\mathbf{w})$ を**図 S.3.1** に示す．図 3.2 と同様，$J(\mathbf{w})$ の等高線が細線で示され，閉じた形の解 $(w_1,\, w_0) = (0.938,\, 0.281)$ が小さな黒丸で示されている．

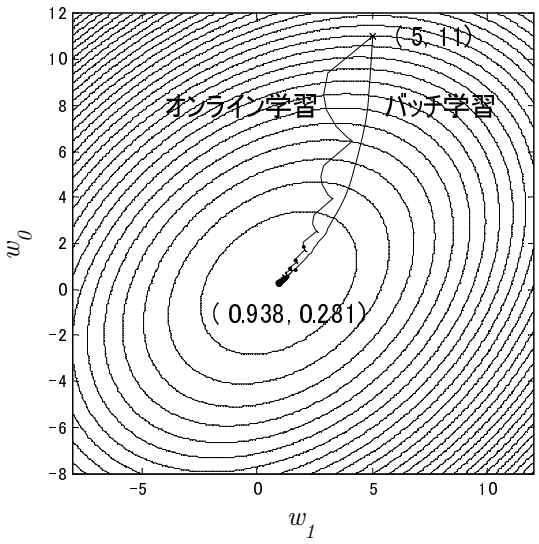


図 S.3.1: ウィドロー・ホフの学習規則（線形分離可能なデータ）

次にウィドロー・ホフの学習規則を適用した結果を示す．重みベクトル $\mathbf{w}$ の初期値を $(w_1,\, w_0) = (5,\, 11)$ に設定し，その他の実験条件は，線形分離不可能

なデータに対する実験と同一とした．実験は，バッチ学習，オンライン学習の双方について行い，収束に至るまでの重みベクトル $\mathbf{w}$ の軌跡を図 S.3.1 に示した．また，収束に至るまでの $J(\mathbf{w})$ の値の変化を**図 S.3.2** に示した．

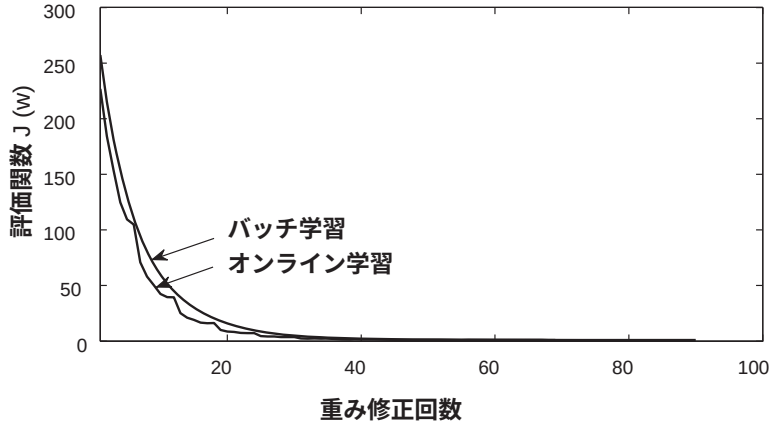


図 S.3.2: ウィドロー・ホフの学習規則の収束過程（線形分離可能なデータ）

以上の結果をみると，$J(\mathbf{w})$ の形状，最適な $\mathbf{w}$，決定境界 x はいずれも線形分離不可能なデータを用いたときの結果とほぼ同じである．すなわち，本実験で用いたデータに関する限り，二乗誤差を最小にするという評価では，両者の結果に大きな差は見られないことがわかる．

【演習問題 3.4】

本文の 2.3 節 ［1］で述べたように，重みベクトル $\mathbf{w} = (w_0, w_1, \ldots, w_d)$ は，$(d+1)$ 次元の重み空間上の 1 点を表している．また同時に $\mathbf{w}$ は，原点 O を始点とし，上記の点を終点とする有向線分をも表している．以下では，$\mathbf{w}$ と表記することで，これらの両方を表すことにする．また，特徴ベクトル $\mathbf{x}$ についても同様に扱う．

本問題を 2 次元重み空間を例として示したのが**図 S.3.3** である．入力パターン $\boldsymbol{x}$ に対して，超平面（この図では直線）$g(\boldsymbol{x}) = \mathbf{w}^t\mathbf{x} = 0$ が定まる．図で示す

ように，超平面は原点 O を通り，$\mathbf{x}$ を法線方向として持つ．この超平面上に存在する重みベクトル $\mathbf{w}$ はすべて $\mathbf{w}^t\mathbf{x} = 0$ を満たす．

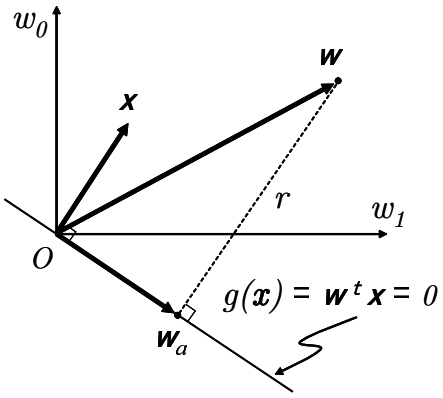


図 S.3.3: 重み空間における超平面と重みベクトル

ここで，重み空間上の任意の重みベクトル $\mathbf{w}$ を，超平面の法線方向に射影して得られるベクトルを $\mathbf{w}_a$ とする．求めるべき r は，$\mathbf{w}_a$（の終点）と $\mathbf{w}$（の終点）の距離である．したがって，$\mathbf{w}$ は法線ベクトル $\mathbf{x}$ を用いて

$$\mathbf{w} = \mathbf{w}_a \pm r\,\frac{\mathbf{x}}{||\mathbf{x}||} \tag{S.3.24}$$

と書ける．上式の複号 $\pm$ は，超平面を介して $\mathbf{w}$ が $\mathbf{x}$ と同じ側にあるときは $+$ を，反対側にあるときは $-$ を取ることを示している．また，$\mathbf{w}_a$ は超平面上にあるので，下式が成り立つ．

$$\mathbf{w}_a^t\mathbf{x} = 0 \tag{S.3.25}$$

式 (S.3.24) を転置して右から $\mathbf{x}$ を乗じ，続いて式 (S.3.25) を用いると

$$\mathbf{w}^t\mathbf{x} = \mathbf{w}_a^t\mathbf{x} \pm r\,\frac{\mathbf{x}^t\mathbf{x}}{||\mathbf{x}||} \tag{S.3.26}$$

$$= \pm r\,||\mathbf{x}|| \tag{S.3.27}$$

が得られる．これより

$$|\mathbf{w}^t\mathbf{x}| = r\,||\mathbf{x}|| \tag{S.3.28}$$

となり，最終的に

$$r = \frac{|\mathbf{w}^t\mathbf{x}|}{||\mathbf{x}||} \tag{S.3.29}$$

が得られる．

第4章

【演習問題 4.1】

式 (4.21) を再掲すると

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathcal{X}; \boldsymbol{\theta}) = \sum_{k=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{x}_k; \boldsymbol{\theta}) = \mathbf{0} \tag{S.4.1}$$

である．ここでパターン $\boldsymbol{x}_k$ は，平均ベクトル $\mathbf{m}$，共分散行列 $\boldsymbol{\Sigma}$ の正規分布に従うとすると，式 (4.8) より

$$\begin{aligned} p(\boldsymbol{x}_k; \boldsymbol{\theta}) &= p(\boldsymbol{x}_k; \mathbf{m}, \boldsymbol{\Sigma}) \\ &= \frac{1}{(2\pi)^{d/2} \left|\boldsymbol{\Sigma}\right|^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_k - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_k - \mathbf{m}) \right\} \end{aligned} \tag{S.4.2}$$

と書ける．したがって，対数をとることにより

$$\begin{aligned} &\log p(\boldsymbol{x}_k; \mathbf{m}, \boldsymbol{\Sigma}) \\ &= -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\boldsymbol{x}_k - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_k - \mathbf{m}) \end{aligned} \tag{S.4.3}$$

となる．

まず，$\mathbf{m}$ の推定を行う．共分散行列 $\boldsymbol{\Sigma}$ は対称行列であるので $\boldsymbol{\Sigma}^{-1}$ も対称行列となる．したがって，上式を $\mathbf{m}$ で微分すると，**付録 A.2** の式 (A.2.5) より

$$\frac{\partial}{\partial \mathbf{m}} \log p(\boldsymbol{x}_k; \mathbf{m}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_k - \mathbf{m}) \tag{S.4.4}$$

が得られる．上式を式 (S.4.1) に代入し，左から $\boldsymbol{\Sigma}$ を乗ずることにより

$$\sum_{k=1}^{n} (\boldsymbol{x}_k - \mathbf{m}) = \mathbf{0} \tag{S.4.5}$$

が得られ，これより $\mathbf{m}$ の推定値 $\hat{\mathbf{m}}$ は

$$\hat{\mathbf{m}} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{x}_k \tag{S.4.6}$$

となり，式 (4.22) が導出できた．

次に $\boldsymbol{\Sigma}$ の推定を行う．**付録 A.2** の式 (A.2.12) より

$$\left|\boldsymbol{\Sigma}^{-1}\right|^{-1} = |\boldsymbol{\Sigma}| \tag{S.4.7}$$

が成り立ち，式 (A.2.11) より

$$(\boldsymbol{x}_k - \mathbf{m})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_k - \mathbf{m}) = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_k - \mathbf{m})(\boldsymbol{x}_k - \mathbf{m})^t\right) \tag{S.4.8}$$

が成り立つので，式 (S.4.3) は次のように $\boldsymbol{\Sigma}^{-1}$ の関数として書き換えられる．

$$\begin{aligned}&\log p(\boldsymbol{x}_k; \mathbf{m}, \boldsymbol{\Sigma}) \\ &= -\frac{d}{2}\log(2\pi) + \frac{1}{2}\log\left|\boldsymbol{\Sigma}^{-1}\right| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_k - \mathbf{m})(\boldsymbol{x}_k - \mathbf{m})^t\right)\end{aligned} \tag{S.4.9}$$

ここで，最尤推定により $\boldsymbol{\Sigma}$ を求めることは，最尤推定により $\boldsymbol{\Sigma}^{-1}$ を求めることと等価である．そこで式 (S.4.9) を $\boldsymbol{\Sigma}^{-1}$ で微分すると，式 (A.2.10)，式 (A.2.7) を用いて

$$\frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} \log p(\boldsymbol{x}_k; \mathbf{m}, \boldsymbol{\Sigma}) = \frac{1}{2}\boldsymbol{\Sigma} - \frac{1}{2}(\boldsymbol{x}_k - \mathbf{m})(\boldsymbol{x}_k - \mathbf{m})^t \tag{S.4.10}$$

となる．上式を式 (S.4.1) に代入すると

$$\sum_{k=1}^{n} \left(\boldsymbol{\Sigma} - (\boldsymbol{x}_k - \mathbf{m})(\boldsymbol{x}_k - \mathbf{m})^t\right) = \mathbf{0} \tag{S.4.11}$$

が得られる．上式と式 (S.4.6) より，$\boldsymbol{\Sigma}$ の推定値 $\hat{\boldsymbol{\Sigma}}$ は

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n} (\boldsymbol{x}_k - \hat{\mathbf{m}})(\boldsymbol{x}_k - \hat{\mathbf{m}})^t \tag{S.4.12}$$

となり，式 (4.23) が導出できた．

【演習問題 4.2】

まず，最適な $\boldsymbol{w}$ を求めるため，式 (4.44) と同様の計算を行うと

$$\frac{\partial J}{\partial \tilde{\sigma}_i^2} = -k_i \cdot \frac{(k_1\tilde{m}_1^2 + k_2\tilde{m}_2^2)}{(k_1\tilde{\sigma}_1^2 + k_2\tilde{\sigma}_2^2)^2} \qquad (i = 1, 2) \tag{S.4.13}$$

が得られ，式 (4.42) より

$$s = \frac{k_1}{k_1 + k_2}, \qquad 1 - s = \frac{k_2}{k_1 + k_2}$$

が得られる．したがって，式 (4.41) より，最適な $\boldsymbol{w}$ は

$$\boldsymbol{w} = a \cdot (k_1\boldsymbol{\Sigma}_1 + k_2\boldsymbol{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{S.4.14}$$

となる．ただし，a は任意の定数である．

次に最適な w_0 を求めるため

$$\frac{\partial J}{\partial \tilde{m}_i} = \frac{2k_i\tilde{m}_i}{k_1\tilde{\sigma}_1^2 + k_2\tilde{\sigma}_2^2} \qquad (i = 1, 2) \tag{S.4.15}$$

を式 (4.39) に代入して

$$k_1\tilde{m}_1 + k_2\tilde{m}_2 = 0 \tag{S.4.16}$$

となる．式 (4.30) より

$$\tilde{m}_i = \boldsymbol{w}^t\mathbf{m}_i + w_0 \qquad (i = 1, 2) \tag{S.4.17}$$

であるので，式 (S.4.16) より

$$\boldsymbol{w}^t(k_1\mathbf{m}_1 + k_2\mathbf{m}_2) + (k_1 + k_2)w_0 = 0 \tag{S.4.18}$$

が得られる．したがって，すでに求められた式 (S.4.14) の $\boldsymbol{w}$ を上式に代入することにより，最適な w_0 は

$$\begin{aligned} w_0 &= -\frac{\boldsymbol{w}^t(k_1\mathbf{m}_1 + k_2\mathbf{m}_2)}{k_1 + k_2} \\ &= -\frac{(k_1\mathbf{m}_1^t + k_2\mathbf{m}_2^t)\boldsymbol{w}}{k_1 + k_2} && \text{(S.4.19)} \\ &= -a \cdot \frac{(k_1\mathbf{m}_1^t + k_2\mathbf{m}_2^t)(k_1\boldsymbol{\Sigma}_1 + k_2\boldsymbol{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)}{k_1 + k_2} && \text{(S.4.20)} \end{aligned}$$

となる．ただし，a は，式 (S.4.14) の a と同値の定数である．

【演習問題 4.3 *】

第 4 章では評価関数の最大化を目指したが，ここで定義された評価関数 J は二乗誤差を表しているので，最小化を目指すことになる．最大化すべき評価関数は負号を付けて $-J$ とすればよいので，最大化，最小化に本質的な違いはなく，4.3 節 ［1］で述べた手法はそのまま適用できる．

ここで，最小化すべき評価関数として $J_0(\mathbf{w}) \stackrel{\text{def}}{=} (2/n)J(\mathbf{w})$ を定義すると

$$
\begin{aligned}
J_0(\mathbf{w}) &\stackrel{\text{def}}{=} (2/n)J(\mathbf{w}) && \text{(S.4.21)}\\
&= \frac{1}{n}\sum_{p=1}^{n}(g(\boldsymbol{x}_p) - b_p)^2 \\
&= \mathrm{E}\left\{(g(\boldsymbol{x}_p) - b_p)^2\right\} \\
&= \mathrm{E}\left\{g(\boldsymbol{x}_p)^2\right\} - 2\,\mathrm{E}\left\{b_p g(\boldsymbol{x}_p)\right\} + \mathrm{const} && \text{(S.4.22)}
\end{aligned}
$$

となる．ただし，const は $\mathbf{w}$ に依存しない定数である．式 (4.29) より

$$
\tilde{m}_i = \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_i}\{g(\boldsymbol{x}_p)\} \qquad (i = 1, 2) \qquad \text{(S.4.23)}
$$

であり，式 (4.31)，(S.4.23) より

$$
\begin{aligned}
\tilde{\sigma}_i^2 &= \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_i}\{(g(\boldsymbol{x}_p) - \tilde{m}_i)^2\} && \text{(S.4.24)}\\
&= \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_i}\{g(\boldsymbol{x}_p)^2\} - 2\tilde{m}_i \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_i}\{g(\boldsymbol{x}_p)\} + \tilde{m}_i^2 \\
&= \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_i}\{g(\boldsymbol{x}_p)^2\} - \tilde{m}_i^2 \qquad (i = 1, 2) && \text{(S.4.25)}
\end{aligned}
$$

となる．式 (S.4.25) より，式 (S.4.22) 右辺の第 1 項は

$$
\begin{aligned}
\mathrm{E}\left\{g(\boldsymbol{x}_p)^2\right\} &= P(\omega_1) \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_1}\{g(\boldsymbol{x}_p)^2\} + P(\omega_2) \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_2}\{g(\boldsymbol{x}_p)^2\} && \text{(S.4.26)}\\
&= P(\omega_1)\cdot(\tilde{\sigma}_1^2 + \tilde{m}_1^2) + P(\omega_2)\cdot(\tilde{\sigma}_2^2 + \tilde{m}_2^2) && \text{(S.4.27)}
\end{aligned}
$$

と書け，式 (S.4.23)，および教師信号 b_p の定義式 (3.35) より，式 (S.4.22) 右辺の第 2 項は

$$
\begin{aligned}
\mathrm{E}\left\{b_p g(\boldsymbol{x}_p)\right\} &= P(\omega_1) \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_1}\{g(\boldsymbol{x}_p)\} + P(\omega_2) \mathop{\mathrm{E}}_{\boldsymbol{x}_p \in \omega_2}\{-g(\boldsymbol{x}_p)\} && \text{(S.4.28)}\\
&= P(\omega_1)\tilde{m}_1 - P(\omega_2)\tilde{m}_2 && \text{(S.4.29)}
\end{aligned}
$$

と書ける．式 (S.4.27)，(S.4.29) を式 (S.4.22) に代入し

$$
\begin{aligned}
J_0(\mathbf{w}) &= (2/n)J(\mathbf{w}) \\
&= P(\omega_1)\cdot(\tilde{\sigma}_1^2 + \tilde{m}_1^2) + P(\omega_2)\cdot(\tilde{\sigma}_2^2 + \tilde{m}_2^2) \\
&\qquad -2(P(\omega_1)\tilde{m}_1 - P(\omega_2)\tilde{m}_2) + \mathrm{const} && \text{(S.4.30)}
\end{aligned}
$$

が得られる．すなわち，$J_0(\mathbf{w})$ は $\tilde{m}_1$，$\tilde{m}_2$，$\tilde{\sigma}_1^2$，$\tilde{\sigma}_2^2$ の関数となるので， 4.3 節［1］の手法が使える．

まず，最適な $\boldsymbol{w}$ を求めるには，式 (4.42) に従い

$$\frac{\partial J_0}{\partial \tilde{\sigma}_1^2} = P(\omega_1) \tag{S.4.31}$$

$$\frac{\partial J_0}{\partial \tilde{\sigma}_2^2} = P(\omega_2) \tag{S.4.32}$$

より

$$s = \frac{P(\omega_1)}{P(\omega_1) + P(\omega_2)} = P(\omega_1), \qquad 1 - s = P(\omega_2)$$

となるので，式 (4.41) より，最適な $\boldsymbol{w}$ として

$$\begin{aligned} \boldsymbol{w} &= a \cdot (P(\omega_1)\boldsymbol{\Sigma}_1 + P(\omega_2)\boldsymbol{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \\ &= a \cdot {\boldsymbol{\Sigma}_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \end{aligned} \tag{S.4.33}$$

が得られる．上式では，式 (6.113) を用いた．また，a は任意の定数である．

次に最適な w_0 を求めるため

$$\frac{\partial J_0}{\partial \tilde{m}_1} = 2P(\omega_1)\tilde{m}_1 - 2P(\omega_1) \tag{S.4.34}$$

$$\frac{\partial J_0}{\partial \tilde{m}_2} = 2P(\omega_2)\tilde{m}_2 + 2P(\omega_2) \tag{S.4.35}$$

を式 (4.39) に代入して

$$P(\omega_1)\tilde{m}_1 + P(\omega_2)\tilde{m}_2 - P(\omega_1) + P(\omega_2) = 0 \tag{S.4.36}$$

が得られる．式 (4.30) より

$$\tilde{m}_i = \boldsymbol{w}^t \mathbf{m}_i + w_0 \qquad (i = 1, 2) \tag{S.4.37}$$

であるので，上式を式 (S.4.36) に代入すると

$$\begin{aligned} &\boldsymbol{w}^t(P(\omega_1)\,\mathbf{m}_1 + P(\omega_2)\,\mathbf{m}_2) + (P(\omega_1) + P(\omega_2))w_0 \\ &\quad -P(\omega_1) + P(\omega_2) = 0 \end{aligned} \tag{S.4.38}$$

となり，さらに

$$\mathbf{m} = P(\omega_1)\,\mathbf{m}_1 + P(\omega_2)\,\mathbf{m}_2 \tag{S.4.39}$$

$$P(\omega_1) + P(\omega_2) = 1 \tag{S.4.40}$$

を用いると，式 (S.4.38) は

$$\boldsymbol{w}^t\mathbf{m} + w_0 - P(\omega_1) + P(\omega_2) = 0 \tag{S.4.41}$$

と書き換えられる．したがって，すでに求められた式 (S.4.33) の $\boldsymbol{w}$ を上式に代入することにより，最適な w_0 は

$$w_0 = -\mathbf{m}^t\boldsymbol{w} + P(\omega_1) - P(\omega_2) \tag{S.4.42}$$

$$= -a \cdot \mathbf{m}^t{\boldsymbol{\Sigma}_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2) + P(\omega_1) - P(\omega_2) \tag{S.4.43}$$

となる．ここで a は，式 (S.4.14) の a と同値の定数である．

なお，式 (S.4.33)，(S.4.43) の $\boldsymbol{\Sigma}_W$ の代わりに $\boldsymbol{\Sigma}_T$ を用いて $\boldsymbol{w}$，w_0 を表したのが式 (9.21)，(9.22) であり，これらは式 (S.4.33)，(S.4.43) と一致する．詳細は**演習問題 9.2** の解を参照されたい．

【演習問題 4.4】

得られた 6 種の線形識別関数に，パターン $\boldsymbol{x}_1 = (2, 9)^t$ を入力した結果を**表 S.4.1** に示す．

表 S.4.1: 多数決法によるパターン $\boldsymbol{x}_1$ の得票結果

識別関数	値	符号	ω_1	ω_2	ω_3	ω_4
$g_{12}(\boldsymbol{x}_1)$	14	+	○			
$g_{13}(\boldsymbol{x}_1)$	32	+	○			
$g_{14}(\boldsymbol{x}_1)$	−7	−				○
$g_{23}(\boldsymbol{x}_1)$	3	+		○		
$g_{24}(\boldsymbol{x}_1)$	−21	−				○
$g_{34}(\boldsymbol{x}_1)$	−39	−				○
得票数			2	1	0	3

各線形識別関数の値と符号が示され，さらに各クラスがどの識別関数に対して得票したかを○印で示した．最終行には式 (4.53) で示したクラス毎の得票数 $N(i)$ $(i = 1, \ldots, 4)$ が得票結果として記されている．式 (4.53) を適用すると，$\underset{i}{\mathrm{argmax}} \{N(i)\} = 4$ となるので，$\boldsymbol{x}_1 \in \omega_4$ と識別される．

同様にして，パターン $\boldsymbol{x}_2 = (2, 11)^t$ に対する得票結果を**表 S.4.2** に示す．

表 S.4.2: 多数決法によるパターン $\boldsymbol{x}_2$ の得票結果

識別関数	値	符号	ω_1	ω_2	ω_3	ω_4
$g_{12}(\boldsymbol{x}_2)$	24	+	○			
$g_{13}(\boldsymbol{x}_2)$	30	+	○			
$g_{14}(\boldsymbol{x}_2)$	5	+	○			
$g_{23}(\boldsymbol{x}_2)$	1	+		○		
$g_{24}(\boldsymbol{x}_2)$	−19	−				○
$g_{34}(\boldsymbol{x}_2)$	−25	−				○
得票数			3	1	0	2

パターン $\boldsymbol{x}_2$ に対しては，$\underset{i}{\mathrm{argmax}} \{N(i)\} = 1$ となるので，$\boldsymbol{x}_2 \in \omega_1$ と識別される．

本演習問題では，6 種の線形識別関数を用いた多数決法によりパターンの識別を行った．一方，**演習問題 2.1** では，クラス毎に得られる 4 種の線形識別関数により，同じパターンの識別を行った．この 4 種の線形識別関数から得られる決定境界は，まさしく本演習問題の 6 種の線形識別関数と同一である．

クラス毎に線形識別関数が得られる場合には，多数決法を導入するまでもない．一方，本演習問題のように，すべてのクラス対 $\omega_i - \omega_j$ に対して識別関数 $g_{ij}(\boldsymbol{x})$ が得られている場合には多数決法が力を発揮する．たとえば，フィッシャの方法と多数決法を組み合わせるという手法は有効である．

【演習問題 4.5*】

式 (4.74) の $L(n,d)$ は，d 次元空間上で n 個のパターンを $(d-1)$ 次元超平面で ω_1，ω_2 の区別なく単純に二分する方法の数の 2 倍である．なぜなら，このような二分法の各々に対して ω_1, ω_2 を入れ替えた割付方法が存在するからである．例えば 2 次元特徴空間上に分布する 4 パターン $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4$ が一般位置にある場合は，**図 S.4.1** に示したように超平面 l_1〜l_7 による分割方法が可能であり

$$L(4,2) = 2 \cdot 7 = 14 \tag{S.4.44}$$

となる．

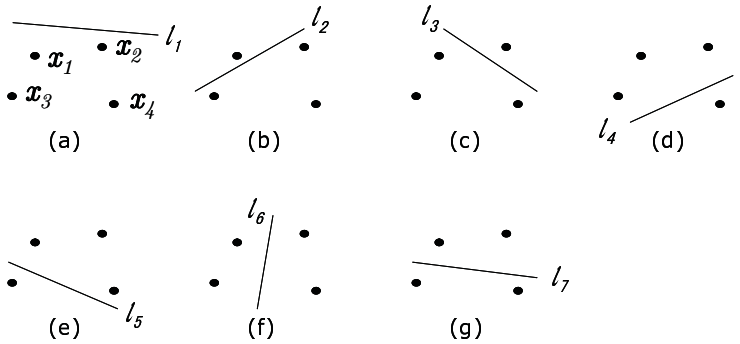


図 S.4.1: 可能な線形二分法

合計 $(n-1)$ 個のパターン $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n-1}$ からなる集合 $\mathcal{X}^{(n-1)}$ を考えよう．これらは d 次元特徴空間上で一般位置にあるとする．パターン集合 $\mathcal{X}^{(n-1)}$ の線形二分法の数は $L(n-1,d)$ である．この集合にさらに n 番目のパターン $\boldsymbol{x}_n$ が加わった集合を $\mathcal{X}^{(n)}$ とする．ただし，$\boldsymbol{x}_n$ は $\mathcal{X}^{(n)}$ 内で一般位置を占めるように選ぶ．ここで，$\boldsymbol{x}_n$ を通る超平面を設定し，集合 $\mathcal{X}^{(n-1)}$ の線形二分法を次の 2 種に分けて考える．

(1) $\boldsymbol{x}_n$ を通る超平面で可能な二分法

(2) $\boldsymbol{x}_n$ を通る超平面では不可能な二分法

図 S.4.1 の 4 パターンに $\boldsymbol{x}_5$ が加わった例が，**図 S.4.2** に示されている．両図

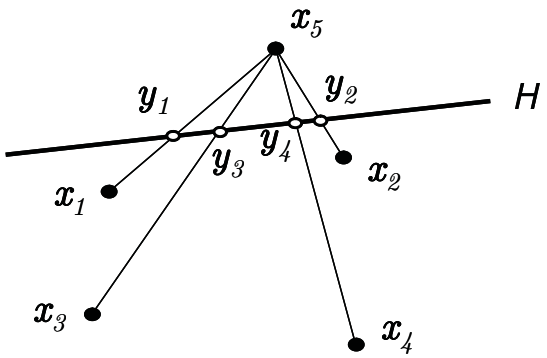


図 S.4.2: パターン $\boldsymbol{x}_5$ が加わったときの線形二分法

より明らかなように，図 S.4.1 の超平面 l_1, l_2, l_3, l_6 は，図 S.4.2 の $\boldsymbol{x}_5$ を通る超平面で実現できる．しかし，超平面 l_4, l_5, l_7 は $\boldsymbol{x}_5$ を通る超平面では実現できない．

上記 (1) の数を $L_{\boldsymbol{x}_n}(n-1,d)$ で表すと，(2) の数は，$L(n-1,d)-L_{\boldsymbol{x}_n}(n-1,d)$ であり，(2) はすべて $\mathcal{X}^{(n)}$ の線形二分法となっている．一方，(1) の超平面は，それを少し移動させることにより，$\boldsymbol{x}_n$ を ω_1, ω_2 のいずれに含めることもできる．したがって，その線形二分法の数は $2 \cdot L_{\boldsymbol{x}_n}(n-1,d)$ である．集合 $\mathcal{X}^{(n)}$ の線形二分法の数は，これらで完全に網羅されるので

$$\begin{aligned} L(n,d) &= L(n-1,d) - L_{\boldsymbol{x}_n}(n-1,d) + 2 \cdot L_{\boldsymbol{x}_n}(n-1,d) \\ &= L(n-1,d) + L_{\boldsymbol{x}_n}(n-1,d) \end{aligned} \tag{S.4.45}$$

となる．

次に $L_{\boldsymbol{x}_n}(n-1,d)$ を求めてみよう．点 $\boldsymbol{x}_n$ を通り，かつ $\mathcal{X}^{(n-1)}$ 内の各点を通る直線（1 次元超平面）を引くと，合計 $(n-1)$ 本の直線が得られる．集合 $\mathcal{X}^{(n)}$ 内の各点は一般位置にあるので，$(d-1)$ 次元超平面 H を適切に設定すれば，H はこれらの直線と $(n-1)$ 個の交点を持つことができる．これらの交点を $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-1}$ とする．図 S.4.2 の例では，$\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4$ が白丸で示されている．点 $\boldsymbol{x}_n$ を通る $(d-1)$ 次元超平面による $\mathcal{X}^{(n-1)}$ の線形二分法の数 $L_{\boldsymbol{x}_n}(n-1,d)$ は，H における $(d-2)$ 次元超平面による $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-1}$ の線

形二分法の数 $L(n-1, d-1)$ に等しい．すなわち

$$L_{\boldsymbol{x}_n}(n-1, d) = L(n-1, d-1) \tag{S.4.46}$$

が成り立つ．上式と式 (S.4.45) より

$$L(n, d) = L(n-1, d) + L(n-1, d-1) \tag{S.4.47}$$

が得られる．

【演習問題 4.6】

式 (4.75) が成り立つことを具体例で確かめるため，$n = 4$，$d = 2$ としてみる．すると，式 (4.75) の上式より

$$\begin{aligned} L(4, 2) &= 2 \cdot \sum_{j=0}^{2} {}_3C_j \qquad \text{(S.4.48)} \\ &= 2\left({}_3C_0 + {}_3C_1 + {}_3C_2\right) \\ &= 2(1 + 3 + 3) \\ &= 14 \qquad \text{(S.4.49)} \end{aligned}$$

が得られ，式 (S.4.44) と一致する．以下では，一般的な場合について式 (4.75) が成り立つことを示す．

式 (4.75) において

$$L(1, d) = 2 \tag{S.4.50}$$

は明らかであるので，$n = 1$ のとき式 (4.75) は成り立つ．

次に，$n = k$ のとき式 (4.75) が成り立つと仮定すると

$$L(k, d) = \begin{cases} 2 \cdot \displaystyle\sum_{j=0}^{d} {}_{k-1}C_j & (k > d) \qquad (a) \\ 2^k & (k \leq d) \qquad (b) \end{cases} \tag{S.4.51}$$

が得られる．式 (4.74) は，すべての n に対して成り立つので，$n = k + 1$ として

$$L(k+1, d) = L(k, d) + L(k, d-1) \tag{S.4.52}$$

が得られる．ここで，k の値を 2 通りの場合に分けて扱う．

(1) $k > d-1$ のとき

この場合は，さらに $k = d$ と $k > d$ の場合に分けて考える．

(i) $k = d$ のとき

この場合，$L(k,d)$ に対しては，式 (S.4.51) (b) の条件が当てはまるので

$$L(k,d) = 2^k \tag{S.4.53}$$

となる．また，$L(k,d-1)$ に対しては，式 (S.4.51) (a) の条件が当てはまるので

$$\begin{aligned} L(k,d-1) &= 2 \cdot \sum_{j=0}^{d-1} {}_{k-1}C_j && \text{(S.4.54)} \\ &= 2 \cdot \sum_{j=0}^{k-1} {}_{k-1}C_j && \text{(S.4.55)} \\ &= 2 \cdot 2^{k-1} && \text{(S.4.56)} \\ &= 2^k && \text{(S.4.57)} \end{aligned}$$

が成り立つ．なお，式 (S.4.55) から式 (S.4.56) への変形には

$$\sum_{j=0}^{k} {}_kC_j = 2^k \tag{S.4.58}$$

を用いた[*5]．

式 (S.4.53) と式 (S.4.57) を式 (S.4.52) に代入することにより，

$$L(k+1,d) = 2^k + 2^k = 2^{k+1} \tag{S.4.59}$$

[*5] 式 (S.4.58) は，二項定理

$$(x+y)^k = \sum_{j=0}^{k} {}_kC_j x^{k-j} y^j$$

において，$x = y = 1$ と置くことによって得られる．

が得られる．

(ii) $k > d$ のとき

この場合は，$L(k,d)$ および $L(k,d-1)$ の双方に対して式 (S.4.51) (a) の条件が当てはまるので

$$L(k,d) = 2 \cdot \sum_{j=0}^{d} {}_{k-1}C_j \tag{S.4.60}$$

$$L(k,d-1) = 2 \cdot \sum_{j=0}^{d-1} {}_{k-1}C_j \tag{S.4.61}$$

となる．これらを式 (S.4.52) に代入することにより，下式が得られる．

$$L(k+1,d) = 2 \left(\sum_{j=0}^{d} {}_{k-1}C_j + \sum_{j=0}^{d-1} {}_{k-1}C_j \right) \tag{S.4.62}$$

$$= 2 \left(1 + \sum_{j=1}^{d} {}_{k-1}C_j + \sum_{j=1}^{d} {}_{k-1}C_{j-1} \right) \tag{S.4.63}$$

$$= 2 \left(1 + \sum_{j=1}^{d} \left({}_{k-1}C_j + {}_{k-1}C_{j-1} \right) \right) \tag{S.4.64}$$

$$= 2 \left(1 + \sum_{j=1}^{d} {}_{k}C_j \right) \tag{S.4.65}$$

$$= 2 \cdot \sum_{j=0}^{d} {}_{k}C_j \tag{S.4.66}$$

なお，式 (S.4.64) から式 (S.4.65) への変形には，公式

$${}_{k}C_j = {}_{k-1}C_j + {}_{k-1}C_{j-1} \tag{S.4.67}$$

を用いた．

ここで式 (S.4.66) において $k = d$ とすると，式 (S.4.58) より

$$2 \cdot \sum_{j=0}^{d} {}_{k}C_j = 2 \cdot \sum_{j=0}^{k} {}_{k}C_j = 2 \cdot 2^k = 2^{k+1} \tag{S.4.68}$$

となり，式 (S.4.59) と同じ結果となる．したがって，式 (S.4.59) と式 (S.4.66) をまとめて

$$L(k+1,d) = 2 \cdot \sum_{j=0}^{d} {}_kC_j \qquad (k > d-1) \tag{S.4.69}$$

とすることができる．

(2) $k \leq d-1$ のとき

この場合は，$L(k,d)$ および $L(k,d-1)$ の双方に対して式 (S.4.51) (b) の条件が当てはまるので

$$L(k,d) = 2^k \tag{S.4.70}$$
$$L(k,d-1) = 2^k \tag{S.4.71}$$

となる．これらを式 (S.4.52) に代入することにより，

$$L(k+1,d) = 2^k + 2^k = 2^{k+1} \qquad (k \leq d-1) \tag{S.4.72}$$

となる．

式 (S.4.69)，(S.4.72) をまとめると

$$L(k+1,d) = \begin{cases} 2 \cdot \displaystyle\sum_{j=0}^{d} {}_kC_j & (k > d-1) \\ 2^{k+1} & (k \leq d-1) \end{cases} \tag{S.4.73}$$

となる．上式は，式 (4.75) において $n = k+1$ とした式にほかならない．すなわち，式 (4.75) が $n = k$ に対して成り立つなら，$n = k+1$ に対しても成り立つことが示され，数学的帰納法によりすべての $n \geq 1$ に対して式 (4.75) が成り立つことが証明された．

【演習問題 4.7】

式 (4.67) において $n = 2(d+1)$ とおくと

$$p(n,d) = 2^{-2d-2} \cdot 2 \cdot \sum_{j=0}^{d} {}_{2d+1}C_j \tag{S.4.74}$$

$$= 2^{-2d-2} \cdot \left(\sum_{j=0}^{d} {}_{2d+1}C_j + \sum_{j=d+1}^{2d+1} {}_{2d+1}C_{2d+1-j} \right) \tag{S.4.75}$$

$$= 2^{-2d-2} \cdot \left(\sum_{j=0}^{d} {}_{2d+1}C_j + \sum_{j=d+1}^{2d+1} {}_{2d+1}C_j \right) \tag{S.4.76}$$

$$= 2^{-2d-2} \cdot \sum_{j=0}^{2d+1} {}_{2d+1}C_j \tag{S.4.77}$$

$$= 2^{-2d-2} \cdot 2^{2d+1} \tag{S.4.78}$$

$$= \frac{1}{2} \tag{S.4.79}$$

上式で，式 (S.4.75) から式 (S.4.76) への変形には

$${}_kC_j = {}_kC_{k-j} \tag{S.4.80}$$

を用い，式 (S.4.77) から式 (S.4.78) への変形には式 (S.4.58) を用いた．

第5章

【演習問題 5.1】

問 (1)

特徴抽出法 1, 2 によって得られた 2 次元特徴ベクトルをプロットした結果を**図 S.5.1** に示す．図では，クラス ω_1，ω_2 のパターンをそれぞれ●と○で示した．

問 (2)

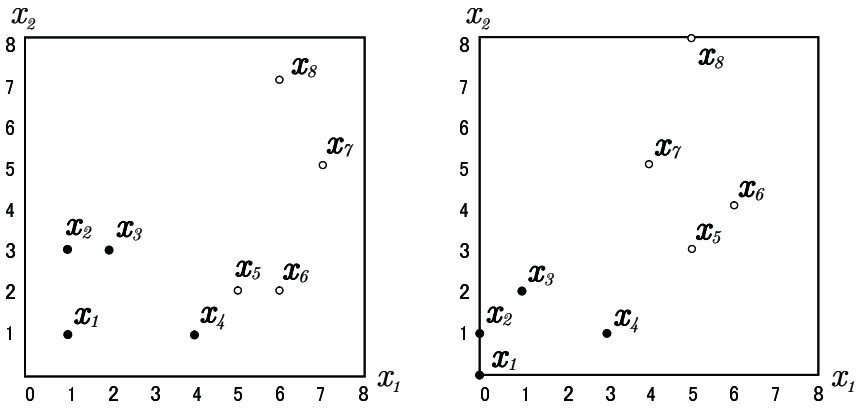
クラス数は $c = 2$ であり，全パターン数は $n = 8$，各クラスのパターン数は $n_1 = n_2 = 4$ である．

まず，特徴抽出法 1 における各クラスの平均ベクトル $\mathbf{m}_1$, $\mathbf{m}_2$ と，全パターンの平均ベクトル $\mathbf{m}$ は

$$\mathbf{m}_1 = (\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4)/4 = (2, 2)^t \tag{S.5.1}$$

$$\mathbf{m}_2 = (\boldsymbol{x}_5 + \boldsymbol{x}_6 + \boldsymbol{x}_7 + \boldsymbol{x}_8)/4 = (6, 4)^t \tag{S.5.2}$$

$$\mathbf{m} = (\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4 + \boldsymbol{x}_5 + \boldsymbol{x}_6 + \boldsymbol{x}_7 + \boldsymbol{x}_8)/8 = (4, 3)^t \tag{S.5.3}$$



(a) 特徴抽出法 1　　(b) 特徴抽出法 2

図 S.5.1: 特徴抽出法の比較

と求められる．式 (5.1) より

$$\sigma_W{}^2 = \frac{1}{8}\sum_{i=1}^{2}\sum_{\boldsymbol{x}\in\mathcal{X}_i}(\boldsymbol{x}-\mathbf{m}_i)^t(\boldsymbol{x}-\mathbf{m}_i) \tag{S.5.4}$$

$$= \frac{1}{8}\sum_{i=1}^{2}\sum_{\boldsymbol{x}\in\mathcal{X}_i}\|\boldsymbol{x}-\mathbf{m}_i\|^2 \tag{S.5.5}$$

$$= \frac{1}{8}(\overbrace{2+2+1+5}^{\omega_1}+\overbrace{5+4+2+9}^{\omega_2}) = 15/4 \tag{S.5.6}$$

が得られ，式 (5.2) より

$$\sigma_B{}^2 = \frac{1}{8}\sum_{i=1}^{2} 4\cdot(\mathbf{m}_i-\mathbf{m})^t(\mathbf{m}_i-\mathbf{m}) \tag{S.5.7}$$

$$= \frac{1}{2}\sum_{i=1}^{2}\|\mathbf{m}_i-\mathbf{m}\|^2 \tag{S.5.8}$$

$$= \frac{1}{2}(5+5) = 5 \tag{S.5.9}$$

が得られる．これらより，特徴抽出法 1 のクラス間分散・クラス内分散比 $J_{\sigma 1}$ は

$$J_{\sigma 1} = \frac{\sigma_B{}^2}{\sigma_W{}^2} = 4/3 \tag{S.5.10}$$

となる．

特徴抽出法 2 についても同様にして

$$\mathbf{m}_1 = (1,1)^t, \qquad \mathbf{m}_2 = (5,5)^t, \qquad \mathbf{m} = (3,3)^t \tag{S.5.11}$$

と求められ

$$\sigma_W{}^2 = \frac{1}{8}(\overbrace{2+1+1+4}^{\omega_1}+\overbrace{4+2+1+9}^{\omega_2}) = 3 \tag{S.5.12}$$

$$\sigma_B{}^2 = \frac{1}{2}(8+8) = 8 \tag{S.5.13}$$

を得る．これらより，特徴抽出法 2 のクラス間分散・クラス内分散比 $J_{\sigma 2}$ は

$$J_{\sigma 2} = \frac{{\sigma_B}^2}{{\sigma_W}^2} = 8/3 \tag{S.5.14}$$

となる．

式 (S.5.10)，(S.5.14) より

$$J_{\sigma 2} > J_{\sigma 1} \tag{S.5.15}$$

であるので，特徴抽出法 2 の方が優れている．

【演習問題 5.2】

区間 $[0, 4]$ で一様分布となる 1 次元関数を $u(x)$ とすると

$$u(x) = \begin{cases} a & (0 \leq x \leq 4) \\ 0 & (\text{その他}) \end{cases} \tag{S.5.16}$$

と書ける．ただし，a は定数である．上式の $u(x)$ は確率密度関数であるので，次式を満たさなくてはならない．

$$\int_{-\infty}^{\infty} u(x)dx = \int_0^4 a\, dx = a\,[x]_0^4 = 4a = 1 \tag{S.5.17}$$

これより $a = 1/4$ が得られ，$u(x)$ は下式で表される．

$$u(x) = \begin{cases} 1/4 & (0 \leq x \leq 4) \\ 0 & (\text{その他}) \end{cases} \tag{S.5.18}$$

同様にして，区間 $[2, 5]$ で一様分布となる 1 次元関数を $v(x)$ とすると

$$\int_{-\infty}^{\infty} v(x)dx = \int_2^5 v(x)dx = 1 \tag{S.5.19}$$

より，$v(x)$ は下式で表される．

$$v(x) = \begin{cases} 1/3 & (2 \leq x \leq 5) \\ 0 & (\text{その他}) \end{cases} \tag{S.5.20}$$

クラス ω_1 の確率密度関数を $p(\boldsymbol{x}|\omega_1)$ とすると，x_1 と x_2 は独立であるので，式 (S.5.18) より

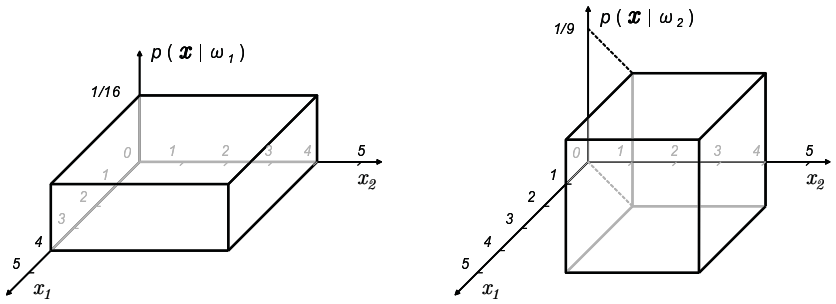
$$p(\boldsymbol{x}|\omega_1) = p(x_1, x_2|\omega_1) = p(x_1|\omega_1) \cdot p(x_2|\omega_1) \tag{S.5.21}$$

$$= u(x_1) \cdot u(x_2) = \begin{cases} 1/16 & (0 \leq x_1, x_2 \leq 4) \\ 0 & (\text{その他}) \end{cases} \tag{S.5.22}$$

と書ける．同様にして，式 (S.5.20) より

$$p(\boldsymbol{x}|\omega_2) = v(x_1) \cdot v(x_2) = \begin{cases} 1/9 & (2 \leq x_1, x_2 \leq 5) \\ 0 & (\text{その他}) \end{cases} \tag{S.5.23}$$

と書ける．**図 S.5.2** に，確率密度関数 $p(\boldsymbol{x}|\omega_1)$，$p(\boldsymbol{x}|\omega_2)$ を示した．



(a) クラス ω_1　　(b) クラス ω_2

図 S.5.2: 2 クラスの確率密度関数

ベイズ誤り確率の式 (5.15) を，ベイズの定理の式 (5.7) を用いて書き換えると下式が得られる．

$$e_B = \int \min\{P(\omega_1)P(\boldsymbol{x}|\omega_1),\, P(\omega_2)P(\boldsymbol{x}|\omega_2)\}\, d\boldsymbol{x} \tag{S.5.24}$$

ここで，

$$P(\omega_1) = 2/3, \qquad P(\omega_2) = 1/3 \tag{S.5.25}$$

である．

図 S.5.3 に，$P(\omega_1)P(\boldsymbol{x}|\omega_1)$，$P(\omega_2)P(\boldsymbol{x}|\omega_2)$ のカバーする領域を 2 次元的に示し，さらにそれらの関数を x_1 軸および x_2 軸へ射影した結果を示した．

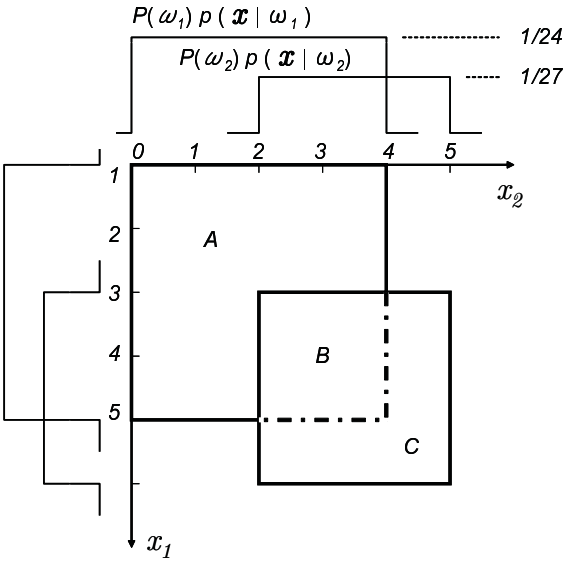


図 S.5.3: ベイズ誤り確率の算出

二つの確率密度関数の値域を考えると，積分の範囲は図の A，B，C の領域に限定してよいことがわかる．ここで簡単のため

$$Q(\boldsymbol{x}) \stackrel{\text{def}}{=} \min\{P(\omega_1)P(\boldsymbol{x}|\omega_1),\ P(\omega_2)P(\boldsymbol{x}|\omega_2)\} \tag{S.5.26}$$

と定義すると，式 (S.5.22)，(S.5.23)，(S.5.25) より

$$Q(\boldsymbol{x}) = \begin{cases} \min\left\{\dfrac{2}{3}\cdot\dfrac{1}{16}, \quad 0\right\} = 0 & (\boldsymbol{x} \in A) \\ \min\left\{\dfrac{2}{3}\cdot\dfrac{1}{16}, \quad \dfrac{1}{3}\cdot\dfrac{1}{9}\right\} = 1/27 & (\boldsymbol{x} \in B) \\ \min\left\{0, \quad \dfrac{1}{3}\cdot\dfrac{1}{9}\right\} = 0 & (\boldsymbol{x} \in C) \end{cases} \tag{S.5.27}$$

となる．これより，ベイズ誤り確率 e_B は，式 (S.5.24) を用いて

$$e_B = \int_{A+B+C} Q(\boldsymbol{x})\, d\boldsymbol{x} \tag{S.5.28}$$

$$= \int_A Q(\boldsymbol{x})\, d\boldsymbol{x} + \int_B Q(\boldsymbol{x})\, d\boldsymbol{x} + \int_C Q(\boldsymbol{x})\, d\boldsymbol{x} \tag{S.5.29}$$

$$= \int_B Q(\boldsymbol{x})\, d\boldsymbol{x} \tag{S.5.30}$$

$$= \int_2^4 \int_2^4 \frac{1}{27}\, dx_1\, dx_2 \tag{S.5.31}$$

$$= \frac{4}{27} \tag{S.5.32}$$

となる．また，ベイズ決定則を適用したときの決定境界は，図の太い一点鎖線のごとくなる．

【演習問題 5.3 *】

ここでは $x, x_1, x_2, \ldots$ などを特徴値と同時にパターンそのものを表すのに用いることにする．ベイズの定理の式 (4.4) より

$$P(\omega_1|x) = x \tag{S.5.33}$$

$$P(\omega_2|x) = 1 - x \tag{S.5.34}$$

が得られる．

入力パターン x を NN 法で識別したときの誤り確率 $e_n(x,x')$ は，式 (5.26) より

$$e_n(x,x') = P(\omega_1|x)\,P(\omega_2|x') + P(\omega_2|x)\,P(\omega_1|x') \tag{S.5.35}$$
$$= x(1-x') + x'(1-x) \tag{S.5.36}$$

と表される．これを x' で平均した結果を $e_n(x)$ と置くと，式 (S.5.36) より

$$e_n(x) = \int_0^1 e_n(x,x')\,q(x,x')\,dx' \tag{S.5.37}$$
$$= \int_0^1 \{x + (1-2x)x'\}\,q(x,x')\,dx' \tag{S.5.38}$$
$$= x + (1-2x)\int_0^1 x'\,q(x,x')\,dx' \tag{S.5.39}$$
$$= x + (1-2x)\,I \tag{S.5.40}$$

となる．ここで $q(x,x')$ は未知パターン x に対してその最近傍が x' となる確率を表す[*6]．また I は

$$I \stackrel{\text{def}}{=} \int_0^1 x'\,q(x,x')\,dx' \tag{S.5.41}$$

である．次に $q(x,x')$ を求めてみよう．

まず $0 \le x \le 1/2$ の場合を考える．ここで x' の位置を**図 S.5.4** に示されるような $[0,x]$，$[x,2x]$，$[2x,1]$ の 3 通りの場合に分けて考える．未知パターン x に対して，その最近傍が x' となるためには，n 個のプロトタイプの一つが座標値 x' に存在し，残り $(n-1)$ 個が図の太線の区間に存在しなくてはならない．そのような確率を求めることにより，次式を得る．

$$q(x,x') = \begin{cases} n(1-2x+2x')^{n-1} & (0 \le x' \le x) \\ n(1+2x-2x')^{n-1} & (x \le x' \le 2x) \\ n(1-x')^{n-1} & (2x \le x' \le 1) \end{cases} \tag{S.5.42}$$

明らかに

$$\int_0^1 q(x,x')\,dx' = 1 \tag{S.5.43}$$

[*6] $q(x,x')$ は確率密度関数であるから，“最近傍が $(x', x'+dx')$ に存在する確率を $q(x,x')dx'$ で表す” というのが正しい．

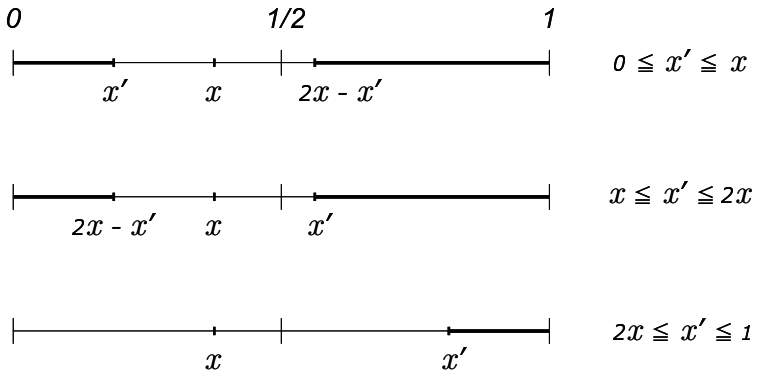


図 S.5.4: プロトタイプの存在可能な範囲 ($0 \leq x \leq 1/2$ の場合)

が成り立つ．式 (S.5.42) を式 (S.5.41) に代入することにより

$$I = \int_0^x n(1-2x+2x')^{n-1}\, x'\, dx' + \int_x^{2x} n(1+2x-2x')^{n-1}\, x'\, dx' + \int_{2x}^1 n(1-x')^{n-1}\, x'\, dx' \tag{S.5.44}$$

$$= x + \frac{1}{n+1}\,(1-2x)^n\,\{(n-1)x+1\} \qquad \left(0 \leq x \leq \frac{1}{2}\right) \tag{S.5.45}$$

となる．

一方，$1/2 \leq x \leq 1$ の場合は，x' の位置を $[0, 2x-1]$，$[2x-1, x]$，$[x, 1]$ の 3 通りに分けて考えることにより

$$q(x,x') = \begin{cases} n(x')^{n-1} & (0 \leq x' \leq 2x-1) \\ n(1-2x+2x')^{n-1} & (2x-1 \leq x' \leq x) \\ n(1+2x-2x')^{n-1} & (x \leq x' \leq 1) \end{cases} \tag{S.5.46}$$

を得る．式 (S.5.45) と同様にして

$$I = x + \frac{1}{n+1}\,(2x-1)^n\,\{(n-1)x-n\} \qquad \left(\frac{1}{2} \leq x \leq 1\right) \tag{S.5.47}$$

となる．式 (S.5.45)，(S.5.47) を式 (S.5.40) に代入し整理すると

$0 \le x \le 1/2$ のとき

$$e_n(x) = \frac{1}{2}\left\{-\frac{n-1}{n+1}(1-2x)^{n+2} + (1-2x)^{n+1} - (1-2x)^2 + 1\right\} \tag{S.5.48}$$

$1/2 \le x \le 1$ のとき

$$e_n(x) = \frac{1}{2}\left\{-\frac{n-1}{n+1}(2x-1)^{n+2} + (2x-1)^{n+1} - (2x-1)^2 + 1\right\} \tag{S.5.49}$$

となる．したがって，これらのプロトタイプを用いた NN 法の誤り確率 e_n は $2x-1=t$ と置換することにより

$$\begin{aligned} e_n &= \int_0^1 e_n(x)\, p(x)\, dx && \text{(S.5.50)} \\ &= \int_0^{1/2} e_n(x)\, p(x)\, dx + \int_{1/2}^1 e_n(x)\, p(x)\, dx && \text{(S.5.51)} \\ &= \frac{1}{2}\int_0^1 \left(-\frac{n-1}{n+1}\, t^{n+2} + t^{n+1} - t^2 + 1\right) dt && \text{(S.5.52)} \\ &= \frac{1}{3} + \frac{3n+5}{2(n+1)(n+2)(n+3)} && \text{(S.5.53)} \end{aligned}$$

と求められる．ここで式 (5.47) を用いた．

第６章

【演習問題 6.1 *】

式 (6.36) までは同じで，違いは式 (6.37) 以降である．すなわち，分散 $\tilde{\sigma}^2(\mathbf{A})$ は，式 (6.36) より

$$\tilde{\sigma}^2(\mathbf{A}) = \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}} \left(\mathbf{A}^t(\boldsymbol{x}-\mathbf{m})\right)^t \left(\mathbf{A}^t(\boldsymbol{x}-\mathbf{m})\right) \tag{S.6.1}$$

$$= \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}} (\boldsymbol{x}-\mathbf{m})^t \mathbf{A}\mathbf{A}^t(\boldsymbol{x}-\mathbf{m}) \tag{S.6.2}$$

となる．式 (6.30) より

$$\mathbf{A}\mathbf{A}^t = (\mathbf{u}_1,\ldots,\mathbf{u}_{\bar{d}}) \begin{pmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_{\bar{d}}^t \end{pmatrix} = \sum_{i=1}^{\bar{d}} \mathbf{u}_i\mathbf{u}_i^t \tag{S.6.3}$$

であるから，上式を式 (S.6.2) に代入して

$$\tilde{\sigma}^2(\mathbf{A}) = \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}} (\boldsymbol{x}-\mathbf{m})^t \left(\sum_{i=1}^{\bar{d}} \mathbf{u}_i\mathbf{u}_i^t\right)(\boldsymbol{x}-\mathbf{m}) \tag{S.6.4}$$

$$= \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{i=1}^{\bar{d}} (\boldsymbol{x}-\mathbf{m})^t\mathbf{u}_i\mathbf{u}_i^t(\boldsymbol{x}-\mathbf{m})$$

$$= \frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{i=1}^{\bar{d}} \mathbf{u}_i^t(\boldsymbol{x}-\mathbf{m})(\boldsymbol{x}-\mathbf{m})^t\mathbf{u}_i$$

$$= \sum_{i=1}^{\bar{d}} \mathbf{u}_i^t \left(\frac{1}{n}\sum_{\boldsymbol{x}\in\mathcal{X}} (\boldsymbol{x}-\mathbf{m})(\boldsymbol{x}-\mathbf{m})^t\right)\mathbf{u}_i$$

$$= \sum_{i=1}^{\bar{d}} \mathbf{u}_i^t\boldsymbol{\Sigma}\mathbf{u}_i \tag{S.6.5}$$

が得られる．上式の $\boldsymbol{\Sigma}$ は，式 (6.40) で定義される共分散行列である．

ここで，式 (6.32) の制約，すなわちベクトル表記で

$$\mathbf{u}_i^t\mathbf{u}_i = 1 \qquad (i=1,\ldots,\tilde{d}) \tag{S.6.6}$$

の制約の下で式 (S.6.5) を最大にする $\mathbf{A}$ を求める．そのためには，ラグランジュの未定乗数法により

$$L = \sum_{i=1}^{\bar{d}} \mathbf{u}_i^t \mathbf{\Sigma} \mathbf{u}_i - \sum_{i=1}^{\bar{d}} \lambda_i (\mathbf{u}_i^t \mathbf{u}_i - 1) \tag{S.6.7}$$

と置き，$\partial L / \partial \mathbf{u}_k = \mathbf{0}$ とすることにより（ベクトルの微分は**付録 A.2** 参照）

$$\mathbf{\Sigma} \mathbf{u}_k = \lambda_k \mathbf{u}_k \qquad (k = 1, \ldots, \tilde{d}) \tag{S.6.8}$$

が得られ，式 (6.45) の固有値問題に帰着される．

【演習問題 6.2】

問 (1)

学習パターンの平均ベクトル $\mathbf{m}$ と共分散行列 $\mathbf{\Sigma}$ は次式のごとく求められる．

$$\mathbf{m} = \frac{1}{6} \sum_{k=1}^{6} \boldsymbol{x}_k = (6,\, 6)^t \tag{S.6.9}$$

$$\mathbf{\Sigma} = \frac{1}{6} \sum_{k=1}^{6} (\boldsymbol{x}_k - \mathbf{m})(\boldsymbol{x}_k - \mathbf{m})^t = \frac{1}{3} \begin{pmatrix} 41 & 19 \\ 19 & 25 \end{pmatrix} \tag{S.6.10}$$

求めるべき $\mathbf{u}$ は $\mathbf{\Sigma}$ の固有ベクトルであるので下式を満たす．

$$\mathbf{\Sigma} \mathbf{u} = \lambda \mathbf{u} \tag{S.6.11}$$

固有値 λ は，$\mathbf{\Sigma}$ の固有方程式

$$\det(\lambda \mathbf{I} - \mathbf{\Sigma}) = 0 \tag{S.6.12}$$

の根である．そこで，$\mathbf{\Sigma}$ の固有方程式を求めると，λ に関する二次方程式

$$9\,\lambda^2 - 198\,\lambda + 664 = 0 \tag{S.6.13}$$

となり，これを解くと λ の解として

$$\lambda_1 = 11 + \frac{5}{3}\sqrt{17} \tag{S.6.14}$$

$$\lambda_2 = 11 - \frac{5}{3}\sqrt{17} \tag{S.6.15}$$

を得る．固有値 λ_1，λ_2 をそれぞれ式 (S.6.11) に代入し，さらに正規化条件

$$\|\mathbf{u}\| = 1 \tag{S.6.16}$$

を適用することにより，固有値 λ_1，λ_2 に対応する固有ベクトルとして

$$\mathbf{u}_1 = (0.833, 0.553)^t \tag{S.6.17}$$
$$\mathbf{u}_2 = (-0.553, 0.833)^t \tag{S.6.18}$$

を得る．求めるべき固有ベクトル $\mathbf{u}$ は，最大固有値 λ_1 に対応する固有ベクトル $\mathbf{u}_1$ であるので，下式となる．

$$\mathbf{u} = \mathbf{u}_1 = (0.833, 0.553)^t \tag{S.6.19}$$

問 (2)

図 S.6.1 に，$\mathbf{u}$ によって定まる射影軸を示す．平均ベクトル $\mathbf{m}$ が図中の*で示されている．

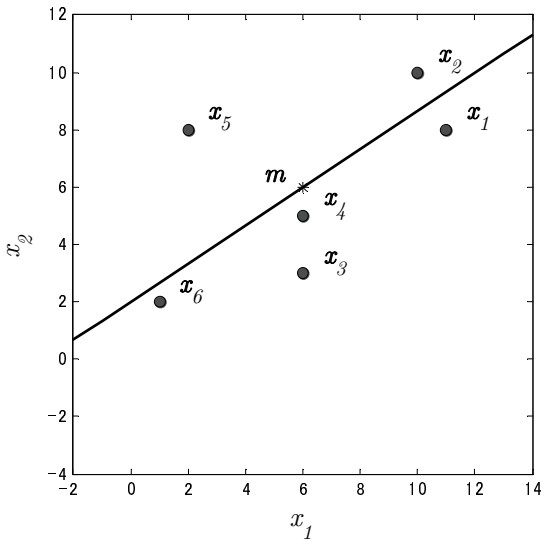


図 S.6.1: KL 展開で求めた射影軸

【演習問題 6.3】

問 (1)

クラス ω_1，ω_2 の平均ベクトル $\mathbf{m}_1$，$\mathbf{m}_2$ は，それぞれ

$$\mathbf{m}_1 = (\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3)/3 = (9, 7)^t \tag{S.6.20}$$
$$\mathbf{m}_2 = (\boldsymbol{x}_4 + \boldsymbol{x}_5 + \boldsymbol{x}_6)/3 = (3, 5)^t \tag{S.6.21}$$

となり，これより

$$\begin{aligned}\mathbf{S}_1 &= \sum_{\boldsymbol{x}\in\omega_1} (\boldsymbol{x} - \mathbf{m}_1)(\boldsymbol{x} - \mathbf{m}_1)^t \\ &= \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix} + \begin{pmatrix} 9 & 12 \\ 12 & 16 \end{pmatrix} = \begin{pmatrix} 14 & 17 \\ 17 & 26 \end{pmatrix}\end{aligned} \tag{S.6.22}$$

$$\begin{aligned}\mathbf{S}_2 &= \sum_{\boldsymbol{x}\in\omega_2} (\boldsymbol{x} - \mathbf{m}_2)(\boldsymbol{x} - \mathbf{m}_2)^t \\ &= \begin{pmatrix} 9 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix} + \begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix} = \begin{pmatrix} 14 & 3 \\ 3 & 18 \end{pmatrix}\end{aligned} \tag{S.6.23}$$

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 = \begin{pmatrix} 28 & 20 \\ 20 & 44 \end{pmatrix} = 4 \cdot \begin{pmatrix} 7 & 5 \\ 5 & 11 \end{pmatrix} \tag{S.6.24}$$

が得られ，2×2 の行列の逆行列計算法[*7] により

$${\mathbf{S}_W}^{-1} = \frac{1}{4} \cdot \frac{1}{52} \cdot \begin{pmatrix} 11 & -5 \\ -5 & 7 \end{pmatrix} = \frac{1}{208} \cdot \begin{pmatrix} 11 & -5 \\ -5 & 7 \end{pmatrix} \tag{S.6.25}$$

と計算できる．これらの結果より，フィッシャーの方法の式 (6.106) を用いて[*8]

$$\begin{aligned}\boldsymbol{w} &= {\mathbf{S}_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \\ &= \frac{1}{208} \cdot \begin{pmatrix} 11 & -5 \\ -5 & 7 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = \frac{1}{26} \cdot \begin{pmatrix} 7 \\ -2 \end{pmatrix}\end{aligned} \tag{S.6.26}$$

[*7] $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ のとき $\mathbf{A}^{-1} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \Big/ |\mathbf{A}|$

ただし，行列式 $|\mathbf{A}| = a_{11}\, a_{22} - a_{12}\, a_{21}$

[*8] 式 (6.106) に付した脚注に従い，等号とした

となり，正規化された $\boldsymbol{w}$ として下式が得られる．

$$\boldsymbol{w}/\|\boldsymbol{w}\| = \frac{1}{\sqrt{53}}\,(7,-2)^t = (0.962,\ -0.275)^t \qquad \text{(S.6.27)}$$

式 (S.6.26) より，y の勾配は $-2/7$ であることがわかる．**図 S.6.2** に，クラス ω_1，ω_2 のパターンをそれぞれ●と○で示し，射影軸 y と，y に射影したパターンを示す．図からパターンが y 軸上で誤りなく識別できることが確認できる．

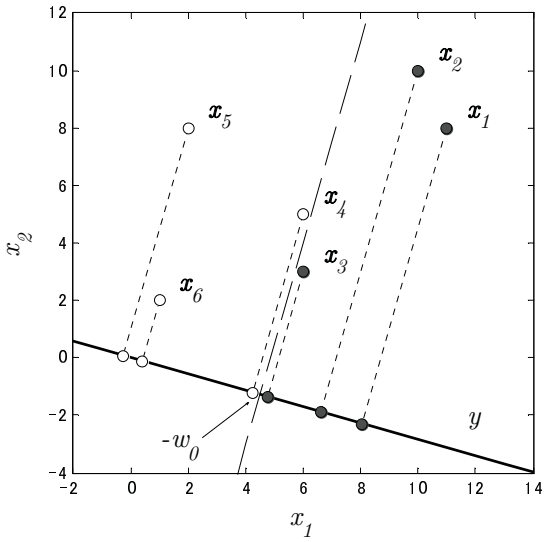


図 S.6.2: フィッシャーの方法で求めた射影軸

問 (2)

すでに**演習問題 6.2** で求めた KL 展開の主軸（図 S.6.1）との違いは明らかで，KL 展開の主軸は，クラス間を分離するための射影軸ではないことがわかる．

問 (3)

図の射影軸に各パターンを投影したときの射影値は，$\boldsymbol{x}_3$ が 4.94，$\boldsymbol{x}_4$ が 4.40 であった．図より，式 (6.110) の $-w_0$ としては，両値の間に設定すればよいことがわかるので，$-w_0 = (4.94 + 4.40)/2 = 4.67$ とした．式 (6.111) より，両クラスを分離する決定境界

$$g(\boldsymbol{x}) = \boldsymbol{w}^t \boldsymbol{x} + w_0 = 0 \tag{S.6.28}$$

が得られ，それが図 S.6.2 の鎖線で示されている[*9]．**演習問題 2.2** では，同データに対し，パーセプトロンの学習規則を適用して決定境界を求めた．その結果とは異なるが，上記決定境界も二つのクラスを正しく分離できている．

問 (4)

図 S.6.3 に，平均 $\mathbf{m}_1$，$\mathbf{m}_2$ の位置が*で示され，さらに，$\mathbf{m}_1 - \mathbf{m}_2$ の方向が太い矢印で示されている．射影軸 y は，この矢印に平行に設定されており，y 軸上で両クラスが正しく分離できていないことがわかる．

【演習問題 6.4】

問 (1)

原空間での特徴ベクトルを $\boldsymbol{x}$，変換後の特徴ベクトルを $\boldsymbol{x}'$ で表すと，

$$\boldsymbol{x}' = \mathbf{A}^t \boldsymbol{x} \tag{S.6.29}$$

が成り立つ．

原空間でのクラス ω_i の平均ベクトルを $\mathbf{m}_i$，変動行列を $\mathbf{S}_i$ とする（$i = 1, 2$, 以下同）．また，原空間での全パターンの平均ベクトルを $\mathbf{m}$，クラス内変動行列，クラス間変動行列をそれぞれ $\mathbf{S}_W$，$\mathbf{S}_B$ とする．

一方，変換後の空間で，上記に対応する各ベクトルをそれぞれ $\mathbf{m}_i{}'$，$\mathbf{m}'$，各行列をそれぞれ $\mathbf{S}_i{}'$，$\mathbf{S}_W{}'$，$\mathbf{S}_B{}'$ とする．さらに $\mathcal{X}_i$ に対応する変換後のパターン集合を $\mathcal{X}_i{}'$ とする．明らかに下式が成り立つ．

$$\mathbf{m}_i{}' = \mathbf{A}^t \mathbf{m}_i \qquad (i = 1, 2) \tag{S.6.30}$$

$$\mathbf{m}' = \mathbf{A}^t \mathbf{m} \tag{S.6.31}$$

[*9] 定数 w_0 の決め方については，4.3 節 ［1］で述べた．本問題に対して，例えば式 (4.46) を用いると，両クラスを正しく識別することができない．むしろ，ここで示したような目視による決定方法の方が優れている場合もある．

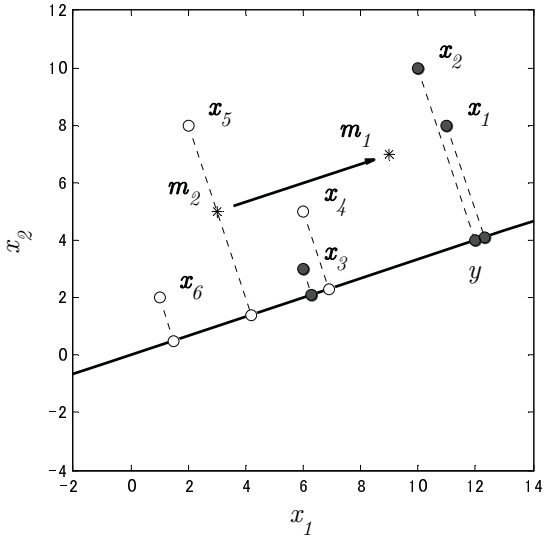


図 S.6.3: $\mathbf{m}_1 - \mathbf{m}_2$ の射影軸

上式，および式 (6.79)，(S.6.29) より

$$\mathbf{S}_i{}' = \sum_{\boldsymbol{x}' \in \mathcal{X}_i{}'} (\boldsymbol{x}' - \mathbf{m}_i{}')(\boldsymbol{x}' - \mathbf{m}_i{}')^t \tag{S.6.32}$$

$$= \sum_{\boldsymbol{x} \in \mathcal{X}_i} (\mathbf{A}^t \boldsymbol{x} - \mathbf{A}^t \mathbf{m}_i)(\mathbf{A}^t \boldsymbol{x} - \mathbf{A}^t \mathbf{m}_i)^t \tag{S.6.33}$$

$$= \mathbf{A}^t \left(\sum_{\boldsymbol{x} \in \mathcal{X}_i} (\boldsymbol{x} - \mathbf{m}_i)(\boldsymbol{x} - \mathbf{m}_i)^t \right) \mathbf{A} \tag{S.6.34}$$

$$= \mathbf{A}^t \mathbf{S}_i \mathbf{A} \qquad (i = 1,\ 2) \tag{S.6.35}$$

が得られ，上式および式 (6.80) より

$$\mathbf{S}_W{}' = \mathbf{S}_1{}' + \mathbf{S}_2{}' \tag{S.6.36}$$

$$= \mathbf{A}^t (\mathbf{S}_1 + \mathbf{S}_2) \mathbf{A} \tag{S.6.37}$$

$$= \mathbf{A}^t \mathbf{S}_W \mathbf{A} \tag{S.6.38}$$

が得られる．

式 (6.106) より，α を定数として

$$\boldsymbol{w} = \alpha\, \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{S.6.39}$$

と書ける．同様に $\boldsymbol{v}$ は β を定数として

$$\begin{aligned}
\boldsymbol{v} &= \beta \left(\mathbf{S}_W{}'\right)^{-1} (\mathbf{m}_1{}' - \mathbf{m}_2{}') && \text{(S.6.40)}\\
&= \beta\, (\mathbf{A}^t \mathbf{S}_W \mathbf{A})^{-1} \cdot \mathbf{A}^t(\mathbf{m}_1 - \mathbf{m}_2) && \text{(S.6.41)}\\
&= \beta\, \mathbf{A}^{-1}\mathbf{S}_W^{-1}(\mathbf{A}^t)^{-1} \cdot \mathbf{A}^t(\mathbf{m}_1 - \mathbf{m}_2) && \text{(S.6.42)}\\
&= \beta\, \mathbf{A}^{-1}\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) && \text{(S.6.43)}\\
&= (\beta/\alpha)\, \mathbf{A}^{-1}\boldsymbol{w} && \text{(S.6.44)}
\end{aligned}$$

となり，

$$\boldsymbol{v} \propto \mathbf{A}^{-1}\boldsymbol{w} \tag{S.6.45}$$

となる．なお，変換行列 $\mathbf{A}$ は，式 (6.3) に示すごとく対角行列であり，その対角成分はすべて非零であるので，逆行列 $\mathbf{A}^{-1}$ は必ず存在する．

問 (2)

クラス ω_1，ω_2 のパターン数をそれぞれ n_1，n_2，全パターン数を n とする．式 (6.83) より

$$\begin{aligned}
\mathbf{S}_B{}' &= \frac{n_1 n_2}{n}(\mathbf{m}_1{}' - \mathbf{m}_2{}')(\mathbf{m}_1{}' - \mathbf{m}_2{}')^t && \text{(S.6.46)}\\
&= \frac{n_1 n_2}{n}(\mathbf{A}^t\mathbf{m}_1 - \mathbf{A}^t\mathbf{m}_2)(\mathbf{A}^t\mathbf{m}_1 - \mathbf{A}^t\mathbf{m}_2)^t && \text{(S.6.47)}\\
&= \frac{n_1 n_2}{n}\mathbf{A}^t(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t\mathbf{A} && \text{(S.6.48)}\\
&= \mathbf{A}^t\mathbf{S}_B\mathbf{A} && \text{(S.6.49)}
\end{aligned}$$

を得る．

原特徴空間で求めた最適な軸 $\boldsymbol{w}$ は，フィッシャーの評価基準を最小化する．そのときの評価値は，式 (6.98) より

$$J_S(\boldsymbol{w}) = \frac{\boldsymbol{w}^t\, \mathbf{S}_B\, \boldsymbol{w}}{\boldsymbol{w}^t\, \mathbf{S}_W\, \boldsymbol{w}} \tag{S.6.50}$$

と書ける．

一方，変換後の特徴空間で求めた最適な軸 $\boldsymbol{v}$ の評価値 $J_S(\boldsymbol{v})$ は，式 (S.6.44)，(S.6.49) より下式のごとくなる．

$$J_S(\boldsymbol{v}) = \frac{\boldsymbol{v}^t \, \mathbf{S}_B{}' \, \boldsymbol{v}}{\boldsymbol{v}^t \, \mathbf{S}_W{}' \, \boldsymbol{v}} \tag{S.6.51}$$

$$= \frac{(\mathbf{A}^{-1}\,\boldsymbol{w})^t \, \mathbf{A}^t \, \mathbf{S}_B \, \mathbf{A} \, (\mathbf{A}^{-1}\,\boldsymbol{w})}{(\mathbf{A}^{-1}\,\boldsymbol{w})^t \, \mathbf{A}^t \, \mathbf{S}_W \, \mathbf{A} \, (\mathbf{A}^{-1}\,\boldsymbol{w})} \tag{S.6.52}$$

$$= \frac{\boldsymbol{w}^t (\mathbf{A}^{-1})^t \, \mathbf{A}^t \, \mathbf{S}_B \, \mathbf{A} \, \mathbf{A}^{-1} \, \boldsymbol{w}}{\boldsymbol{w}^t (\mathbf{A}^{-1})^t \, \mathbf{A}^t \, \mathbf{S}_W \, \mathbf{A} \, \mathbf{A}^{-1} \, \boldsymbol{w}} \tag{S.6.53}$$

$$= \frac{\boldsymbol{w}^t \, \mathbf{S}_B \, \boldsymbol{w}}{\boldsymbol{w}^t \, \mathbf{S}_W \, \boldsymbol{w}} \tag{S.6.54}$$

$$= J_S(\boldsymbol{w}) \tag{S.6.55}$$

したがって，フィッシャーの評価基準値は，正規化前後で変わらないことが確かめられる．すなわち，フィッシャーの方法により最適な軸を求める際，正規化の処理は不要であることがわかる．

【演習問題 6.5】

(1) 式 (6.177) の証明

式 (6.177) の $J_1(\mathbf{A})$ を最大にする $\mathbf{A}$ を求めることは，式 (6.176) の固有値問題

$$\boldsymbol{\Sigma}_B \mathbf{A} = \boldsymbol{\Sigma}_W \mathbf{A} \boldsymbol{\Lambda} \tag{S.6.56}$$

に帰着することが知られている． 上式は

$$\boldsymbol{\Sigma}_W^{-1} \boldsymbol{\Sigma}_B \mathbf{A} = \mathbf{A} \boldsymbol{\Lambda} \tag{S.6.57}$$

と書き直せ，固有値問題の解 $\mathbf{A}$ が得られたときは，$\boldsymbol{\Sigma}_W^{-1} \boldsymbol{\Sigma}_B$ の d 個の固有値のうち $\hat{d}$ 個が $\boldsymbol{\Lambda}$ の対角成分となることがわかる．

式 (S.6.56) の両辺に左から $\mathbf{A}^t$ をかけると

$$\mathbf{A}^t \boldsymbol{\Sigma}_B \mathbf{A} = \mathbf{A}^t \boldsymbol{\Sigma}_W \mathbf{A} \boldsymbol{\Lambda} \tag{S.6.58}$$

となるので，式 (6.166)，(6.167) より

$$\tilde{\boldsymbol{\Sigma}}_B = \tilde{\boldsymbol{\Sigma}}_W \boldsymbol{\Lambda} \tag{S.6.59}$$

が得られる．

ここで，$\mathbf{A}$ に対して式 (6.175) の正規化条件

$$\tilde{\mathbf{\Sigma}}_W = \mathbf{A}^t \mathbf{\Sigma}_W \mathbf{A} = \mathbf{I} \tag{S.6.60}$$

が課せられているなら，上式と式 (S.6.59) より

$$\tilde{\mathbf{\Sigma}}_B = \tilde{\mathbf{\Sigma}}_W \mathbf{\Lambda} = \mathbf{\Lambda} \tag{S.6.61}$$

が得られる．以上より

$$J_1(\mathbf{A}) = \frac{\mathrm{tr}\left(\tilde{\mathbf{\Sigma}}_B\right)}{\mathrm{tr}\left(\tilde{\mathbf{\Sigma}}_W\right)} = \frac{\mathrm{tr}\left(\mathbf{\Lambda}\right)}{\mathrm{tr}\left(\mathbf{I}\right)} = \frac{1}{\tilde{d}} \sum_{i=1}^{\bar{d}} \lambda_i \tag{S.6.62}$$

が得られる．したがって，$J_1(\mathbf{A})$ を最大にするには，固有値を大きい方から順に $\tilde{d}$ 個選んで $\lambda_1, \ldots, \lambda_{\bar{d}}$ とすればよい．

以上より，式 (6.177) が証明された．ここでは正規化条件の式 (S.6.60) を用いたが，以下の (2) (3) (4) では正規化条件を適用することなく証明できる．

(2) 式 (6.178) の証明

式 (6.170) の $J_2(\mathbf{A})$ を最大にする $\mathbf{A}$ を求めることは，同じく式 (6.176) の固有値問題に帰着することが知られている．式 (S.6.59) は

$$\tilde{\mathbf{\Sigma}}_W^{-1} \tilde{\mathbf{\Sigma}}_B = \mathbf{\Lambda} \tag{S.6.63}$$

と書き直せる．すなわち上式は，式 (S.6.57) で得られた $\mathbf{\Lambda}$ の $\hat{d}$ 個の対角成分が，$\tilde{\mathbf{\Sigma}}_W^{-1} \tilde{\mathbf{\Sigma}}_B$ の固有値であることを示している．

行列のトレースは，行列の固有値の和に一致するので，式 (6.170) より

$$J_2(\mathbf{A}) = \mathrm{tr}\left(\tilde{\mathbf{\Sigma}}_W^{-1} \tilde{\mathbf{\Sigma}}_B\right) = \sum_{i=1}^{\bar{d}} \lambda_i \tag{S.6.64}$$

となる．したがって，$J_2(\mathbf{A})$ を最大にするには，固有値を大きい方から順に $\tilde{d}$ 個選んで $\lambda_1, \ldots, \lambda_{\bar{d}}$ とすればよい．

以上より，式 (6.178) が証明された．

(3) 式 (6.179) の証明

式 (6.171) の $J_3(\mathbf{A})$ を最大にする $\mathbf{A}$ を求めることは，前問と同様，式 (S.6.56) の固有値問題に帰着することが知られている．したがって，式 (S.6.59) が成り立つので

$$\begin{aligned}\det\left(\tilde{\mathbf{\Sigma}}_B\right) &= \det\left(\tilde{\mathbf{\Sigma}}_W\mathbf{\Lambda}\right)\\ &= \det\left(\tilde{\mathbf{\Sigma}}_W\right)\cdot\det\left(\mathbf{\Lambda}\right)\end{aligned} \tag{S.6.65}$$

となる．これより，

$$J_3(\mathbf{A}) = \frac{\det\left(\tilde{\mathbf{\Sigma}}_B\right)}{\det\left(\tilde{\mathbf{\Sigma}}_W\right)} = \det\left(\mathbf{\Lambda}\right) \tag{S.6.66}$$

$$= \prod_{i=1}^{\tilde{d}} \lambda_i \tag{S.6.67}$$

が得られる．式 (S.6.66) から (S.6.67) への変形には，行列式が行列の固有値の積に一致することを用いた（式 (6.174)）．したがって，$J_3(\mathbf{A})$ を最大にするには，固有値を大きい方から順に $\tilde{d}$ 個選んで $\lambda_1,\ldots,\lambda_{\tilde{d}}$ とすればよい．

以上より，式 (6.179) が証明された．

(4) 式 (6.180) の証明

次式で $J_0(\mathbf{A})$ を

$$J_0(\mathbf{A}) \stackrel{\text{def}}{=} \frac{\det\left(\tilde{\mathbf{\Sigma}}_T\right)}{\det\left(\tilde{\mathbf{\Sigma}}_W\right)} \tag{S.6.68}$$

と定義すれば，

$$J_4(\mathbf{A}) = \log\left(J_0(\mathbf{A})\right) \tag{S.6.69}$$

であるので，$J_4(\mathbf{A})$ を最大にすることは $J_0(\mathbf{A})$ を最大にすることと等価である．式 (S.6.68) の最大化も式 (S.6.56) で示した固有値問題に帰着することが知られている．式 (6.161) で示した関係式

$$\mathbf{\Sigma}_T = \mathbf{\Sigma}_W + \mathbf{\Sigma}_B \tag{S.6.70}$$

を式 (S.6.56) に代入して $\mathbf{\Sigma}_B$ を消去すると

$$(\mathbf{\Sigma}_T - \mathbf{\Sigma}_W)\,\mathbf{A} = \mathbf{\Sigma}_W \mathbf{A}\mathbf{\Lambda} \tag{S.6.71}$$

より

$$\begin{aligned}\mathbf{\Sigma}_T\mathbf{A} &= \mathbf{\Sigma}_W\mathbf{A} + \mathbf{\Sigma}_W\mathbf{A}\mathbf{\Lambda} = \mathbf{\Sigma}_W\mathbf{A}\,(\mathbf{I}+\mathbf{\Lambda}) \\ &= \mathbf{\Sigma}_W\mathbf{A}\mathbf{\Lambda}'\end{aligned} \tag{S.6.72}$$

が得られる．ただし

$$\mathbf{\Lambda}' \stackrel{\text{def}}{=} \mathbf{I}+\mathbf{\Lambda} = \begin{pmatrix} \lambda_1+1 & & & 0 \\ & \lambda_2+1 & & \\ & & \ddots & \\ 0 & & & \lambda_{\tilde{d}}+1 \end{pmatrix} \tag{S.6.73}$$

と定義した．

式 (S.6.72) は，次のことを示している．すなわち，$\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_T$ の固有ベクトルは $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_B$ の固有ベクトルと同じであり，固有値は $\mathbf{\Lambda}'$ の対角成分，λ_i+1 である $(i=1,\ldots,\tilde{d})$．式 (S.6.67) より明らかなように，$J_3(\mathbf{A})$ は，$\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_B$ の $\tilde{d}$ 個の固有値の積である．同様にして，$J_0(\mathbf{A})$ は，$\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_T$ の $\tilde{d}$ 個の固有値の積として

$$J_0(\mathbf{A}) = \prod_{i=1}^{\tilde{d}}(\lambda_i+1) \tag{S.6.74}$$

と表される．したがって，式 (S.6.69) より

$$\begin{aligned}J_4(\mathbf{A}) &= \log\big(J_0(\mathbf{A})\big) = \log\left(\prod_{i=1}^{\tilde{d}}(\lambda_i+1)\right) \\ &= \sum_{i}^{\tilde{d}}\log(\lambda_i+1)\end{aligned} \tag{S.6.75}$$

が得られる．これまでと同様，$J_4(\mathbf{A})$ を最大にするには，固有値を大きい方から順に $\tilde{d}$ 個選んで $\lambda_1,\ldots,\lambda_{\tilde{d}}$ とすればよい．

以上より，式 (6.180) が証明された．

【演習問題 6.6】

問 (1)

平均ベクトル $\mathbf{m}$ は

$$\mathbf{m} = \frac{n_1\mathbf{m}_1 + n_2\mathbf{m}_2}{n} \tag{S.6.76}$$

であるので

$$\begin{cases} \mathbf{m}_1 - \mathbf{m} = \dfrac{n\mathbf{m}_1 - n_1\mathbf{m}_1 - n_2\mathbf{m}_2}{n} = \dfrac{n_2(\mathbf{m}_1 - \mathbf{m}_2)}{n} \\ \mathbf{m}_2 - \mathbf{m} = \dfrac{n\mathbf{m}_2 - n_1\mathbf{m}_1 - n_2\mathbf{m}_2}{n} = \dfrac{n_1(\mathbf{m}_2 - \mathbf{m}_1)}{n} \end{cases} \tag{S.6.77}$$

が成り立つ．これらの結果より，式 (6.82) は

$$\begin{aligned} \mathbf{S}_B &= \sum_{i=1,2} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \\ &= n_1\left(\frac{n_2}{n}\right)^2 (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \\ &\qquad + n_2\left(\frac{n_1}{n}\right)^2 (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \\ &= \left(\frac{n_1 n_2^2}{n^2} + \frac{n_2 n_1^2}{n^2}\right)(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \\ &= \frac{n_1 n_2}{n}(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \end{aligned} \tag{S.6.78}$$

となり，式 (6.83) と一致することが証明された．

問 (2)

まず，式 (6.140) と式 (6.141) が等しいことを証明する．式 (6.140) は，すでに式 (6.126) で示されている．ここで，$J_\Sigma(\boldsymbol{w})$ を最大にする $\boldsymbol{w}$ は，式 (6.125) で示した $\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_B$ の最大固有値 λ_1 に対応する固有ベクトルであるので

$$\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_B\boldsymbol{w} = \lambda_1\boldsymbol{w} \tag{S.6.79}$$

が成り立つ．上式に，式 (6.116) の $\mathbf{\Sigma}_B$ を代入すると下式を得る．

$$P(\omega_1)P(\omega_2)\mathbf{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t\boldsymbol{w} = \lambda_1\boldsymbol{w} \tag{S.6.80}$$

ここで $(\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{w}$ がスカラー量であることに注意して上式を変形すると，下式を得る．

$$P(\omega_1)P(\omega_2)\boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \frac{\lambda_1 \boldsymbol{w}}{(\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{w}} \tag{S.6.81}$$

一方，式 (6.136) より，共分散行列が二つのクラスで異なる場合のマハラノビス汎距離 $D_M(\mathbf{m}_1, \mathbf{m}_2)$ は，下式で表される．

$$D_M^2(\mathbf{m}_1, \mathbf{m}_2) = (\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{S.6.82}$$

上式と式 (S.6.81) より

$$\begin{aligned} & P(\omega_1)P(\omega_2)D_M^2(\mathbf{m}_1, \mathbf{m}_2) \\ &= (\mathbf{m}_1 - \mathbf{m}_2)^t P(\omega_1)P(\omega_2)\boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \\ &= \lambda_1 \cdot \frac{(\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{w}}{(\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{w}} \\ &= \lambda_1 \end{aligned} \tag{S.6.83}$$

となり，式 (6.140) と式 (6.141) が等しいことが証明された．

次に式 (6.141) と式 (6.142) が等しいことを証明しよう．
次式は明らかである．

$$P(\omega_1) + P(\omega_2) = 1 \tag{S.6.84}$$

$$\mathbf{m} = P(\omega_1)\,\mathbf{m}_1 + P(\omega_2)\,\mathbf{m}_2 \tag{S.6.85}$$

これより

$$\begin{aligned} \mathbf{m}_1 - \mathbf{m} &= (1 - P(\omega_1))\mathbf{m}_1 - P(\omega_2)\,\mathbf{m}_2 \\ &= P(\omega_2)\,(\mathbf{m}_1 - \mathbf{m}_2) \end{aligned} \tag{S.6.86}$$

が得られ，同様にして

$$\mathbf{m}_2 - \mathbf{m} = -P(\omega_1)\,(\mathbf{m}_1 - \mathbf{m}_2) \tag{S.6.87}$$

が得られる．これらより

$$\begin{aligned} & P(\omega_1)D_M^2(\mathbf{m}_1, \mathbf{m}) \\ &= P(\omega_1)(\mathbf{m}_1 - \mathbf{m})^t \boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}) \\ &= P(\omega_1)P(\omega_2)^2(\mathbf{m}_1 - \mathbf{m}_2)^t \boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \\ &= P(\omega_1)P(\omega_2)^2 D_M^2(\mathbf{m}_1, \mathbf{m}_2) \end{aligned} \tag{S.6.88}$$

が得られ，同様にして

$$P(\omega_2)D_M^2(\mathbf{m}_2,\mathbf{m}) = P(\omega_1)^2P(\omega_2)D_M^2(\mathbf{m}_1,\mathbf{m}_2) \tag{S.6.89}$$

が得られる．これらより

$$\begin{aligned} &P(\omega_1)D_M^2(\mathbf{m}_1,\mathbf{m}) + P(\omega_2)D_M^2(\mathbf{m}_2,\mathbf{m}) \\ &= P(\omega_1)P(\omega_2)\left(P(\omega_1)+P(\omega_2)\right)D_M^2(\mathbf{m}_1,\mathbf{m}_2) \\ &= P(\omega_1)P(\omega_2)D_M^2(\mathbf{m}_1,\mathbf{m}_2) \end{aligned} \tag{S.6.90}$$

となり，式 (6.141) と式 (6.142) が等しいことが証明された．

最後に，制約条件である式 (6.143) の下に式 (6.144) が成り立つことを証明する．式 (6.86) より下式が得られる．

$$\tilde{m}_1 = \boldsymbol{w}^t\mathbf{m}_1, \quad \tilde{m}_2 = \boldsymbol{w}^t\mathbf{m}_2 \tag{S.6.91}$$

式 (6.143) の両辺に左から $\boldsymbol{w}$ を乗じると

$$\boldsymbol{w}\boldsymbol{w}^t\boldsymbol{\Sigma}_W\boldsymbol{w} = \boldsymbol{w} \tag{S.6.92}$$

となるので

$$\boldsymbol{w}\boldsymbol{w}^t\boldsymbol{\Sigma}_W = \mathbf{I} \tag{S.6.93}$$

が成り立ち，これより

$$\boldsymbol{w}\boldsymbol{w}^t = \boldsymbol{\Sigma}_W^{-1} \tag{S.6.94}$$

が得られる．上式を式 (S.6.82) に代入すると

$$D_M^2(\mathbf{m}_1,\mathbf{m}_2) = (\mathbf{m}_1-\mathbf{m}_2)^t\boldsymbol{w}\boldsymbol{w}^t(\mathbf{m}_1-\mathbf{m}_2) \tag{S.6.95}$$

$$= \left(\boldsymbol{w}^t(\mathbf{m}_1-\mathbf{m}_2)\right)^2 \tag{S.6.96}$$

$$= (\tilde{m}_1-\tilde{m}_2)^2 \tag{S.6.97}$$

となり，式 (6.144) が証明された．

問 (3)

式 (6.155) において，

$$f_{ij} = P(\omega_i)P(\omega_j)(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^t$$

とおくと，$f_{ii} = 0$，$f_{ij} = f_{ji}$ であるので，式 (6.155) と式 (6.156) が等しいことは明らかである．そこで，以下では式 (6.154) と式 (6.155) が等しいことのみ証明する．

煩雑になるのを避けるため，$P(\omega_i)$ を P_i と略記する．下式は明らかである．

$$\sum_{k=1}^{c} P_k = 1 \tag{S.6.98}$$

$$\mathbf{m} = \sum_{k=1}^{c} P_k \mathbf{m}_k \tag{S.6.99}$$

以下，二通りの証明法を示す．

（証明法 1）

式 (6.155) $\times 2$ は，下式のごとく変形される．

$$\sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^t \tag{S.6.100}$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j (\mathbf{m}_i\mathbf{m}_i^t - \mathbf{m}_i\mathbf{m}_j^t - \mathbf{m}_j\mathbf{m}_i^t + \mathbf{m}_j\mathbf{m}_j^t) \tag{S.6.101}$$

$$\begin{aligned} = \sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j \Big(& (\mathbf{m}_i\mathbf{m}_i^t - \mathbf{m}_i\mathbf{m}^t - \mathbf{m}\mathbf{m}_i^t + \mathbf{m}\mathbf{m}^t) \\ & + (\mathbf{m}_j\mathbf{m}_j^t - \mathbf{m}_j\mathbf{m}^t - \mathbf{m}\mathbf{m}_j^t + \mathbf{m}\mathbf{m}^t) \\ & - (\mathbf{m}_i\mathbf{m}_j^t - \mathbf{m}_i\mathbf{m}^t - \mathbf{m}\mathbf{m}_j^t + \mathbf{m}\mathbf{m}^t) \\ & - (\mathbf{m}_j\mathbf{m}_i^t - \mathbf{m}_j\mathbf{m}^t - \mathbf{m}\mathbf{m}_i^t + \mathbf{m}\mathbf{m}^t) \Big) \end{aligned} \tag{S.6.102}$$

$$\begin{aligned} = \sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j \Big(& (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t + (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^t \\ & - (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^t - (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \Big) \end{aligned} \tag{S.6.103}$$

$$= \sum_{i=1}^{c} P_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \sum_{j=1}^{c} P_j + \sum_{i=1}^{c} P_i \sum_{j=1}^{c} P_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^t$$

$$-\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})\sum_{j=1}^{c} P_j(\mathbf{m}_j-\mathbf{m})^t-\sum_{j=1}^{c} P_j(\mathbf{m}_j-\mathbf{m})\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})^t \tag{S.6.104}$$

$$=2\cdot\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^t \tag{S.6.105}$$

式 (S.6.98)，式 (S.6.99) を用いると，式 (S.6.104) の各項に対しては以下が成り立つ．

$$\begin{cases} (\text{第 1 項}) \quad =(\text{第 2 項})=\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^t \\ (\text{第 3 項}) \quad =(\text{第 4 項})=\mathbf{0} \end{cases}$$

式 (S.6.104) から式 (S.6.105) への変形では，上式の関係を用いた．

式 (S.6.100)，(S.6.105) より

$$\begin{aligned} &\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^t \\ =\quad &\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c} P_iP_j(\mathbf{m}_i-\mathbf{m}_j)(\mathbf{m}_i-\mathbf{m}_j)^t \end{aligned} \tag{S.6.106}$$

が得られ，式 (6.154) と式 (6.155) は等しいことが証明された．

（証明法 2）

式 (S.6.98)，(S.6.99) より下式を得る．

$$\begin{aligned} \mathbf{m}_i-\mathbf{m}&=\mathbf{m}_i-\sum_{k=1}^{c} P_k\mathbf{m}_k=\mathbf{m}_i\sum_{k=1}^{c} P_k-\sum_{k=1}^{c} P_k\mathbf{m}_k \\ &=\sum_{k=1}^{c} P_k(\mathbf{m}_i-\mathbf{m}_k) \end{aligned} \tag{S.6.107}$$

式 (S.6.107) を用いると，式 (6.154) は以下のごとく書き換えられる．ただし，

途中で式 (S.6.98) を用いた.

$$
\begin{aligned}
&\sum_{i=1}^{c} P_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^t \\
&\quad=\sum_{i=1}^{c} P_i\left(\sum_{k=1}^{c} P_k(\mathbf{m}_i-\mathbf{m}_k)\right)\left(\sum_{l=1}^{c} P_l(\mathbf{m}_i-\mathbf{m}_l)^t\right) \\
&\quad=\sum_{i=1}^{c}\sum_{k=1}^{c}\sum_{l=1}^{c} P_iP_kP_l(\mathbf{m}_i-\mathbf{m}_k)(\mathbf{m}_i-\mathbf{m}_l)^t \\
&\quad=\sum_{i=1}^{c}\sum_{k=1}^{c}\sum_{l=1}^{c} P_iP_kP_l(\mathbf{m}_i\mathbf{m}_i^t+\mathbf{m}_k\mathbf{m}_l^t-\mathbf{m}_i\mathbf{m}_l^t-\mathbf{m}_k\mathbf{m}_i^t) \\
&\quad=\sum_{k=1}^{c} P_k\sum_{l=1}^{c} P_l\sum_{i=1}^{c} P_i\mathbf{m}_i\mathbf{m}_i^t+\sum_{i=1}^{c} P_i\sum_{k=1}^{c}\sum_{l=1}^{c} P_kP_l\mathbf{m}_k\mathbf{m}_l^t \\
&\qquad-\sum_{k=1}^{c} P_k\sum_{i=1}^{c}\sum_{l=1}^{c} P_iP_l\mathbf{m}_i\mathbf{m}_l^t-\sum_{l=1}^{c} P_l\sum_{k=1}^{c}\sum_{i=1}^{c} P_kP_i\mathbf{m}_k\mathbf{m}_i^t \\
&\quad=\sum_{i=1}^{c} P_i\mathbf{m}_i\mathbf{m}_i^t+\sum_{k=1}^{c}\sum_{l=1}^{c} P_kP_l\mathbf{m}_k\mathbf{m}_l^t-\sum_{i=1}^{c}\sum_{l=1}^{c} P_iP_l\mathbf{m}_i\mathbf{m}_l^t-\sum_{k=1}^{c}\sum_{i=1}^{c} P_kP_i\mathbf{m}_k\mathbf{m}_i^t \\
&\quad=\sum_{i=1}^{c} P_i\mathbf{m}_i\mathbf{m}_i^t-\sum_{k=1}^{c}\sum_{l=1}^{c} P_kP_l\mathbf{m}_k\mathbf{m}_l^t \qquad \text{(S.6.108)}
\end{aligned}
$$

一方，式 (6.155) は以下のごとく書き換えられる．同様に，途中で式 (S.6.98) を用いた.

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c} P_iP_j(\mathbf{m}_i-\mathbf{m}_j)(\mathbf{m}_i-\mathbf{m}_j)^t \\
&\quad=\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c} P_iP_j(\mathbf{m}_i\mathbf{m}_i^t+\mathbf{m}_j\mathbf{m}_j^t-\mathbf{m}_i\mathbf{m}_j^t-\mathbf{m}_j\mathbf{m}_i^t) \\
&\quad=\frac{1}{2}\sum_{j=1}^{c} P_j\sum_{i=1}^{c} P_i\mathbf{m}_i\mathbf{m}_i^t+\frac{1}{2}\sum_{i=1}^{c} P_j\sum_{j=1}^{c} P_j\mathbf{m}_j\mathbf{m}_j^t \\
&\qquad-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c} P_iP_j\mathbf{m}_i\mathbf{m}_j^t-\frac{1}{2}\sum_{i=1}^{c}\sum_{j=1}^{c} P_iP_j\mathbf{m}_j\mathbf{m}_i^t
\end{aligned}
$$

$$= \frac{1}{2}\sum_{i=1}^{c} P_i \mathbf{m}_i \mathbf{m}_i^t + \frac{1}{2}\sum_{i=1}^{c} P_j \mathbf{m}_j \mathbf{m}_j^t - \sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j \mathbf{m}_i \mathbf{m}_j^t$$

$$= \sum_{i=1}^{c} P_i \mathbf{m}_i \mathbf{m}_i^t - \sum_{i=1}^{c}\sum_{j=1}^{c} P_i P_j \mathbf{m}_i \mathbf{m}_j^t \tag{S.6.109}$$

式 (S.6.108) と式 (S.6.109) は等しい．したがって，式 (6.154) と式 (6.155) は等しいことが証明された．

問 (4)

大きさ $d \times d$ のクラス内共分散行列 $\mathbf{\Sigma}_W$ を式 (6.113) で定義すると，マハラノビス汎距離 $D_M^2(\mathbf{m}_i, \mathbf{m})$，および $D_M^2(\mathbf{m}_i, \mathbf{m}_j)$ は，式 (6.136) より

$$D_M^2(\mathbf{m}_i, \mathbf{m}) = (\mathbf{m}_i - \mathbf{m})^t \mathbf{\Sigma}_W^{-1} (\mathbf{m}_i - \mathbf{m}) \tag{S.6.110}$$

$$D_M^2(\mathbf{m}_i, \mathbf{m}_j) = (\mathbf{m}_i - \mathbf{m}_j)^t \mathbf{\Sigma}_W^{-1} (\mathbf{m}_i - \mathbf{m}_j) \tag{S.6.111}$$

と書ける．クラス内共分散行列 $\mathbf{\Sigma}_W$ は対称行列であるので，その逆行列 $\mathbf{\Sigma}_W^{-1}$ も対称行列となる．さらに，$\mathbf{\Sigma}_W^{-1}$ が対称行列であるなら，d 次正方行列 $\mathbf{A}$ を用いて

$$\mathbf{\Sigma}_W^{-1} = \mathbf{A}^t \mathbf{A} \tag{S.6.112}$$

と表せる（**演習問題 6.7** 参照）．ここで

$$\mathbf{y} = \mathbf{A}\mathbf{m}, \qquad \mathbf{y}_i = \mathbf{A}\mathbf{m}_i \tag{S.6.113}$$

と置くと，式 (S.6.110) は

$$D_M^2(\mathbf{m}_i, \mathbf{m}) = (\mathbf{m}_i - \mathbf{m})^t \mathbf{A}^t \mathbf{A} (\mathbf{m}_i - \mathbf{m}) \tag{S.6.114}$$

$$= (\mathbf{A}(\mathbf{m}_i - \mathbf{m}))^t (\mathbf{A}(\mathbf{m}_i - \mathbf{m})) \tag{S.6.115}$$

$$= (\mathbf{y}_i - \mathbf{y})^t (\mathbf{y}_i - \mathbf{y}) \tag{S.6.116}$$

と書ける．同様にして，式 (S.6.111) は

$$D_M^2(\mathbf{m}_i, \mathbf{m}_j) = (\mathbf{y}_i - \mathbf{y}_j)^t (\mathbf{y}_i - \mathbf{y}_j) \tag{S.6.117}$$

と書ける．式 (6.181) に式 (S.6.116)，(S.6.117) を代入することにより，証明す

べき式 (6.181) は

$$\sum_{i=1}^{c} P(\omega_i)(\mathbf{y}_i - \mathbf{y})^t(\mathbf{y}_i - \mathbf{y})$$
$$= \sum_{i=1}^{c}\sum_{j<i} P(\omega_i)P(\omega_j)(\mathbf{y}_i - \mathbf{y}_j)^t(\mathbf{y}_i - \mathbf{y}_j) \tag{S.6.118}$$

と書き直すことができる.

一方，問 (3) ですでに証明したように，式 (6.154)，(6.156) より下式が成り立つ.

$$\sum_{i=1}^{c} P(\omega_i)(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t$$
$$= \sum_{i=1}^{c}\sum_{j<i} P(\omega_i)P(\omega_j)(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^t \tag{S.6.119}$$

上式の両辺の trace をとって付録の式 (A.2.11) を用いると，下式のごとく行列の演算がベクトルの内積演算に変換される.

$$\sum_{i=1}^{c} P(\omega_i)(\mathbf{m}_i - \mathbf{m})^t(\mathbf{m}_i - \mathbf{m})$$
$$= \sum_{i=1}^{c}\sum_{j<i} P(\omega_i)P(\omega_j)(\mathbf{m}_i - \mathbf{m}_j)^t(\mathbf{m}_i - \mathbf{m}_j) \tag{S.6.120}$$

式 (S.6.118) は，式 (S.6.120) において $\mathbf{m}_i \to \mathbf{y}_i$，$\mathbf{m}_j \to \mathbf{y}_j$，$\mathbf{m} \to \mathbf{y}$ なる変換を施すことによって得られるので，式 (S.6.118) が成り立つことは明らかである. したがって，式 (6.181) が証明された.

【演習問題 6.7】

最初に第一の式，$\mathbf{A}^t\mathbf{S}\mathbf{A} = \mathbf{I}$ を導出する．実対称行列 $\mathbf{S}$ は直交行列 $\mathbf{U}$ によって次式のごとく対角化できる.

$$\mathbf{U}^t\mathbf{S}\mathbf{U} = \mathbf{\Lambda} \tag{S.6.121}$$

ここで $\mathbf{\Lambda}$ は，$\lambda_1, \ldots, \lambda_d$ を対角成分とする対角行列で，下式で表される．

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{pmatrix} \tag{S.6.122}$$

また，直交行列の定義より，下式が成り立つ．

$$\mathbf{U}^{-1} = \mathbf{U}^t \tag{S.6.123}$$

ここで，$\lambda_1^{-1/2}, \ldots, \lambda_d^{-1/2}$ を対角成分とする対角行列 $\mathbf{\Lambda}_0$ を定義する．すなわち

$$\mathbf{\Lambda}_0 \stackrel{\text{def}}{=} \begin{pmatrix} \lambda_1^{-1/2} & & & 0 \\ & \lambda_2^{-1/2} & & \\ & & \ddots & \\ 0 & & & \lambda_d^{-1/2} \end{pmatrix} \tag{S.6.124}$$

である．式 (S.6.121) の両辺に，左から $\mathbf{\Lambda}_0^t$ $(= \mathbf{\Lambda}_0)$，右から $\mathbf{\Lambda}_0$ を乗ずると

$$\mathbf{\Lambda}_0^t \mathbf{U}^t \mathbf{S} \mathbf{U} \mathbf{\Lambda}_0 = \mathbf{\Lambda}_0^t \mathbf{\Lambda} \mathbf{\Lambda}_0 = \mathbf{I} \tag{S.6.125}$$

が得られる．ここで

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda}_0 \tag{S.6.126}$$

と置くと，式 (S.6.125) は

$$\mathbf{A}^t \mathbf{S} \mathbf{A} = \mathbf{I} \tag{S.6.127}$$

と書き換えられ，第一の式が導出できた．

次に第二の式，$\mathbf{S} = \mathbf{B}^t \mathbf{B}$ を導出する．式 (S.6.127) の両辺に，左から $(\mathbf{A}^t)^{-1}$，右から $\mathbf{A}^{-1}$ を乗ずることにより $\mathbf{S}$ を求めると

$$\mathbf{S} = \left(\mathbf{A}^t\right)^{-1} \mathbf{A}^{-1} \tag{S.6.128}$$

$$= \left(\mathbf{A}^{-1}\right)^t \mathbf{A}^{-1} \tag{S.6.129}$$

となる．ここで

$$\mathbf{B} = \mathbf{A}^{-1} \tag{S.6.130}$$

と置くと，式 (S.6.129) は

$$\mathbf{S} = \mathbf{B}^t\mathbf{B} \tag{S.6.131}$$

と書き換えられ，第二の式が導出できた．なお，$\mathbf{B}$ は式 (S.6.123)，(S.6.126) より

$$\mathbf{B} = \mathbf{A}^{-1} = (\mathbf{U}\mathbf{\Lambda}_0)^{-1} = \mathbf{\Lambda}_0^{-1}\mathbf{U}^t \tag{S.6.132}$$

であり，上式の $\mathbf{\Lambda}_0^{-1}$ は式 (S.6.124) より

$$\mathbf{\Lambda}_0^{-1} = \begin{pmatrix} \lambda_1^{1/2} & & & 0 \\ & \lambda_2^{1/2} & & \\ & & \ddots & \\ 0 & & & \lambda_d^{1/2} \end{pmatrix} \tag{S.6.133}$$

として求められる．

【演習問題 6.8】

問 (1)

与えられたパターンを 2 次元特徴空間上にプロットすると，**図 S.6.4**(a) に示すように楕円形の分布となることがわかる．これらのパターンから各種平均ベクトルを求めると

$$\begin{aligned} \mathbf{m}_1 &= \frac{1}{5}\sum_{p=1}^{5}\boldsymbol{x}_p = \frac{1}{5}(1+3+2-1+0, 0+2-1-2+1)^t \\ &= (1,\ 0)^t \end{aligned} \tag{S.6.134}$$

となり，同様にして以下が得られる．

$$\mathbf{m}_2 = \frac{1}{5}\sum_{p=6}^{10}\boldsymbol{x}_p = (-1,\ 0)^t \tag{S.6.135}$$

$$\mathbf{m} = \frac{1}{10}\sum_{p=1}^{10}\boldsymbol{x}_p = (0,\ 0)^t \tag{S.6.136}$$

定義に従って変動行列，共分散行列を求めると，式 (6.79) より

$$\begin{aligned}
\mathbf{S}_1 &= \sum_{p=1}^{5}(\boldsymbol{x}_p - \mathbf{m}_1)(\boldsymbol{x}_p - \mathbf{m}_1)^t \\
&= \begin{pmatrix} 0 \\ 0 \end{pmatrix}(0,\ 0) + \begin{pmatrix} 2 \\ 2 \end{pmatrix}(2,\ 2) + \begin{pmatrix} 1 \\ -1 \end{pmatrix}(1,\ -1) \\
&\quad + \begin{pmatrix} -2 \\ -2 \end{pmatrix}(-2,\ -2) + \begin{pmatrix} -1 \\ 1 \end{pmatrix}(-1,\ 1) \\
&= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\
&\quad + \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \\
&= 2\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}
\end{aligned} \tag{S.6.137}$$

となる．同様にして

$$\mathbf{S}_2 = 2\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} = \mathbf{S}_1 \tag{S.6.138}$$

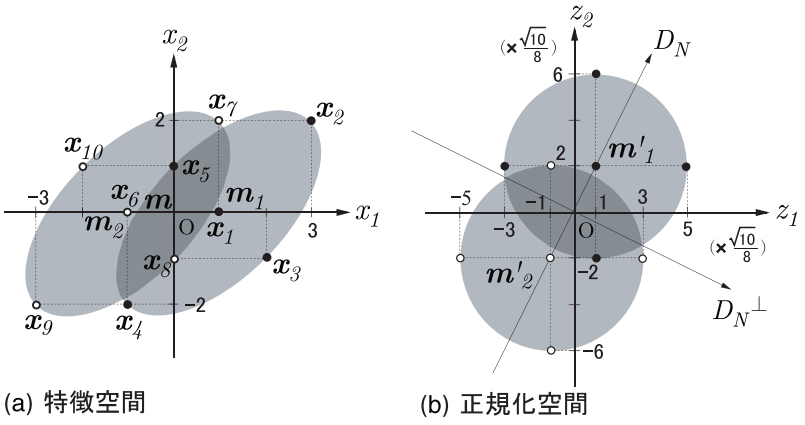


図 S.6.4: 2 クラスの特徴ベクトルの分布

と求められる．また，式 (6.112) より，下式が得られる．

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \frac{1}{5}\mathbf{S}_1 = \frac{2}{5}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \tag{S.6.139}$$

さらに，式 (6.162)，(6.160) より下式が得られる．

$$\mathbf{S}_T = \sum_{p=1}^{10}(\boldsymbol{x}_p - \mathbf{m})(\boldsymbol{x}_p - \mathbf{m})^t = 2\begin{pmatrix} 15 & 6 \\ 6 & 10 \end{pmatrix} \tag{S.6.140}$$

$$\boldsymbol{\Sigma}_T = \frac{1}{10}\sum_{p=1}^{10}(\boldsymbol{x}_p - \mathbf{m})(\boldsymbol{x}_p - \mathbf{m})^t = \frac{1}{5}\begin{pmatrix} 15 & 6 \\ 6 & 10 \end{pmatrix} = \frac{1}{10}\mathbf{S}_T \tag{S.6.141}$$

式 (6.80) より

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 = 4\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \tag{S.6.142}$$

が得られ，式 (6.82) より

$$\begin{aligned}\mathbf{S}_B &= 5\,(\mathbf{m}_1 - \mathbf{m})(\mathbf{m}_1 - \mathbf{m})^t + 5\,(\mathbf{m}_2 - \mathbf{m})(\mathbf{m}_2 - \mathbf{m})^t \\ &= 10\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\end{aligned} \tag{S.6.143}$$

が得られる．ここで，$P(\omega_i) = n_i/n \quad (i = 1, 2)$ であるので，式 (6.113)，(6.115) より，共分散行列はそれぞれ

$$\boldsymbol{\Sigma}_W = \sum_{i=1,2} P(\omega_i)\boldsymbol{\Sigma}_i = \frac{5}{10}\boldsymbol{\Sigma}_1 + \frac{5}{10}\boldsymbol{\Sigma}_2 = \frac{2}{5}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix} \tag{S.6.144}$$

$$\boldsymbol{\Sigma}_B = \sum_{i=1,2} P(\omega_i)(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{S.6.145}$$

となる．上式において式 (6.161)，(6.163) が成立している．また，$\mathbf{m} = \mathbf{0}$ であるので式 (6.56) より次式が成り立つ．

$$\mathbf{R} = \boldsymbol{\Sigma}_T + \mathbf{m}\mathbf{m}^t = \boldsymbol{\Sigma}_T \tag{S.6.146}$$

問 (2)

全共分散行列 $\mathbf{\Sigma}_T$ の固有値 λ_1，λ_2 と，それらに対応する正規直交固有ベクトル $\mathbf{u}_1$，$\mathbf{u}_2$ を求める．これらは固有値問題

$$\mathbf{\Sigma}_T\mathbf{u} = \lambda\mathbf{u} \tag{S.6.147}$$

の解として求められる．固有値 λ は，$\mathbf{\Sigma}_T$ の固有方程式

$$\det(\lambda\mathbf{I} - \mathbf{\Sigma}_T) = 0 \tag{S.6.148}$$

の根（特性根）である．そこで，$\mathbf{\Sigma}_T$ の固有方程式を求めると，λ に関する二次方程式

$$(\lambda - 3)(\lambda - 2) - \frac{36}{25} = 0 \tag{S.6.149}$$

となり，これを解くと λ の解として

$$\lambda_1 = \frac{19}{5}, \quad \lambda_2 = \frac{6}{5} \tag{S.6.150}$$

が得られる．本結果を式 (S.6.147) に代入し，さらに正規化条件

$$\|\mathbf{u}\| = 1 \tag{S.6.151}$$

を適用することにより $\mathbf{u}$ の解として

$$\mathbf{u}_1 = \frac{1}{\sqrt{13}}\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{13}}\begin{pmatrix} 2 \\ -3 \end{pmatrix} \tag{S.6.152}$$

が得られる．したがって，6.3 節で述べた KL 展開によって定まる主軸は $(3, 2)$ 方向であり，**図 S.6.5**(a) の P となる．これはパターン全体の分布の主軸である．

問 (3)

$\mathbf{S}_W$，$\mathbf{\Sigma}_W$ の逆行列を求めると

$$\mathbf{S}_W^{-1} = \frac{1}{64}\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}, \quad \mathbf{\Sigma}_W^{-1} = \frac{5}{32}\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix} \tag{S.6.153}$$

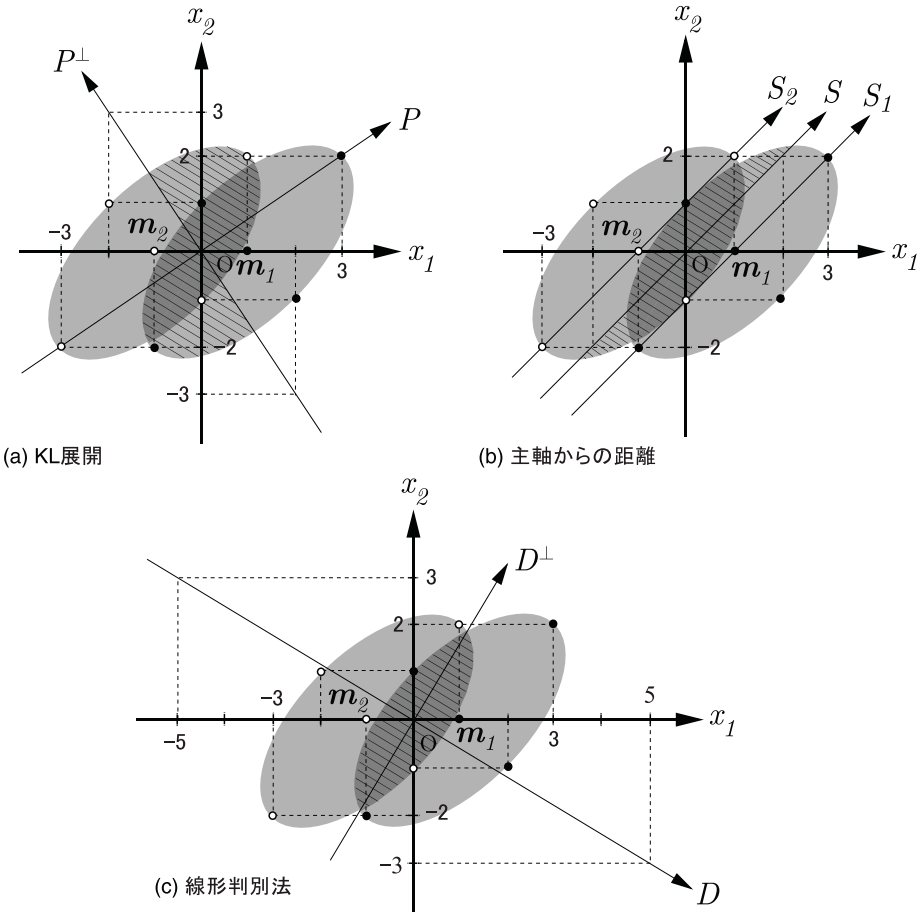


図 S.6.5: 種々の線形変換によって得られる 1 次元部分空間と決定境界

となる[*10]．本問は 2 クラス問題であるので，フィッシャーの方法を適用できる．すなわち，式 (6.106) より，

$$\boldsymbol{w} \propto \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \frac{1}{64}\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

[*10] 大きさ 2×2 の行列の逆行列の求め方は，**演習問題 6.3** を参照のこと．

$$= \frac{1}{32}\begin{pmatrix} 5 \\ -3 \end{pmatrix} \tag{S.6.154}$$

となるので，$\|\boldsymbol{w}\| = 1$ となるよう正規化すると

$$\boldsymbol{w} = \frac{1}{\sqrt{34}}\begin{pmatrix} 5 \\ -3 \end{pmatrix} \tag{S.6.155}$$

が得られる．このベクトル $\boldsymbol{w}$ の方向は**図 S.6.5**(c) における D の方向であり，クラス間分散・クラス内分散比最大という基準において二つのクラスを識別するのに最も適した軸である．

一方，以下のように式 (6.103) の固有値問題を解くことにより，最適な判別軸を求めることもできる．まず $\mathbf{S}_W^{-1}\mathbf{S}_B$ を求めると

$$\mathbf{S}_W^{-1}\mathbf{S}_B = \boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_B = \frac{5}{32}\begin{pmatrix} 5 & 0 \\ -3 & 0 \end{pmatrix} \tag{S.6.156}$$

となる．式 (S.6.156) の左側の等号が成立するのは $P(\omega_i) = n_i/n$ と仮定したためである．このとき $\boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_B$ の固有値と正規直交固有ベクトルを求めると

$$\lambda_1 = \frac{25}{32}, \quad \lambda_2 = 0 \tag{S.6.157}$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{34}}\begin{pmatrix} 5 \\ -3 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{S.6.158}$$

となる．最大固有値 λ_1 に対応する固有ベクトル $\mathbf{u}_1$ は，式 (S.6.155) の $\boldsymbol{w}$ と一致する．

線形判別による部分空間である 1 次元判別軸 D を $\boldsymbol{\Sigma}_W^{-1}\boldsymbol{\Sigma}_B$ の固有ベクトルとして求めたが，式 (6.127)，(6.128) より $\boldsymbol{\Sigma}_W$ または $\boldsymbol{\Sigma}_T$ を用いて

$$\boldsymbol{w} \propto \boldsymbol{\Sigma}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \frac{5}{32}\begin{pmatrix} 5 & -3 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \frac{5}{16}\begin{pmatrix} 5 \\ -3 \end{pmatrix} \tag{S.6.159}$$

$$\boldsymbol{w} \propto \boldsymbol{\Sigma}_T^{-1}(\mathbf{m}_1 - \mathbf{m}_2) = \frac{5}{114}\begin{pmatrix} 10 & -6 \\ -6 & 15 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \frac{10}{57}\begin{pmatrix} 5 \\ -3 \end{pmatrix} \tag{S.6.160}$$

と求めることもできる．

問 (4)

各クラスの平均と全平均とのマハラノビス汎距離の二乗値を平均すると

$$\sum_{i=1}^{c} P(\omega_i) D_M^2(\mathbf{m}_i, \mathbf{m}) = \sum_{i=1}^{c} P(\omega_i)(\mathbf{m}-\mathbf{m}_i)^t \mathbf{\Sigma}_W^{-1}(\mathbf{m}-\mathbf{m}_i) \quad \text{(S.6.161)}$$

$$= \frac{25}{32} \quad \text{(S.6.162)}$$

となり，$\mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_B$ の最大固有値（式 (S.6.157) の λ_1）に等しく，式 (6.142) が成り立っていることがわかる．一方，クラス ω_1，ω_2 の平均間マハラノビス汎距離は

$$D_M^2(\mathbf{m}_1, \mathbf{m}_2) = \begin{pmatrix} 2, & 0 \end{pmatrix} \mathbf{\Sigma}_W^{-1} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \frac{25}{8} \quad \text{(S.6.163)}$$

となる．

問 (5)

1) 第一の変換行列 $\mathbf{A}_1$

6.4 節 ［3］で述べたように，線形判別によって求まる変換行列 $\mathbf{A}$ は，クラス内分散の正規化を行う変換，すなわち $\mathbf{\Sigma}_W$ を白色化する (d, d) 行列 $\mathbf{A}_1$ と，正規化された空間での二つのクラス平均の点に対し KL 展開を行う $(d, \tilde{d})$ 行列 $\mathbf{A}_2$ とに分解することができ

$$\mathbf{A} = \mathbf{A}_1 \mathbf{A}_2 \quad \text{(S.6.164)}$$

と書ける．クラス内共分散行列 $\mathbf{\Sigma}_W$ の固有値と正規直交固有ベクトルを求めると

$$\lambda_1 = \frac{16}{5}, \quad \lambda_2 = \frac{4}{5} \quad \text{(S.6.165)}$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{(S.6.166)}$$

となる．式 (6.182) を満たすための変換行列 $\mathbf{A}_1$ は

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{u}_1, & \mathbf{u}_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{-1/2} = \frac{\sqrt{10}}{8} \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} \quad \text{(S.6.167)}$$

と求められる[*11]．原空間上の点 $\boldsymbol{x} = (x_1, x_2)^t$ を $\mathbf{A}_1$ で変換した点，すなわち正規化空間上での座標 $\mathbf{z} = (z_1, z_2)^t$ は

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \mathbf{A}_1^t \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{\sqrt{10}}{8} \begin{pmatrix} x_1 + x_2 \\ 2x_1 - 2x_2 \end{pmatrix} \tag{S.6.168}$$

で与えられる．**図 S.6.4**(b) は，図 S.6.4(a) に示した原空間の分布を $\mathbf{A}_1$ によって正規化空間上へ変換した分布である．

2) 第二の変換行列 $\mathbf{A}_2$

正規化空間上での各クラスの平均 $\mathbf{m}'_1$，$\mathbf{m}'_2$ （この例の場合，$c = 2$ であるからたかだか 2 個の点）に対して KL 展開を行って主軸を求めるためには，正規化空間におけるクラス平均の共分散行列 $\boldsymbol{\Sigma}_N$ の固有ベクトルを求めればよい．したがって，

$$\boldsymbol{\Sigma}_N = \frac{1}{c}\sum_{i=1}^{c}(\mathbf{m}'_i - \mathbf{m}')(\mathbf{m}'_i - \mathbf{m}')^t = \mathbf{A}_1^t \boldsymbol{\Sigma}_B \mathbf{A}_1 = \frac{5}{32}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \tag{S.6.169}$$

となり，この固有値と正規直交固有ベクトルは

$$\lambda_1 = \frac{25}{32}, \quad \lambda_2 = 0 \tag{S.6.170}$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} \tag{S.6.171}$$

である．したがって, 線形判別によって決まる軸は, 正規化空間上では図 S.6.4(b) の D_N である．このとき

$$\mathbf{A}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{S.6.172}$$

$$\mathbf{A} = \mathbf{A}_1\mathbf{A}_2 = \frac{\sqrt{2}}{8}\begin{pmatrix} 5 \\ -3 \end{pmatrix} \tag{S.6.173}$$

となる．また

$$\boldsymbol{\Lambda} \stackrel{\text{def}}{=} \begin{pmatrix} \lambda_1 \end{pmatrix} = \begin{pmatrix} \frac{25}{32} \end{pmatrix} \tag{S.6.174}$$

[*11] 変換行列 $\mathbf{A}_1$ が式 (S.6.167) の形になることは，**演習問題 6.7** を参照のこと．

とすると

$$\mathbf{A}^t\mathbf{\Sigma}_B\mathbf{A} = \frac{25}{32} = \mathbf{\Lambda} \tag{S.6.175}$$

となり，式 (6.192) が成立する．

$\mathbf{A}_1$ が直交変換でないため D_N は $\mathbf{m}_1'$，$\mathbf{m}_2'$ を通るが (図 S.6.4(b))，原空間上で D が $\mathbf{m}_1$，$\mathbf{m}_2$ を通るとは限らない（図 S.6.5(c)）ことに注意が必要である．正規化空間上で D_N に垂直な方向 $D_N^{\perp}$ (式 (S.6.171) の $\mathbf{u}_2$) は，原空間上では次式で示される $D^{\perp}$ となる．

$$(\mathbf{A}_1^{-1})^t\mathbf{u}_2 = \frac{\sqrt{10}}{5}\begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{\sqrt{2}}{5}\begin{pmatrix} 3 \\ 5 \end{pmatrix} \tag{S.6.176}$$

この $D^{\perp}$ は，原空間上で D への射影が同じ点になる方向であり，確かに D と直交している（図 S.6.5(c)）．

問 (6)

二つの分布の共分散行列，事前確率が等しいとすると，その 2 クラスを識別するための境界は，変換後の 1 次元空間上でクラス平均の中点，すなわち全平均に定めるのが最も良いと考えられる．

線形判別法によって定まる 1 次元空間において，こうして求めた最適な決定境界は図 S.6.5(c) の $D^{\perp}$ である．また，KL 展開よって定まる同様の決定境界は図 S.6.5(a) の $P^{\perp}$ である．一方，最も単純な識別法として各クラスの分布の主軸 S_1，S_2 からの距離によって識別した場合が (b) であり，その境界は S となる[*12]．

試みに，こうして定まった (a)，(b)，(c) の三つの決定境界を比べてみると，与えられた 10 個の特徴ベクトルに対して誤識別をするのはどれにおいても各クラス 1 パターンずつ計 2 パターンである．しかしながら，分布全体の中で誤識別を引き起こす領域 (図中斜線部分) を比べてみると，この例では，(c) の $D^{\perp}$ が最も良い決定境界を与えていることがわかる．

[*12] より一般的には，各クラスごとにクラス中心を原点とした部分空間 (この例では 1 次元空間 S_1 と S_2) を KL 展開により定め，未知パターンから各部分空間への投影距離の大小によって属するクラスを識別する方法であり，投影距離法 [池田 83] と呼ばれる．

また，第 7 章で述べる部分空間法をこの例に適用してみると，クラス自己相関行列は式 (7.15) より

$$\mathbf{R}_1 = \frac{1}{5}\sum_{p=1}^{5} \boldsymbol{x}\boldsymbol{x}^t = \frac{1}{5}\begin{pmatrix} 15 & 6 \\ 6 & 10 \end{pmatrix} \tag{S.6.177}$$

$$\mathbf{R}_2 = \frac{1}{5}\sum_{p=6}^{10} \boldsymbol{x}\boldsymbol{x}^t = \frac{1}{5}\begin{pmatrix} 15 & 6 \\ 6 & 10 \end{pmatrix} \tag{S.6.178}$$

となる．すなわち

$$\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{\Sigma}_T \tag{S.6.179}$$

となることから，各クラスに対して求まる 1 次元軸はともに (b) の S であり，どのパターンに対しても識別不能となる．本例のように各クラスの部分空間が互いに交差しているような場合には部分空間法は原理的に向かない．

第 7 章

【演習問題 7.1】

行列 $\mathbf{X}^t\mathbf{X}$ は $n \times n$ の対称行列であり，その固有値，固有ベクトルが λ_i，$\mathbf{e}_i$ であることから

$$\mathbf{X}^t\mathbf{X}\mathbf{e}_i = \lambda_i \mathbf{e}_i \qquad (i = 1, \ldots, n) \tag{S.7.1}$$

が成り立つ．また，対称行列の固有ベクトルは互いに直交するので

$$\mathbf{e}_i{}^t\mathbf{e}_j = 0 \qquad (i \neq j) \tag{S.7.2}$$

が成り立つ．さらに $\mathbf{e}_i$ を正規直交基底とするため

$$\|\mathbf{e}_i\| = 1 \qquad (i = 1, \ldots, n) \tag{S.7.3}$$

と正規化されているものとする．

式 (S.7.1) の両辺に左から $\mathbf{X}$ を乗じて

$$\mathbf{X}\mathbf{X}^t\mathbf{X}\mathbf{e}_i = \lambda_i \mathbf{X}\mathbf{e}_i \qquad (i = 1, \ldots, n) \tag{S.7.4}$$

となるので，

$$\mathbf{u}_i = c_i \mathbf{X}\mathbf{e}_i \tag{S.7.5}$$

と置くと

$$\mathbf{X}\mathbf{X}^t\mathbf{u}_i = \lambda_i \mathbf{u}_i \qquad (i = 1, \ldots, n) \tag{S.7.6}$$

が得られる．ただし，式 (S.7.5) の c_i は定数である．上式は，$\mathbf{X}\mathbf{X}^t$ の固有値と固有ベクトルが，それぞれ λ_i と $\mathbf{u}_i$ であることを示している．ここで $\mathbf{X}\mathbf{X}^t$ も対称行列であるので，異なる固有ベクトル $\mathbf{u}_i$，$\mathbf{u}_j$ は直交する．これらを正規直交基底とするためには，式 (S.7.3) と同様，そのノルムを 1 とすべく

$$\begin{aligned}
\|\mathbf{u}_i\|^2 &= \mathbf{u}_i^t\mathbf{u}_i \\
&= (c_i\mathbf{X}\mathbf{e}_i)^t\,(c_i\mathbf{X}\mathbf{e}_i) && \text{(S.7.7)} \\
&= c_i^2\ \mathbf{e}_i^t\,\mathbf{X}^t\mathbf{X}\mathbf{e}_i && \text{(S.7.8)} \\
&= c_i^2\ \mathbf{e}_i^t \cdot \lambda_i\mathbf{e}_i && \text{(S.7.9)} \\
&= \lambda_i\, c_i^2 = 1 && \text{(S.7.10)}
\end{aligned}$$

が成り立たなくてはならない．式 (S.7.8) から (S.7.9) への変形には式 (S.7.1) を，また式 (S.7.9) から (S.7.10) への変形には式 (S.7.3) をそれぞれ用いた．

式 (S.7.10) より

$$c_i = \frac{1}{\sqrt{\lambda_i}} \tag{S.7.11}$$

となり，この結果を式 (S.7.5) に代入することにより

$$\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{X}\mathbf{e}_i \tag{S.7.12}$$

が得られる．式の導出は以上である．以下では，本方法が計算の効率化に役立つことを述べる．

行列 $\mathbf{X}$ を構成する n 個の d 次元列ベクトルを $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ とし，これらを d 次元特徴ベクトルのパターン集合とみなす．すなわち

$$\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \tag{S.7.13}$$

と表せる．これより

$$\mathbf{X}\mathbf{X}^t = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \begin{pmatrix} \boldsymbol{x}_1^t \\ \boldsymbol{x}_2^t \\ \vdots \\ \boldsymbol{x}_n^t \end{pmatrix} = \sum_{p=1}^{n} \boldsymbol{x}_p \boldsymbol{x}_p^t \tag{S.7.14}$$

が得られる．上式は定数項を除いて式 (7.15) と一致し，$\mathbf{X}\mathbf{X}^t$ は自己相関行列とみなせる．第 7 章で述べたように，部分空間法では自己相関行列の固有値，固有ベクトルを求めなくてはならない．式 (S.7.14) の自己相関行列の大きさは $d \times d$ である．

一方，行列 $\mathbf{X}^t\mathbf{X}$ の大きさは $n \times n$ であり，n がd より小さい場合には $\mathbf{X}^t\mathbf{X}$ の固有値，固有ベクトルを求める計算量は $\mathbf{X}\mathbf{X}^t$ の場合に比して少なくて済む．しかも式 (S.7.12) によって，$\mathbf{X}^t\mathbf{X}$ の固有ベクトルから自己相関行列の固有ベクトルに変換できる．以上の手法により，自己相関行列の固有値，固有ベクトルをより効率的に求めることができる．

【演習問題 7.2】

クラス ω の部分空間を張る正規直交ベクトル $\mathbf{u}_j$ は，クラス ω に属するパターンの自己相関行列 $\mathbf{R}$ を作成し，固有値展開することで計算できる（ 7.2 節 ）．ここで $\mathbf{R}$ の固有値，固有ベクトルをそれぞれ λ_j，$\mathbf{u}_j$ とすると

$$\mathbf{R}\mathbf{u}_j = \lambda_j \mathbf{u}_j \qquad (j = 1, \ldots, d) \tag{S.7.15}$$

が成り立つ．ただし，$\lambda_1 \geq \lambda_2, \ldots$ とする．自己相関行列の定義より

$$\mathbf{R} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_k \mathbf{y}_k^t = \frac{1}{n}(\mathbf{y}_1, \ldots, \mathbf{y}_n)(\mathbf{y}_1, \ldots, \mathbf{y}_n)^t \tag{S.7.16}$$

である．また，$\mathbf{R}$ はその固有値，固有ベクトル λ_j，$\mathbf{u}_j$ を用いて，下記のように表せる．

$$\mathbf{R} = \sum_{j=1}^{d} \lambda_j \mathbf{u}_j \mathbf{u}_j^t \tag{S.7.17}$$

$$= (\mathbf{u}_1, \ldots, \mathbf{u}_d) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} (\mathbf{u}_1, \ldots, \mathbf{u}_d)^t \tag{S.7.18}$$

式 (S.7.17) は，$\mathbf{R}$ をスペクトル分解した結果であり，式 (S.7.15) が成り立つことは明らかである．

ここで，値の小さい固有値に対応する固有ベクトル成分を省略すれば，近似的に，

$$\mathbf{R} \approx (\mathbf{u}_1, \ldots, \mathbf{u}_{d'}) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{d'} \end{pmatrix} (\mathbf{u}_1, \ldots, \mathbf{u}_{d'})^t \tag{S.7.19}$$

となる．そのため，$\mathbf{y}_k$ と $\mathbf{u}_j$ は次の関係となる．

$$\frac{1}{n}(\mathbf{y}_1, \ldots, \mathbf{y}_n)(\mathbf{y}_1, \ldots, \mathbf{y}_n)^t$$

$$= (\mathbf{u}_1, \ldots, \mathbf{u}_d) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} (\mathbf{u}_1, \ldots, \mathbf{u}_d)^t \tag{S.7.20}$$

$$\approx (\mathbf{u}_1, \ldots, \mathbf{u}_{d'}) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{d'} \end{pmatrix} (\mathbf{u}_1, \ldots, \mathbf{u}_{d'})^t \tag{S.7.21}$$

【演習問題 7.3】

複合類似度 (単純化した式) を，$\boldsymbol{x}^t\mathbf{u}_j$ がスカラーであることに注意し，以下のように変換する．

$$S(\boldsymbol{x}) = \sum_{j=1}^{d} \lambda_j (\boldsymbol{x}^t\mathbf{u}_j)^2 \tag{S.7.22}$$

$$= \sum_{j=1}^{d} \lambda_j (\boldsymbol{x}^t\mathbf{u}_j)(\boldsymbol{x}^t\mathbf{u}_j)^t \tag{S.7.23}$$

$$= (\boldsymbol{x}^t\mathbf{u}_1, \ldots, \boldsymbol{x}^t\mathbf{u}_d) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} (\boldsymbol{x}^t\mathbf{u}_1, \ldots, \boldsymbol{x}^t\mathbf{u}_d)^t \tag{S.7.24}$$

$$= \boldsymbol{x}^t(\mathbf{u}_1, \ldots, \mathbf{u}_d) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} (\mathbf{u}_1, \ldots, \mathbf{u}_d)^t\boldsymbol{x} \tag{S.7.25}$$

$$= \boldsymbol{x}^t\mathbf{R}\boldsymbol{x} \tag{S.7.26}$$

$$= \frac{1}{n}\boldsymbol{x}^t(\mathbf{y}_1, \ldots, \mathbf{y}_n)(\mathbf{y}_1, \ldots, \mathbf{y}_n)^t\boldsymbol{x} \tag{S.7.27}$$

$$= \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{x}^t\mathbf{y}_k)^2 \tag{S.7.28}$$

上式で，式 (S.7.25) から式 (S.7.26) への変形には式 (S.7.18) を，式 (S.7.26) から式 (S.7.27) への変形には式 (S.7.16) を用いた．

つまり，複合類似度 (単純化した式) は，入力ベクトル $\boldsymbol{x}$ と，n 個のパターン $\mathbf{y}_k(k = 1, \ldots, n)$ との間の内積の二乗平均を意味している．

【演習問題 7.4】

パターン $\mathbf{y}_k(k = 1, \ldots, n)$ を部分空間 $\mathbf{u}_1$ へ射影したベクトル長の二乗は $(\mathbf{u}_1^t\mathbf{y}_k)^2$ となるため，その平均値は $\frac{1}{n}\sum_{k=1}^{n}(\mathbf{u_1}^t\mathbf{y}_k)^2$ となる．そこで，$\mathbf{u}_1$ の長

さが 1 であるという制約の下でラグランジュ関数を作ると，$\mathbf{u}_1^t\mathbf{y}_k = \mathbf{y}_k^t\mathbf{u}_1$ はスカラーであるので

$$L(\mathbf{u}_1, \lambda) = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{u}_1^t\mathbf{y}_k)^2 - \lambda(\mathbf{u}_1^t\mathbf{u}_1 - 1) \tag{S.7.29}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\mathbf{u}_1^t\mathbf{y}_k\mathbf{y}_k^t\mathbf{u}_1 - \lambda(\mathbf{u}_1^t\mathbf{u}_1 - 1) \tag{S.7.30}$$

$$= \mathbf{u}_1^t\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{y}_k\mathbf{y}_k^t\right)\mathbf{u}_1 - \lambda(\mathbf{u}_1^t\mathbf{u}_1 - 1) \tag{S.7.31}$$

$$= \mathbf{u}_1^t\mathbf{R}\mathbf{u}_1 - \lambda(\mathbf{u}_1^t\mathbf{u}_1 - 1) \tag{S.7.32}$$

となる．ここで $\mathbf{R}$ は式 (S.7.16) で示したように, $\mathbf{y}_k(k = 1, \ldots, n)$ の自己相関行列である．ラグランジュの未定乗数法により $\mathbf{u}_1$ を計算するためには，$L(\mathbf{u}_1, \lambda)$ が極値をとる $\mathbf{u}_1$ を求めることになり，以下のように計算できる．

$$\frac{\partial L(\mathbf{u}_1, \lambda)}{\partial \mathbf{u}_1} = 2\mathbf{R}\mathbf{u}_1 - 2\lambda\mathbf{u}_1 = 0$$

$$\frac{\partial L(\mathbf{u}_1, \lambda)}{\partial \lambda} = -(\mathbf{u}_1^t\mathbf{u}_1 - 1) = 0$$

つまり

$$\mathbf{R}\mathbf{u}_1 = \lambda\mathbf{u}_1$$

となる固有値問題を解き，最大固有値に対する固有ベクトルを長さ 1 に正規化したベクトルが $\mathbf{u}_1$ となる．

第８章

【演習問題 8.1】

図 S.8.1 に $f(w) = \cos(w)$ および $h(w)$ が，$\pi \leq w \leq 2\pi$ の範囲でプロットされている．観測できるのは，$f(w)$ にノイズが付加された $h(w)$ のみであり，それが図では黒丸で示されている．

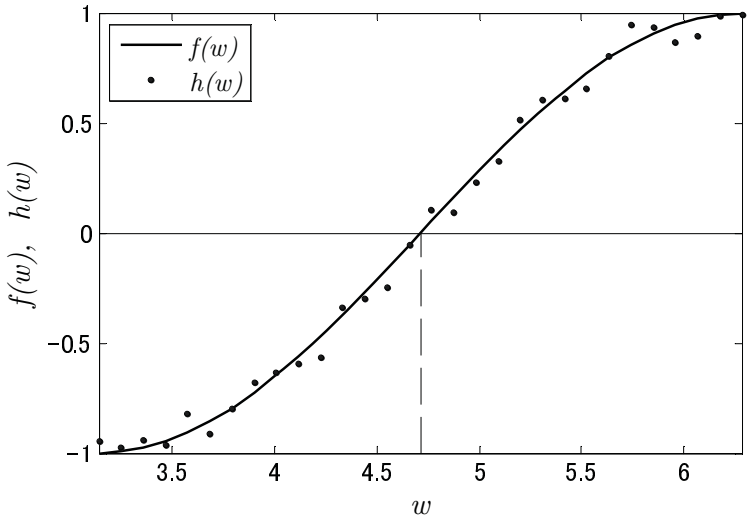


図 S.8.1: 回帰関数 $f(w)$ と観測データ $h(w)$

初期値を $w(1) = 3.5$ に設定し，$\rho(t) = 1/t$ として，式 (8.47) を $t = 1$ から $t = 10$ まで繰り返した結果を**表 S.8.1** に示す．この表からわかるように，$w(t)$ は t の増大とともに $f(w) = 0$ の根 $w = 3\pi/2 = 4.712$ に近づき，繰り返し $t = 10$ において $w(t) = 4.717$ となっている．繰り返し数 t を横軸に取り，$w(t)$ が $3\pi/2$ に近接する過程を**図 S.8.2** に示した．**図 S.8.3** は，図 S.8.1 の一部を拡大した図である．本図と表 S.8.1 を参照しながら，本アルゴリズムの処理過程を追ってみよう．

表 S.8.1: ロビンス・モンローアルゴリズムの計算結果

t	$w(t)$	$f(w(t))$	ξ	$h(w(t))$	$\rho(t)$
1	3.500	-0.936	0.045	-0.891	1.000
2	4.391	-0.316	0.049	-0.267	0.500
3	4.524	-0.187	-0.042	-0.229	0.333
4	4.601	-0.111	-0.094	-0.205	0.250
5	4.652	-0.060	-0.060	-0.120	0.200
6	4.676	-0.036	-0.025	-0.061	0.167
7	4.686	-0.026	-0.097	-0.123	0.143
8	4.704	-0.008	-0.016	-0.025	0.125
9	4.707	-0.005	-0.084	-0.090	0.111
10	4.717	0.005	-0.061	-0.056	0.100

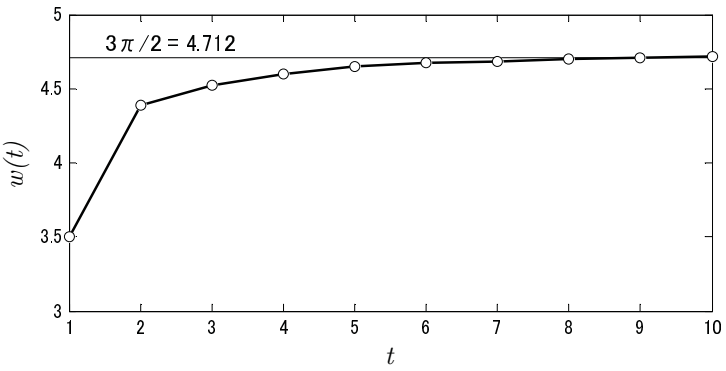


図 S.8.2: 収束過程

まず，初期値 $w(1) = 3.5$ における $f(w)$ の値は表 S.8.1 にも示されているように，$f(w(1)) = -0.936$ である．この値は，図 S.8.3 において白丸で示されている．しかし，実際観測されるのは $f(w(1))$ ではなく，ノイズ $\xi = 0.045$ が付加さ

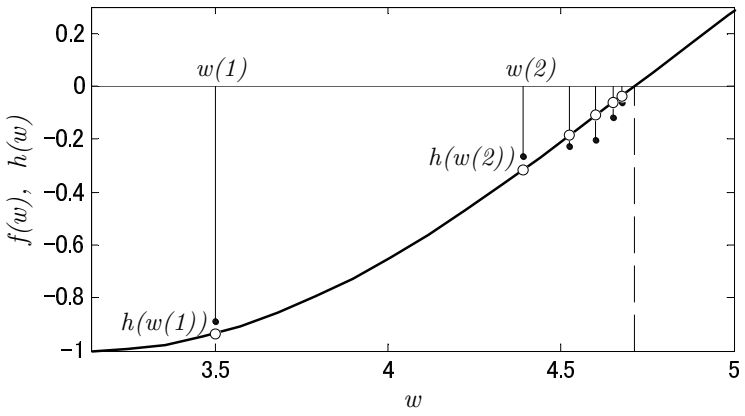


図 S.8.3: 零点探索過程

れた

$$\begin{aligned} h(w(1)) &= f(w(1)) + \xi \\ &= -0.936 + 0.045 \\ &= -0.891 \end{aligned}$$

であり，この値が同図の黒丸で示されている．ここで $t = 1$ として式 (8.47) を適用すると

$$\begin{aligned} \boldsymbol{w}(2) &= w(1) - \rho(1) \cdot h(w(1)) \\ &= 3.5 - 1 \cdot (-0.891) \\ &= 4.391 \end{aligned}$$

が得られ，この値が表 S.8.1 の $w(2)$ の値として記されている．

以下，同様の操作を繰り返すことにより，$w(3)$, $w(4)$, ..., $w(10)$ と求められていくが，図ではこれらの表記は省かれ，その位置のみが $w(6)$ まで示されている．

以上の結果より，繰り返しにより $w(t)$ が限りなく $3\pi/2$ に近づくことが確認できる．

【演習問題 8.2*】

以下では，クラス ω_i を取り上げることにする（$i = 1, \ldots, c$）．クラス ω_i のベイズ識別関数を $G_i(\boldsymbol{x})$ で表すと，式 (5.19) より

$$G_i(\boldsymbol{x}) = P(\omega_i|\boldsymbol{x}) \tag{S.8.1}$$

となる．また，クラス ω_i の線形識別関数 $g_i(\boldsymbol{x})$ は，拡張特徴ベクトル $\mathbf{x}$ と拡張重みベクトル $\mathbf{w}_i$ を用いて

$$g_i(\boldsymbol{x}) = \mathbf{w}_i^t \mathbf{x} \tag{S.8.2}$$

と書ける．

ベイズ識別関数 $G_i(\boldsymbol{x})$ を線形識別関数 $g_i(\boldsymbol{x})$ で近似することを考える．そこで両者の平均二乗誤差を $\mathbf{w}_i$ の関数として $\epsilon^2(\mathbf{w}_i)$ で表すと

$$\epsilon^2(\mathbf{w}_i) = \mathrm{E}\left\{ \left(g_i(\boldsymbol{x}) - G_i(\boldsymbol{x})\right)^2 \mid \boldsymbol{x} \right\} \tag{S.8.3}$$

$$= \mathrm{E}\left\{ \left(\mathbf{w}_i^t \mathbf{x} - G_i(\boldsymbol{x})\right)^2 \mid \boldsymbol{x} \right\} \tag{S.8.4}$$

$$= \int \left(\mathbf{w}_i^t \mathbf{x} - G_i(\boldsymbol{x})\right)^2 p(\boldsymbol{x}) d\boldsymbol{x} \tag{S.8.5}$$

と書け，最良の近似を実現するには，$\epsilon^2(\mathbf{w}_i)$ を最小にする $\mathbf{w}_i$ を求めればよい．しかし，$G_i(\boldsymbol{x})$ を直接知ることはできないので，上式を計算することはできない．そこで，以下では 3.1 節 ［1］で導入した教師信号を用いることにより，この問題の解決を図る．

すでに 3.1 節 ［1］で示したように，n 個の学習パターンのうち，p 番目のパターン $\boldsymbol{x}_p$ を識別したとき，線形識別関数 $g_i(\boldsymbol{x})$ に対しては教師信号 b_{ip} が与えられる．教師信号 b_{ip} は，$g_i(\boldsymbol{x}_p)$ として望ましい値を示しており，学習の過程では $g_i(\boldsymbol{x}_p)$ をできるだけ b_{ip} に近づけるよう，重み $\mathbf{w}_i$ の修正を行うことになる．

ここで，問題をより一般的に扱えるよう，線形識別関数 $g_i(\boldsymbol{x})$ に対する教師信号を $b_i(\boldsymbol{x})$ で表し，下式のごとく定義する．

$$b_i(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \omega_i) \\ 0 & (\boldsymbol{x} \notin \omega_i) \end{cases} \qquad \left(\boldsymbol{x} \in \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}\right) \tag{S.8.6}$$

すなわち，上で引用した b_{ip} は，$b_i(\boldsymbol{x}_p)$ を略記した教師信号である．

下式が成り立つことは明らかである．

$$\begin{cases} P\big(b_i(\boldsymbol{x}) = 1 \mid \boldsymbol{x}\big) = & P(\omega_i|\boldsymbol{x}) \\ P\big(b_i(\boldsymbol{x}) = 0 \mid \boldsymbol{x}\big) = & 1 - P(\omega_i|\boldsymbol{x}) \end{cases} \tag{S.8.7}$$

したがって，$\boldsymbol{x}$ が与えられたときの $b_i(\boldsymbol{x})$ の期待値 $\mathrm{E}\left\{b_i(\boldsymbol{x})|\boldsymbol{x}\right\}$ は

$$\begin{aligned} \mathrm{E}\left\{b_i(\boldsymbol{x})|\boldsymbol{x}\right\} &= \sum_{b_i(\boldsymbol{x})} b_i(\boldsymbol{x}) P\big(b_i(\boldsymbol{x}) \mid \boldsymbol{x}\big) \\ &= 1 \cdot P(\omega_i|\boldsymbol{x}) + 0 \cdot (1 - P(\omega_i|\boldsymbol{x})) \\ &= P(\omega_i|\boldsymbol{x}) \\ &= G_i(\boldsymbol{x}) \end{aligned} \tag{S.8.8}$$

となり，ベイズ識別関数と一致する．

ここで，式 (S.8.8) について考察してみよう．あるパターン $\boldsymbol{x}$ が与えられたとき，教師信号 $b_i(\boldsymbol{x})$ としては $b_i(\boldsymbol{x}) = P(\omega_i|\boldsymbol{x}) < 1$ とするのが望ましい．教師信号が 1 より小さいのは，パターン $\boldsymbol{x}$ が他のクラス ω_j $(j \neq i)$ に属する可能性もあるからである．特に，パターン $\boldsymbol{x}$ が特徴空間上でクラス ω_i とクラス ω_j の境界近くに存在している場合にはその可能性はさらに高まる．しかし，学習パターンは，通常その所属クラス名と共に与えられるので，教師信号は常に 1 または 0 である．先に述べたように，$G_i(\boldsymbol{x})$ を直接知る手段が無いので，学習過程では $G_i(\boldsymbol{x})$ を二値の教師信号 $b_i(\boldsymbol{x})$ で代用しているわけである．

このように考えると，式 (S.8.8) は次のように解釈することができる．すなわち，1 または 0 をとる教師信号 $b_i(\boldsymbol{x})$ は，$G_i(\boldsymbol{x})$ にノイズ ξ が重畳した信号

$$b_i(\boldsymbol{x}) = G_i(\boldsymbol{x}) + \xi \tag{S.8.9}$$

であり，かつ $\mathrm{E}\{\xi\} = 0$ であるので，式 (S.8.8) が成り立つ．

以上述べたことを踏まえた上で，式 (S.8.4) の $G_i(\boldsymbol{x})$ を $b_i(\boldsymbol{x})$ に置き換え，$g_i(\boldsymbol{x})$ と $b_i(\boldsymbol{x})$ の平均二乗誤差 $\epsilon_0^2(\mathbf{w}_i)$ を求めてみると

$$\begin{aligned} &\epsilon_0^2(\mathbf{w}_i) \\ &= \mathrm{E}\left\{\left(\mathbf{w}_i^t\mathbf{x} - b_i(\boldsymbol{x})\right)^2 \mid \boldsymbol{x}\right\} \end{aligned} \tag{S.8.10}$$

$$= \sum_{j=1}^{c} P(\omega_j) \cdot \mathrm{E}\left\{ \left(\mathbf{w}_i^t \mathbf{x} - b_i(\boldsymbol{x})\right)^2 \mid \boldsymbol{x} \in \omega_j \right\} \tag{S.8.11}$$

$$= P(\omega_i) \cdot \mathrm{E}\left\{ \left(\mathbf{w}_i^t \mathbf{x} - 1\right)^2 \mid \boldsymbol{x} \in \omega_i \right\} + \sum_{j \neq i} P(\omega_j) \cdot \mathrm{E}\left\{ \left(\mathbf{w}_i^t \mathbf{x} - 0\right)^2 \mid \boldsymbol{x} \in \omega_j \right\} \tag{S.8.12}$$

$$= P(\omega_i) \int (\mathbf{w}_i^t \mathbf{x} - 1)^2 p(\boldsymbol{x}|\omega_i) d\boldsymbol{x} + \sum_{j \neq i} P(\omega_j) \int (\mathbf{w}_i^t \mathbf{x})^2 p(\boldsymbol{x}|\omega_j) d\boldsymbol{x} \tag{S.8.13}$$

$$= \int \left((\mathbf{w}_i^t \mathbf{x})^2 - 2\mathbf{w}_i^t \mathbf{x} + 1 \right) P(\omega_i) p(\boldsymbol{x}|\omega_i) d\boldsymbol{x} + \int \sum_{j \neq i} (\mathbf{w}_i^t \mathbf{x})^2 P(\omega_j) p(\boldsymbol{x}|\omega_j) d\boldsymbol{x} \tag{S.8.14}$$

$$= \int (\mathbf{w}_i^t \mathbf{x})^2 \sum_{j=1}^{c} P(\omega_j) p(\boldsymbol{x}|\omega_j) d\boldsymbol{x} - 2 \int \mathbf{w}_i^t \mathbf{x} P(\omega_i) p(\boldsymbol{x}|\omega_i) d\boldsymbol{x} + \int P(\omega_i) p(\boldsymbol{x}|\omega_i) d\boldsymbol{x} \tag{S.8.15}$$

$$= \int (\mathbf{w}_i^t \mathbf{x})^2 p(\boldsymbol{x}) d\boldsymbol{x} - 2 \int \mathbf{w}_i^t \mathbf{x} G_i(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} + \int G_i(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \tag{S.8.16}$$

$$= \int \left((\mathbf{w}_i^t \mathbf{x} - G_i(\boldsymbol{x}) \right)^2 p(\boldsymbol{x}) d\boldsymbol{x} - \int G_i(\boldsymbol{x})^2 p(\boldsymbol{x}) d\boldsymbol{x} + \int G_i(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} \tag{S.8.17}$$

$$= \epsilon^2(\mathbf{w}_i) + \int G_i(\boldsymbol{x}) \left(1 - G_i(\boldsymbol{x})\right) p(\boldsymbol{x}) d\boldsymbol{x} \tag{S.8.18}$$

となる．なお，式 (S.8.15) から式 (S.8.16) への変形には

$$\sum_{j=1}^{c} P(\omega_j) p(\boldsymbol{x}|\omega_j) = p(\boldsymbol{x}) \tag{S.8.19}$$

の関係と，ベイズの定理と式 (S.8.1) より得られる

$$P(\omega_i) p(\boldsymbol{x}|\omega_i) = P(\omega_i|\boldsymbol{x}) p(\boldsymbol{x}) = G_i(\boldsymbol{x}) p(\boldsymbol{x}) \tag{S.8.20}$$

を用いた．

式 (S.8.18) の第 2 項は $\mathbf{w}_i$ に依存しないので，$\epsilon_0^2(\mathbf{w}_i)$ を最小にする $\mathbf{w}_i$ は $\epsilon^2(\mathbf{w}_i)$ をも最小にする．したがって，求めるべき $\mathbf{w}_i$ は

$$\frac{\partial}{\partial \mathbf{w}_i} \epsilon_0^2(\mathbf{w}_i) = 0 \tag{S.8.21}$$

を満たす．一方，式 (S.8.10) を $\mathbf{w}_i$ で微分することにより

$$\mathrm{E}\left\{2\left(\mathbf{w}_i^t\mathbf{x}-b_i(\boldsymbol{x})\right)\mathbf{x}\right\}=\frac{\partial}{\partial\mathbf{w}_i}\epsilon_0^2(\mathbf{w}_i) \tag{S.8.22}$$

が得られる．ここで，上式を式 (8.44) と比較すると，以下のような対応関係があることがわかる．

$$\begin{aligned} h &\leftrightarrow 2\left(\mathbf{w}_i^t\mathbf{x}-b_i(\boldsymbol{x})\right)\mathbf{x} \\ f &\leftrightarrow \frac{\partial}{\partial\mathbf{w}_i}\epsilon_0^2(\mathbf{w}_i) \end{aligned}$$

すなわち，$\frac{\partial}{\partial\mathbf{w}_i}\epsilon_0^2(\mathbf{w}_i)$ は $2\left(\mathbf{w}_i^t\mathbf{x}-b_i(\boldsymbol{x})\right)\mathbf{x}$ の回帰関数である．かつ，求めるべき $\mathbf{w}_i$ は，式 (S.8.21) の根であるので，ロビンス・モンローアルゴリズムが適用できる．したがって，$\mathbf{w}_i$ は

$$\mathbf{w}_i(t+1)=\mathbf{w}_i(t)-2\rho(t)\left(\mathbf{w}_i^t(t)\mathbf{x}-b_i(\boldsymbol{x})\right)\mathbf{x} \tag{S.8.23}$$

の繰り返しによって求めることができる[*13]．

式 (S.8.23) は，式 (3.31) のウィドロー・ホフの学習規則に他ならない．

[*13] なお，h，f とも多変数の関数であるので，本文の 8.3 節 末尾に述べた注意が必要である．

第９章

【演習問題 9.1】

以下に式 (3.40) を再掲する．

$$\mathbf{w} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{b} \tag{S.9.1}$$

上式で $\mathbf{X}$ は式 (3.15) で定義されたパターン行列，$\mathbf{b}$ は式 (3.35)，(3.38) で定義された教師ベクトルであり，下式で表される．

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)^t \tag{S.9.2}$$

$$\mathbf{b} = (b_1, b_2, \ldots, b_n)^t \tag{S.9.3}$$

$$b_p = \begin{cases} 1 & (\boldsymbol{x}_p \in \omega_1) \\ -1 & (\boldsymbol{x}_p \in \omega_2) \end{cases} \qquad (p = 1, \ldots, n) \tag{S.9.4}$$

式 (S.9.1) を変形すると

$$\mathbf{X}^t\mathbf{X}\mathbf{w} = \mathbf{X}^t\mathbf{b} \tag{S.9.5}$$

が得られる．ここで

$$\mathbf{X}^t\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \begin{pmatrix} \mathbf{x}_1^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix} = \sum_{p=1}^{n} \mathbf{x}_p\mathbf{x}_p^t \tag{S.9.6}$$

$$\mathbf{X}^t\mathbf{b} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{p=1}^{n} b_p\mathbf{x}_p \tag{S.9.7}$$

であるので，これらを式 (S.9.5) に代入した後，両辺に $1/n$ を掛けると

$$\left(\frac{1}{n}\sum_{p=1}^{n} \mathbf{x}_p\mathbf{x}_p^t\right)\mathbf{w} = \frac{1}{n}\sum_{p=1}^{n} b_p\mathbf{x}_p \tag{S.9.8}$$

が得られる．ここで，式 (9.8) より

$$\frac{1}{n}\sum_{p=1}^{n} \mathbf{x}_p\mathbf{x}_p^t = \mathop{\mathrm{E}}_{\boldsymbol{x}}\{\mathbf{x}\mathbf{x}^t\} = \mathbf{R}_0 \tag{S.9.9}$$

であり，また式 (S.9.4) より

$$\begin{aligned}
&\frac{1}{n}\sum_{p=1}^{n} b_p \mathbf{x}_p \\
&= \frac{1}{n}\sum_{\boldsymbol{x}_p \in \omega_1} \mathbf{x}_p - \frac{1}{n}\sum_{\boldsymbol{x}_p \in \omega_2} \mathbf{x}_p \\
&= \frac{n_1}{n}\frac{1}{n_1}\sum_{\boldsymbol{x}_p \in \omega_1} \mathbf{x}_p - \frac{n_2}{n}\frac{1}{n_2}\sum_{\boldsymbol{x}_p \in \omega_2} \mathbf{x}_p \\
&= P(\omega_1)\frac{1}{n_1}\sum_{\boldsymbol{x}_p \in \omega_1} \begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix} - P(\omega_2)\frac{1}{n_2}\sum_{\boldsymbol{x}_p \in \omega_2} \begin{pmatrix} 1 \\ \boldsymbol{x}_p \end{pmatrix} \\
&= P(\omega_1)\begin{pmatrix} 1 \\ \mathbf{m}_1 \end{pmatrix} - P(\omega_2)\begin{pmatrix} 1 \\ \mathbf{m}_2 \end{pmatrix} \\
&= \begin{pmatrix} P(\omega_1) - P(\omega_2) \\ P(\omega_1)\,\mathbf{m}_1 - P(\omega_2)\,\mathbf{m}_2 \end{pmatrix}
\end{aligned} \tag{S.9.10}$$

が得られる．ただし，n_1，n_2 は，それぞれクラス ω_1，ω_2 のパターン数である．上式は，式 (9.10) において $b_1 = 1$，$b_2 = -1$ とした結果と一致するので

$$\frac{1}{n}\sum_{p=1}^{n} b_p \mathbf{x}_p = \mathbf{r} \tag{S.9.11}$$

としてよい．式 (S.9.9)，(S.9.11) を式 (S.9.8) に代入すると

$$\mathbf{R}_0 \mathbf{w} = \mathbf{r} \tag{S.9.12}$$

となり，式 (9.12) が導出できる．したがって，以後の処理は式 (9.13) 以降と全く同じであり，最適な $\boldsymbol{w}$，w_0 は式 (9.21)，式 (9.22) と一致する．

【演習問題 9.2*】

すでに示した**演習問題 4.3** の解法の手順で，式 (S.4.32) まではそのまま利用する．その後は式 (4.41) ではなく式 (4.40)，すなわち

$$\boldsymbol{w} = \frac{1}{2}\cdot\frac{\partial J}{\partial \tilde{m}_1}\left(\frac{\partial J}{\partial \tilde{\sigma}_1^2}\boldsymbol{\Sigma}_1 + \frac{\partial J}{\partial \tilde{\sigma}_2^2}\boldsymbol{\Sigma}_2\right)^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \tag{S.9.13}$$

を用いる．ここで式 (S.4.31)，(S.4.32)，(S.4.34) より

$$\frac{\partial J_0}{\partial \tilde{\sigma}_1^2} = P(\omega_1) \tag{S.9.14}$$

$$\frac{\partial J_0}{\partial \tilde{\sigma}_2^2} = P(\omega_2) \tag{S.9.15}$$

$$\frac{\partial J_0}{\partial \tilde{m}_1} = 2\,P(\omega_1)\,(\tilde{m}_1 - 1) \tag{S.9.16}$$

であるので，これらを式 (S.9.13) に代入し，さらに

$$\mathbf{m}_d \stackrel{\text{def}}{=} \mathbf{m}_1 - \mathbf{m}_2 \tag{S.9.17}$$

と置くと

$$\begin{aligned} \boldsymbol{w} &= P(\omega_1)\,(\tilde{m}_1 - 1)\,(P(\omega_1)\boldsymbol{\Sigma}_1 + P(\omega_2)\boldsymbol{\Sigma}_2)^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \\ &= P(\omega_1)(1 - \tilde{m}_1)\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d \end{aligned} \tag{S.9.18}$$

を得る．

一方，w_0 は，式 (S.4.42) より

$$w_0 = -\mathbf{m}^t\boldsymbol{w} + P(\omega_1) - P(\omega_2) \tag{S.9.19}$$

である．式 (4.30) と式 (S.9.19) より

$$\begin{aligned} \tilde{m}_1 &= \boldsymbol{w}^t\mathbf{m}_1 + w_0 \\ &= (\mathbf{m}_1 - \mathbf{m})^t\boldsymbol{w} + P(\omega_1) - P(\omega_2) \\ &= P(\omega_2)\,\mathbf{m}_d^t\,\boldsymbol{w} + 1 - 2P(\omega_2) \end{aligned} \tag{S.9.20}$$

が成り立つ．ただし

$$P(\omega_1) + P(\omega_2) = 1 \tag{S.9.21}$$

$$\mathbf{m} = P(\omega_1)\,\mathbf{m}_1 + P(\omega_2)\,\mathbf{m}_2 \tag{S.9.22}$$

を用いた．式 (S.9.20) を式 (S.9.18) に代入し

$$\boldsymbol{w} = P(\omega_1)\,P(\omega_2)\,(2 - \mathbf{m}_d^t\boldsymbol{w})\,\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d \tag{S.9.23}$$

を得る．ここで

$$a = P(\omega_1)\,(1 - \tilde{m}_1) \tag{S.9.24}$$

と置くと，式 (S.9.18) は

$$\boldsymbol{w} = a \cdot \boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d \tag{S.9.25}$$

と書ける．式 (S.9.25) を式 (S.9.23) に代入すると

$$a \cdot \boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d = P(\omega_1)\,P(\omega_2)\,(2 - a \cdot \mathbf{m}_d^t\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d)\,\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d \tag{S.9.26}$$

となり，両辺の共通のベクトル $\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d$ にかかる係数を比較することにより

$$a = P(\omega_1)\,P(\omega_2)\,(2 - a \cdot \mathbf{m}_d^t\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d) \tag{S.9.27}$$

が成り立つことがわかる．これより

$$a = \frac{2\,P(\omega_1)\,P(\omega_2)}{1 + P(\omega_1)\,P(\omega_2)\mathbf{m}_d^t\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d} \tag{S.9.28}$$

と求められ，式 (S.9.25) より

$$\boldsymbol{w} = \frac{2\,P(\omega_1)\,P(\omega_2)}{1 + P(\omega_1)\,P(\omega_2)\mathbf{m}_d^t\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d} \cdot \boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d \tag{S.9.29}$$

が得られる．

一方，式 (6.134) で示したように

$$\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d = \left(1 + P(\omega_1)P(\omega_2)\mathbf{m}_d^t\boldsymbol{\Sigma}_W^{-1}\mathbf{m}_d\right)\boldsymbol{\Sigma}_T^{-1}\mathbf{m}_d \tag{S.9.30}$$

の関係があるので，上式を式 (S.9.29) に代入すると

$$\begin{aligned}\boldsymbol{w} &= 2\,P(\omega_1)\,P(\omega_2)\boldsymbol{\Sigma}_T^{-1}\mathbf{m}_d \\ &= 2\,P(\omega_1)\,P(\omega_2)\boldsymbol{\Sigma}_T^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\end{aligned} \tag{S.9.31}$$

となり，式 (9.21) が得られる．また，上式を式 (S.9.19) に代入することにより

$$w_0 = -2\,P(\omega_1)\,P(\omega_2)\,\mathbf{m}^t\boldsymbol{\Sigma}_T^{-1}(\mathbf{m}_1 - \mathbf{m}_2) + P(\omega_1) - P(\omega_2) \tag{S.9.32}$$

となり，式 (9.22) が得られる．

【演習問題 9.3】

教師信号 b_{ip} $(i = 1, \ldots, c, \quad p = 1, \ldots, n)$ を要素とする $c \times n$ の行列 $\mathbf{B}$ を次式のごとく定義する．

$$\mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{c1} & b_{c2} & \ldots & b_{cn} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1^t \\ \mathbf{b}_2^t \\ \vdots \\ \mathbf{b}_c^t \end{pmatrix} = (\boldsymbol{\tau}_1,\ \boldsymbol{\tau}_2,\ \ldots,\ \boldsymbol{\tau}_n) \tag{S.9.33}$$

上式で，$\mathbf{b}_i$ は式 (9.43) で定義されており，$\boldsymbol{\tau}_p$ はパターン $\boldsymbol{x}_p$ に対する教師ベクトルで，式 (3.3) で示したように

$$\boldsymbol{\tau}_p = (b_{1p},\ b_{2p},\ \ldots, b_{cp})^t \tag{S.9.34}$$

で表される．式 (3.10)，(9.44) を用いると，式 (9.42) は

$$\begin{aligned} J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c) &= \frac{1}{2} \sum_{i=1}^{c} \|\mathbf{X}\mathbf{w}_i - \mathbf{b}_i\|^2 \\ &= \frac{1}{2} \sum_{p=1}^{n} \sum_{i=1}^{c} (\mathbf{w}_i^t \mathbf{x}_p - b_{ip})^2 \\ &= \frac{1}{2} \sum_{p=1}^{n} \|\Psi(\boldsymbol{x}_p) - \boldsymbol{\tau}_p\|^2 \end{aligned} \tag{S.9.35}$$

と書ける．パターン $\boldsymbol{x}_p$ がクラス ω_i に所属するなら，その教師ベクトル $\boldsymbol{\tau}_p$ は

$$\boldsymbol{\tau}_p = \mathbf{t}_i = (\overset{1}{0}, \ldots, 0, \overset{i}{1}, 0, \ldots, \overset{c}{0}) \tag{S.9.36}$$

である．式 (S.9.35) の $\boldsymbol{\tau}_p$ の代わりに $\mathbf{t}_i$ を使用すると，式 (S.9.35) は

$$J(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_c) = \frac{1}{2} \sum_{i=1}^{c} \sum_{\boldsymbol{x}_p \in \omega_i} \|\Psi(\boldsymbol{x}_p) - \mathbf{t}_i\|^2 \tag{S.9.37}$$

$$= \frac{1}{2} \sum_{p=1}^{n} \sum_{i=1}^{c} \|\Psi(\boldsymbol{x}_p) - \mathbf{t}_i\|^2 \cdot 1(\boldsymbol{x}_p \in \omega_i) \tag{S.9.38}$$

となり，式 (9.45) と一致する．なお，式 (S.9.37) の $\sum_{\boldsymbol{x}_p \in \omega_i}$ は，ω_i に所属するパターン $\boldsymbol{x}_p$ のみを対象とした加算であることを示す．