

VI-1-3. 線形判別分析.

日常生活で、私たちは、過去の経験から得た情報を使って様々なことを判断しています。たとえば、今日の天気（気温、気圧、湿度、風向きなど）から、明日の天気を予想します。私たちは無意識に過去の天気良かった日や雨だった日の一日前の天気を記憶して、それらを総合して、翌日雨になるか晴れになるかという臨界値を持っているのです。これは一種の判別分析です。つまり、天気になる日の前の日か天気ではない日の前の日かというのを判別しています。判別分析とは、何が起きたかを知っている過去のデータを利用して、何が起こるかを予想するための判別スコア（臨界値）を作ることです。こういう判断は実際には複雑ですが、その簡単なものが線形判別分析です。線形判別分析では判別スコアを変数の線形結合によってつくります。単純な線形判別分析では、二つの母集団の間に等分散性を仮定します。つまり、天気良かった日の前の日も天気が悪くなかった日でもデータの分散は同じだと考えているのです。等分散性が前提となるので、等分散でない場合には適用できないというのが、線形判別分析の弱点です。等分散を前提としない判別分析もあります。あることが起こる確率の分析であるロジスティック回帰分析も判別分析の一つだと考えることもできますがこれも等分散を前提にしていません。ロジスティック回帰では、超越関数を含む連立微分方程式を解くので、代数的に解を求めることが出来ません。試行錯誤的な数値微分法（たとえば勾配降下法）を使わなくてはならないので、コンピュータの計算速度に依存しています。その他にもありますが、EM アルゴリズムを使ったりして、やはりコンピュータの計算速度に依存しています。これらの判別分析を実感として理解するには、コンピュータのプログラミングをしなければなりません。それはそれで長い解説になるので、別の章で解説することにして、ここでは、考え方の基礎として、等分散性を前提にした線形判別分析について説明します。

VI-1-3-2. 判別スコア

いくつかの部分集団 $(1, 2, \dots, k, \dots, m)$ を含む母集団の過去のデータがあるとき、私たちは新しく得られたデータが、どの部分集団に属するのかを推定したくなります。そういう目的のために、その判断のために判別スコアと判別のための指標（臨界値）を作ります。最も簡単なスコアの作り方は線形結合です。

$$a_1x_1 + a_2x_2 + \dots + a_px_p = z$$

z : 判別スコア

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix},$$

$$Z = (a_1 \quad a_2 \quad \cdots \quad a_p) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = A^T X$$

点 $(x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$ と平面 $a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p = 0$ の距離は、以下の点と平面の距離の公式（注参照）で表せます。

$$\frac{|a_0' + a_1'x_{i1} + a_2'x_{i2} + \cdots + a_p'x_{ip}|}{\sqrt{\sum_{l=1}^p a_l'^2}} = d$$

これを以下のように変形して、

$$\left| \frac{a_0'}{\sqrt{\sum_{l=1}^p a_l'^2}} + \frac{a_1'}{\sqrt{\sum_{l=1}^p a_l'^2}}x_{i1} + \frac{a_2'}{\sqrt{\sum_{l=1}^p a_l'^2}}x_{i2} + \cdots + \frac{a_p'}{\sqrt{\sum_{l=1}^p a_l'^2}}x_{ip} \right| = d$$

として、

$$\frac{a_0'}{\sqrt{\sum_{l=1}^p a_l'^2}} = 0$$

$$\frac{a_1'}{\sqrt{\sum_{l=1}^p a_l'^2}} = a_1$$

$$\frac{a_2'}{\sqrt{\sum_{l=1}^p a_l'^2}} = a_2$$

⋮

$$\frac{a_p'}{\sqrt{\sum_{l=1}^p a_l'^2}} = a_p$$

と書くと、次の式が原点を通る超平面となり、

$$a_1x_1 + a_2x_2 + \cdots + a_px_p = 0$$

それぞれの点と超平面の距離は以下の式で表せます。

$$|a_1x_{i1} + a_2x_{i2} + \cdots + a_px_{ip}| = d$$

ここでデータ d_{ki} を次のように記述することにします。

k = 部分集団の番号 $(1, \dots, m)$

i : 部分集団内の標本番号, $(1, \dots, n_k)$

$$d_{ki} = \begin{pmatrix} x_{ki1} \\ \vdots \\ x_{kip} \end{pmatrix}$$

負の方向まで考えると、

$$a_1 x_{ki1} + a_2 x_{ki2} + \dots + a_p x_{kip} = Z_{ki}$$

となりスコア Z_{ki} は原点を含む平面とその平面外の空間上の点の距離ということになります。別の言い方をすると、判別スコア Z_{ki} はある点から平面への法線の長さです。図 68 のように、この法線への個々の点からの写像を考えます。この法線上の写像が、図 69 のように、それぞれの部分集団ごとに正規分布しているというのが線形判別分析のモデルです。ですから、点と平面の距離が計算できれば良いことになります。

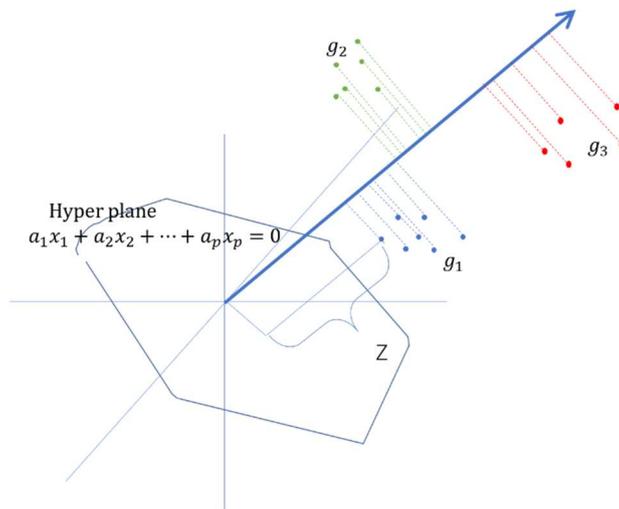


図 68 原点を含む平面の法線へのデータの点への写像

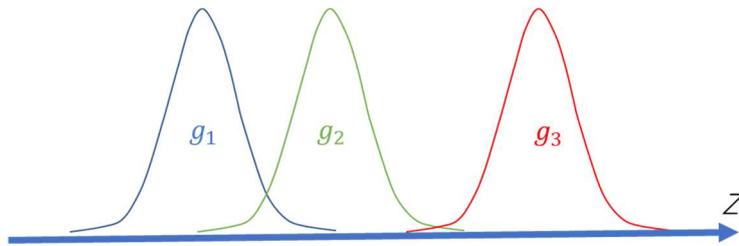
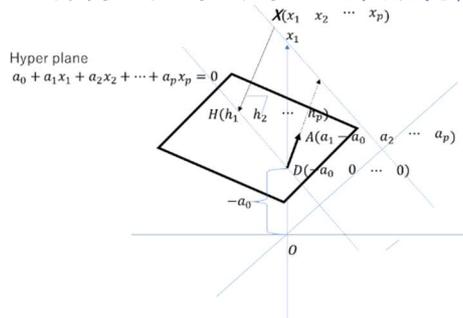


図 69 判別スコア Z の法線への写像の分布。

注 (点と平面の距離の公式)

点と平面の距離の公式を忘れていた人もいたかもしれないので、公式を作ります。



図の点 X と超平面 $a_1x_1 + a_2x_2 + \dots + a_px_p$ の距離を求める。一般に法線の式は、 $\overline{DA}: (a_1 \ a_2 \ \dots \ a_p)$ の実数倍、 $t\overline{DA}: (ta_1 \ ta_2 \ \dots \ ta_p)$ と表すことが出来る。点 X 超平面に下した垂線の脚を $H: (h_1 \ h_2 \ \dots \ h_p)$ とする。 \overline{AH} は超平面に直交するので、超平面の法線の一つである。ここで、向きを反対にして、 \overline{HA} について考える。

$$\overline{HA} = \overline{OA} - \overline{OH} = (x_1 - h_1 \ x_2 - h_2 \ \dots \ x_p - h_p)$$

だから、これを

$$m\overline{DA}: (ma_1 \ ma_2 \ \dots \ ma_p)$$

とすると

$$(x_1 - h_1 \ x_2 - h_2 \ \dots \ x_p - h_p) = (ma_1 \ ma_2 \ \dots \ ma_p)$$

となつて、

$$h_1 = x_1 - ma_1$$

$$h_2 = x_2 - ma_2$$

⋮

$$h_p = x_p - ma_p$$

となり、

$$H: (x_1 - ma_1 \ x_2 - ma_2 \ \dots \ x_p - ma_p)$$

と表すことが出来る。

H は超平面上の点だから、超平面の式にしたがう。

$$\begin{aligned} a_0 + a_1(x_1 - ma_1) + a_2(x_2 - ma_2) + \dots + a_p(x_p - ma_p) &= 0 \\ a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p - ma_1^2 - ma_2^2 - \dots - ma_p^2 &= 0 \\ a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p &= m(a_1^2 + a_2^2 + \dots + a_p^2) \end{aligned}$$

だから、

$$m = \frac{a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p}{a_1^2 + a_2^2 + \dots + a_p^2}$$

となり、 m を求めることが出来る。ところで、 \overline{HA} の長さは、

$$|\overline{HA}| = \sqrt{(ma_1)^2 + (ma_2)^2 + \dots + (ma_p)^2} = |m| \sqrt{a_1^2 + a_2^2 + \dots + a_p^2}$$

だから、 $|m|$ を代入して、

$$\begin{aligned} |\overline{HA}| &= \frac{|a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p|}{a_1^2 + a_2^2 + \dots + a_p^2} \sqrt{a_1^2 + a_2^2 + \dots + a_p^2} \\ &= \frac{a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p}{\sqrt{a_1^2 + a_2^2 + \dots + a_p^2}} \end{aligned}$$

となる。

VI-1-3-3. 判別スコアの分散の分析.

判別分析の公式を導く方法は、細かい違いまで考えるといろいろとあるのですが、大別すると、直接、判別スコアの分散を考えやり方と、超空間上のデータの分布を線形代数的にベクトルや行列で表して距離を最小化する方法があります。ここでは、まず、分散分析的に判別関数とその臨界値を導く方法を紹介し、次に線形代数的な方法を紹介します。次のようなデータセットがあったとします。

部分集団	標本番号	データ	変数の値
			$i \quad \dots \quad p$
1	1	\mathbf{d}_{11}	$= (d_{111} \quad \dots \quad d_{11p})^T$
	2	\mathbf{d}_{12}	$= (d_{121} \quad \dots \quad d_{12p})^T$
⋮	⋮	⋮	⋮

$$\begin{array}{ll}
1 & n_1 \\
\text{部分集団 1 の合計} & \mathbf{d}_{1n_1} = (d_{1n_11} \quad \dots \quad d_{1n_1p})^T \\
& \sum_{i=1}^{n_1} \mathbf{d}_{1i} = (\sum_{i=1}^{n_1} d_{1i1} \quad \dots \quad \sum_{i=1}^{n_1} d_{1ip})^T \\
\text{部分集団 1 の平均} & \bar{\mathbf{d}}_1 = \left(\frac{\sum_{i=1}^{n_1} d_{1i1}}{n_1} = \bar{d}_{11} \quad \dots \quad \frac{\sum_{i=1}^{n_1} d_{1ip}}{n_1} = \bar{d}_{1p} \right)^T
\end{array}$$

$$\begin{array}{ll}
2 & 1 \\
2 & 2 \\
\vdots & \vdots \\
2 & n_2 \\
\text{部分集団 2 の合計} & \mathbf{d}_{2n_2} = (d_{2n_21} \quad \dots \quad d_{2n_2p})^T \\
& \sum_{i=1}^{n_2} \mathbf{d}_{2i} = (\sum_{i=1}^{n_2} d_{2i1} \quad \dots \quad \sum_{i=1}^{n_2} d_{2ip})^T \\
\text{部分集団 2 の平均} & \bar{\mathbf{d}}_2 = \left(\frac{\sum_{i=1}^{n_2} d_{2i1}}{n_2} = \bar{d}_{21} \quad \dots \quad \frac{\sum_{i=1}^{n_2} d_{2ip}}{n_2} = \bar{d}_{2p} \right)^T
\end{array}$$

⋮

$$\begin{array}{ll}
m & 1 \\
m & 2 \\
\vdots & \vdots \\
m & n_m \\
\text{部分集団 m の合計} & \mathbf{d}_{mn_m} = (d_{mn_m1} \quad \dots \quad d_{mn_m p})^T \\
& \sum_{i=1}^{n_m} \mathbf{d}_{mi} = (\sum_{i=1}^{n_m} d_{mi1} \quad \dots \quad \sum_{i=1}^{n_m} d_{mip})^T \\
\text{部分集団 m の平均} & \bar{\mathbf{d}}_m = \left(\frac{\sum_{i=1}^{n_m} d_{mi1}}{n_m} = \bar{d}_{m1} \quad \dots \quad \frac{\sum_{i=1}^{n_m} d_{mip}}{n_m} = \bar{d}_{mp} \right)^T
\end{array}$$

$$\text{全体の合計} \quad N = \sum_{k=1}^m n_k \quad \sum_{k=1}^m \sum_{i=1}^{n_k} \mathbf{d}_{ki} = (\sum_{k=1}^m \sum_{i=1}^{n_k} d_{ki1} \quad \dots \quad \sum_{k=1}^m \sum_{i=1}^{n_k} d_{kip})^T$$

$$\text{全体の平均} \quad \frac{\sum_{k=1}^m \sum_{i=1}^{n_k} \mathbf{d}_{ki}}{N} = \bar{\mathbf{d}} = \left(\frac{\sum_{k=1}^m \sum_{i=1}^{n_k} d_{ki1}}{N} = \bar{d}_1 \quad \dots \quad \frac{\sum_{k=1}^m \sum_{i=1}^{n_k} d_{kip}}{N} = \bar{d}_p \right)^T$$

全平均: $\bar{\mathbf{d}}$

部分集団の平均: $\bar{\mathbf{d}}_i$, ($i = 1, \dots, m$)

計算を簡便化するために、すべてのデータを平均値からの距離のデータに変換します。

$$x_{kij} = d_{kij} - \bar{d}_j$$

変形したデータセットは以下のようになります。

部分集団番号	標本番号	データ	変数
			$i \quad \dots \quad p$
1	1	\mathbf{x}_{11}	$= (x_{111} \quad \dots \quad x_{11p})^T$
	2	\mathbf{x}_{12}	$= (x_{121} \quad \dots \quad x_{12p})^T$
⋮	⋮	⋮	⋮

1	n_1	$\mathbf{x}_{1n_1} = (x_{1n_11} \quad \dots \quad x_{1n_1p})^T$
部分集団 1 の合計		$\sum_{i=1}^{n_1} \mathbf{x}_{1i} = (\sum_{i=1}^{n_1} x_{1i1} \quad \dots \quad \sum_{i=1}^{n_1} x_{1ip})^T$
部分集団 1 の平均		$\bar{\mathbf{x}}_1 = \left(\frac{\sum_{i=1}^{n_1} x_{1i1}}{n_1} = \bar{x}_{11} \quad \dots \quad \frac{\sum_{i=1}^{n_1} x_{1ip}}{n_1} = \bar{x}_{1p} \right)^T$
<hr/>		
2	1	$\mathbf{x}_{21} = (x_{211} \quad \dots \quad x_{21p})^T$
2	2	$\mathbf{x}_{22} = (x_{221} \quad \dots \quad x_{22p})^T$
⋮	⋮	⋮
2	n_2	$\mathbf{x}_{2n_2} = (x_{2n_21} \quad \dots \quad x_{2n_2p})^T$
部分集団 2 の合計		$\sum_{i=1}^{n_2} \mathbf{x}_{2i} = (\sum_{i=1}^{n_2} x_{2i1} \quad \dots \quad \sum_{i=1}^{n_2} x_{2ip})^T$
部分集団 2 の平均		$\bar{\mathbf{x}}_2 = \left(\frac{\sum_{i=1}^{n_2} x_{2i1}}{n_2} = \bar{x}_{21} \quad \dots \quad \frac{\sum_{i=1}^{n_2} x_{2ip}}{n_2} = \bar{x}_{2p} \right)^T$
<hr/>		
⋮		
<hr/>		
m	1	$\mathbf{x}_{m1} = (x_{m11} \quad \dots \quad x_{m1p})^T$
m	2	$\mathbf{x}_{m2} = (x_{m21} \quad \dots \quad x_{m2p})^T$
⋮	⋮	⋮
m	n_m	$\mathbf{x}_{mn_m} = (x_{mn_m1} \quad \dots \quad x_{mn_m p})^T$
部分集団 m の合計		$\sum_{i=1}^{n_m} \mathbf{x}_{mi} = (\sum_{i=1}^{n_m} x_{mi1} \quad \dots \quad \sum_{i=1}^{n_m} x_{mip})^T$
部分集団 m の平均		$\bar{\mathbf{x}}_m = \left(\frac{\sum_{i=1}^{n_m} x_{mi1}}{n_m} = \bar{x}_{m1} \quad \dots \quad \frac{\sum_{i=1}^{n_m} x_{mip}}{n_m} = \bar{x}_{mp} \right)^T$
全体の合計	N	$\sum_{k=1}^m \sum_{i=1}^{n_k} \mathbf{x}_{ki} = (0 \quad \dots \quad 0)^T$
全体の平均		$\bar{\mathbf{x}} = (0 \quad \dots \quad 0)^T$

ここで、それぞれの部分集団のデータ分布の分散は同じだと仮定します（等分散性の仮定）。私たちが問われているのは、これらのデータの分布が部分集団間の平均値の違いを最も強調する射影先の線を見つけることです。これは、平均値の分散と部分集団内の分散の比を最大化するということから、F 値の最大化です。

まず、全体の平均と各部分集団の平均値について考えます。

$$Z_{ki} = \mathbf{A}^T \mathbf{x}_{ki} = (a_1 x_{ki1} + a_2 x_{ki2} + \dots + a_p x_{kip})$$

ですから、全体の平均は

$$\bar{Z} = \mathbf{A}^T \bar{\mathbf{x}} = 0$$

部分集団の平均は

$$\begin{aligned}\bar{Z}_k &= A^T \bar{x}_k = (a_1 \bar{x}_{k1} + a_2 \bar{x}_{k2} + \dots + a_p \bar{x}_{kip}) = \frac{1}{n_k} \sum_{i=1}^{n_k} (a_1 x_{ki1} + a_2 x_{ki2} + \dots + a_p x_{kip}) \\ &= (a_1 \quad \dots \quad a_p) \begin{pmatrix} \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki} \\ \vdots \\ \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kip} \end{pmatrix}\end{aligned}$$

データを部分集団の平均値と部分集団の平均値からの差に分けます。部分集団からの平均値からの差は、部分集団で説明できないランダムな変動（説明できない残差）と考えます。

$$Z_{ki} = \bar{Z}_k + e_{ki}$$

そうすると、

$$e_{ki} = Z_{ki} - \bar{Z}_k = A^T (x_{ki} - \bar{x}_k) = A^T \begin{pmatrix} x_{ki} - \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki1} \\ \vdots \\ x_{kip} - \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kip} \end{pmatrix}$$

これが、各部分集団の平均値からの差です。全体平均からの差の平方和は次の通りです。

$$\begin{aligned}SS_{total} &= \sum_{k=1}^m \sum_{i=1}^{n_k} Z_{ki}^2 = \sum_{k=1}^m \sum_{i=1}^{n_k} (e_{ki} + \bar{Z}_k)^2 \\ &= \sum_{k=1}^m \sum_{i=1}^{n_k} e_{ki}^2 + 2 \sum_{k=1}^m \bar{Z}_k \sum_{i=1}^{n_k} e_{ki} + \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2 \\ &= \sum_{k=1}^m \sum_{i=1}^{n_k} e_{ki}^2 + \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2 \\ &\quad \left(\because \sum_{i=1}^{n_k} e_{ki} = 0 \right)\end{aligned}$$

$$SS_{total} = SS_{residual} + SS_{subpopulation}$$

平均値からの距離の二乗の総和 $\sum_{i=1}^{n_k} e_{ki}^2$ は部分集団の平方和ですから、次のように平方和として書けます。

$$\begin{aligned}\sum_{i=1}^{n_k} e_{ki}^2 &= SS_k \\ SS_{residual} &= \sum_{k=1}^m SS_k = \sum_{k=1}^m (n_k - 1) \sigma_{residual}^2 = \sum_{k=1}^m \sum_{i=1}^{n_k} e_{ki}^2\end{aligned}$$

$$\sum_{k=1}^m (n_k - 1) : \text{残差の自由度}$$

$$SS_{total} = \sum_{k=1}^m SS_k + \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2$$

$$SS_{subpopulatoin} = \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2$$

$$SS_{total} = SS_{residual} + SS_{subpopulatoin}$$

$$SS_{total} = \sum_{k=1}^m SS_k + \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2$$

$$SS_{subpopulatoin} = \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2$$

$$SS_{total} = SS_{residual} + SS_{subpopulatoin}$$

$$\sigma^2_{subpopulation} = \frac{1}{(m-1)} \sum_{k=1}^m \sum_{i=1}^{n_k} \bar{Z}_k^2$$

$$\sigma^2_{residual} = \frac{\sum_{k=1}^m SS_k}{\sum_{k=1}^m (n_k - 1)}$$

$$\sigma^2_{subpopulation} = \frac{SS_{subpopulatoin}}{m-1}$$

これらの分散比が F 値です。

$$F = \frac{\sigma^2_{subpopulation}}{\sigma^2_{residual}} = \frac{SS_{subpopulatoin} \sum_{k=1}^m (n_k - 1)}{m-1 \sum_{k=1}^m SS_k} = \left(\frac{\sum_{k=1}^m (n_k - 1)}{m-1} \right) \left(\frac{SS_{subpopulatoin}}{\sum_{k=1}^m SS_k} \right)$$

この場合 $\left(\frac{\sum_{k=1}^m (n_k - 1)}{m-1} \right)$ は一定で、判別スコアに関係がありません。

$\left(\frac{SS_{subpopulatoin}}{\sum_{k=1}^m SS_k} \right)$ の最大化だけを考えればよいということです。 $\left(\frac{SS_{subpopulatoin}}{SS_{residual}} \right)$ を \mathbf{A} の関数と

考えて、これを最大化するために \mathbf{A} を最適化するので、次のように書きます。

$$f(\mathbf{A}) = \left(\frac{SS_{subpopulatoin}}{SS_{residual}} \right)$$

まず、部分集団の平方和から考えます。

$$SS_{subpopulatoin} = \frac{\sum_{k=1}^m n_k (\bar{Z}_k)^2}{N}$$

$$\bar{Z}_k = \mathbf{A}^T \bar{\mathbf{x}}_k$$

$$\bar{\mathbf{x}}_k = \begin{pmatrix} \bar{x}_{k1} \\ \bar{x}_{k2} \\ \vdots \\ \bar{x}_{kp} \end{pmatrix} = \begin{pmatrix} \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki1} \\ \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki} \\ \vdots \\ \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kip} \end{pmatrix}$$

煩雑なので、平均値という意味で μ を使ってベクトルの中を単純な記号にします。

$$\mu_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kij}$$

$$\boldsymbol{\mu}_k = \begin{pmatrix} \mu_{k1} \\ \mu_{k2} \\ \vdots \\ \mu_{kp} \end{pmatrix}$$

$$\bar{\mathbf{Z}}_k = \mathbf{A}^T \bar{\mathbf{x}}_k = \mathbf{A}^T \boldsymbol{\mu}_k$$

$$(\bar{\mathbf{Z}}_k)^2 = (\mathbf{A}^T \bar{\mathbf{X}}_k)^2 = \mathbf{A}^T \boldsymbol{\mu}_k (\mathbf{A}^T \boldsymbol{\mu}_k)^T = \mathbf{A}^T \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \mathbf{A}$$

$$\mathbf{A}^T \boldsymbol{\mu}_k = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \end{pmatrix} \begin{pmatrix} \mu_{k1} \\ \mu_{k2} \\ \vdots \\ \mu_{kp} \end{pmatrix}$$

$$\boldsymbol{\mu}_k^T \mathbf{A} = \begin{pmatrix} \mu_{k1} & \mu_{k2} & \cdots & \mu_{kp} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

$$\boldsymbol{\mu}_k \boldsymbol{\mu}_k^T = \begin{pmatrix} \mu_{k1}^2 & \mu_{k1}\mu_{k2} & \cdots & \mu_{k1}\mu_{kp} \\ \mu_{k2}\mu_{k1} & \mu_{k2}^2 & \cdots & \mu_{k2}\mu_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{kp}\mu_{k1} & \mu_{kp}\mu_{k2} & \cdots & \mu_{kp}^2 \end{pmatrix}$$

$$\mathbf{A}^T \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \mathbf{A} = \mathbf{A}^T \begin{pmatrix} \mu_{k1}^2 & \mu_{k1}\mu_{k2} & \cdots & \mu_{k1}\mu_{kp} \\ \mu_{k2}\mu_{k1} & \mu_{k2}^2 & \cdots & \mu_{k2}\mu_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{kp}\mu_{k1} & \mu_{kp}\mu_{k2} & \cdots & \mu_{kp}^2 \end{pmatrix} \mathbf{A}$$

$$\mathbf{A}^T \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \mathbf{A} + \mathbf{A}^T \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T \mathbf{A} = \mathbf{A}^T (\boldsymbol{\mu}_k \boldsymbol{\mu}_k^T + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \mathbf{A}$$

$$\sum_{k=1}^m (\bar{\mathbf{Z}}_k)^2 = \mathbf{A}^T \left(\sum_{k=1}^m n_k (\boldsymbol{\mu}_k \boldsymbol{\mu}_k^T + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \right) \mathbf{A}$$

$$\begin{aligned}
&= \mathbf{A}^T \begin{pmatrix} \sum_{k=1}^m \mu_{k1}^2 & \sum_{k=1}^m \mu_{k1}\mu_{k2} & \cdots & \sum_{k=1}^m \mu_{k1}\mu_{kp} \\ \sum_{k=1}^m \mu_{k2}\mu_{k1} & \sum_{k=1}^m \mu_{k2}^2 & \cdots & \sum_{k=1}^m \mu_{k2}\mu_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^m \mu_{kp}\mu_{k1} & \sum_{k=1}^m \mu_{kp}\mu_{k2} & \cdots & \sum_{k=1}^m \mu_{kp}^2 \end{pmatrix} \mathbf{A} \\
&= \mathbf{A}^T \begin{pmatrix} \mu_{11} & \mu_{21} & \cdots & \mu_{m1} \\ \mu_{12} & \mu_{22} & \cdots & \mu_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{1p} & \mu_{2p} & \cdots & \mu_{mp} \end{pmatrix} \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & \mu_{mp} \end{pmatrix} \mathbf{A}
\end{aligned}$$

このように変形すると $(\sum_{k=1}^m (\boldsymbol{\mu}_k \boldsymbol{\mu}_k^T + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T))$ が、分散共分散行列だとわかります。これを、行列 \mathbf{M} とします。

$$\begin{aligned}
\mathbf{M} &= \sum_{k=1}^m \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T = \begin{pmatrix} \sum_{k=1}^m \mu_{k1}^2 & \sum_{k=1}^m \mu_{k1}\mu_{k2} & \cdots & \sum_{k=1}^m \mu_{k1}\mu_{kp} \\ \sum_{k=1}^m \mu_{k2}\mu_{k1} & \sum_{k=1}^m \mu_{k2}^2 & \cdots & \sum_{k=1}^m \mu_{k2}\mu_{kp} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^m \mu_{kp}\mu_{k1} & \sum_{k=1}^m \mu_{kp}\mu_{k2} & \cdots & \sum_{k=1}^m \mu_{kp}^2 \end{pmatrix} \\
SS_{\text{subpopulation}} &= \sum_{k=1}^m (\bar{Z}_k)^2 = \mathbf{A}^T \mathbf{M} \mathbf{A}
\end{aligned}$$

次に残差の平方和を考えます。

$$\begin{aligned}
SS_k &= \sum_{i=1}^{n_k} e_{ki}^2 = \sum_{i=1}^{n_k} (\mathbf{A}^T (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)) (\mathbf{A}^T (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k))^T = \sum_{i=1}^{n_k} \mathbf{A}^T (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k) (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T \mathbf{A} \\
&= \mathbf{A}^T \left(\sum_{i=1}^{n_k} (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k) (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T \right) \mathbf{A}
\end{aligned}$$

煩雑なので、残差の行列 $\mathbf{x}_{ki} - \bar{\mathbf{x}}_k$ を次のように書くことにします。

$$\mathbf{x}_{ki} - \bar{\mathbf{x}}_k = \begin{pmatrix} x_{ki} - \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki} \\ \vdots \\ x_{kip} - \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kip} \end{pmatrix} = \begin{pmatrix} \phi_{ki} \\ \vdots \\ \phi_{kip} \end{pmatrix}$$

$$\begin{aligned}
(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T &= \begin{pmatrix} \phi_{ki1} \\ \vdots \\ \phi_{kip} \end{pmatrix} (\phi_{ki1} \quad \cdots \quad \phi_{kip}) = \begin{pmatrix} \phi_{ki1}^2 & \phi_{ki1}\phi_{ki2} & \cdots & \phi_{ki1}\phi_{kip} \\ \phi_{ki2}\phi_{ki1} & \phi_{ki2}^2 & \cdots & \phi_{ki2}\phi_{kip} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{kip}\phi_{ki1} & \phi_{kip}\phi_{ki2} & \cdots & \phi_{kip}^2 \end{pmatrix} \\
\sum_{k=1}^m \sum_{i=1}^{n_k} (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T &= \begin{pmatrix} \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki1}^2 & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki1}\phi_{ki2} & \cdots & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki1}\phi_{kip} \\ \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki2}\phi_{ki1} & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki2}^2 & \cdots & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{ki2}\phi_{kip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{kip}\phi_{ki1} & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{kip}\phi_{ki2} & \cdots & \sum_{k=1}^m \sum_{i=1}^{n_k} \phi_{kip}^2 \end{pmatrix}
\end{aligned}$$

$\sum_{k=1}^m \sum_{i=1}^{n_k} (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T$ も残差全体の分散共分散行列です。この行列を \mathbf{V} と表します。

$$\mathbf{V} = \sum_{k=1}^m \sum_{i=1}^{n_k} (\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)(\mathbf{x}_{ki} - \bar{\mathbf{x}}_k)^T$$

平均の分散 \mathbf{M} と同様に残差の平方和は次のようになります。

$$SS_{residual} = \sum_{k=1}^m SS_k = \mathbf{A}^T \mathbf{V} \mathbf{A}$$

したがって

$$f(\mathbf{A}) = \frac{\mathbf{A}^T \mathbf{M} \mathbf{A}}{\mathbf{A}^T \mathbf{F} \mathbf{A}}$$

となり、次の偏微分方程式を解いて、 \mathbf{A} を求めます。

$$\frac{df(\mathbf{A})}{d\mathbf{A}} = 0$$

式 75

変動要因別に分散を分けて、偏微分法定式を解いて F 比を最大化する係数ベクトル \mathbf{A} を求めることが出来ました。ある点の座標を、次の式の \mathbf{x} に代入すると、その点を含む座標の原点からその超平面への距離になります。逆に言えば、原点を含む超平面（基準超平面）から点 \mathbf{x} への距離です。これが判別得点です。

$$d = \mathbf{A}^T \mathbf{x} = \text{DS (discriminant score)}$$

次に、部分集団を分ける判別得点の臨界値（境界値）を考えますが、実は、判別得点の境界値を決める一般的な法則はありません。良い境界値とは事柄によって、分析者の判断で決めるものです。たとえば、化学物質の毒性などでは、0 リスクに近い安全側に境界値を作りますが、馬券を買うときには、ある程度お金を失うリスクを引き受けて、境界値を判断します。中立的な境界値の設定の一つの例は、二つのグループの中心の midpoint を通る基準超平面と並行した超平面を選択することです。線形判別分析では、すべてのグループが同じ形・同じ大

きさのデータ分布を持っていることを仮定しています。ですから、中心間の中点を通る超平面が、どちらに属すると判断しても、同じ危険率になる超平面なのです。

$$A \text{ と } B \text{ の選択の境界値 } (p(A) = p(B)) = \frac{d_A + d_B}{2} = \mathbf{A}^T \left(\frac{\mathbf{x}_A + \mathbf{x}_B}{2} \right)$$

これで、線形判別分析の考え方はわかりましたが、 $f(\mathbf{A})$ が複雑になると微分は面倒かもしれません。もっとも、分母も分子も2次式だから、丁寧にやればできないことはありません。大した手間ではないと考える人もいるでしょう。それよりも、ただに2次式を作るのにわざわざ分散共分散行列を作るのは回り道です。計算手順の説明だけならば、そんな必要はないのです。わざわざ、分散共分散行列を作ったのは、次の空間幾何学的な説明のためです。その説明をする前に、手順をしっかりと覚えるために具体的な例を挙げて計算してみます。

練習

次のように8つのデータがあったとします。そのうち4つは部分集団1に属し、その他は部分集団2に属するとします。データは2変数データです。この二つのグループを判別する判別スコアを作ります。

データセット

部分集団	標本番号	データ	
		d_1	d_2
1	1	5	8
1	2	7	4
1	3	8	5
1	4	8	7
2	1	5	5
2	2	7	2
2	3	4	3
2	4	4	6
全平均		6	5

$$\bar{d}_1 = \frac{5 + 7 + 8 + 8 + 5 + 7 + 4 + 4}{8} = \frac{48}{8} = 6$$

$$\bar{d}_2 = \frac{8 + 4 + 5 + 7 + 5 + 2 + 3 + 6}{8} = \frac{40}{8} = 5$$

全平均のベクトル

$$\bar{\mathbf{d}} = \begin{pmatrix} 6 \\ 5 \end{pmatrix}$$

部分集団1の平均

$$\bar{d}_{11} = \frac{5 + 7 + 8 + 8}{4} = \frac{28}{4} = 7$$

$$\bar{d}_{12} = \frac{8 + 4 + 5 + 7}{4} = \frac{24}{4} = 6$$

部分集団 1 の平均ベクトル

$$\bar{\mathbf{d}}_1 = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$$

部分集団 2 の平均

$$\bar{d}_{21} = \frac{5 + 7 + 4 + 4}{4} = \frac{20}{4} = 5$$

$$\bar{d}_{22} = \frac{5 + 2 + 3 + 6}{4} = \frac{16}{4} = 4$$

$$\bar{\mathbf{d}}_2 = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

各データの全平均からの偏差

$$\mathbf{x}_{ki} = \mathbf{d}_{ki} - \bar{\mathbf{d}} = \begin{pmatrix} d_{ki1} \\ d_{ki2} \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} d_{ki1} - 6 \\ d_{ki2} - 5 \end{pmatrix}$$

$$\mathbf{x}_{ki} = (\mathbf{d}_{ki} - \bar{\mathbf{d}}_k) + (\bar{\mathbf{d}}_k - \bar{\mathbf{d}})$$

$$\bar{\mathbf{d}}_k - \bar{\mathbf{d}} = \begin{pmatrix} \bar{d}_{k1} \\ \bar{d}_{k2} \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix}$$

$$\mathbf{d}_{ki} - \bar{\mathbf{d}}_k = \begin{pmatrix} d_{ki1} \\ d_{ki2} \end{pmatrix} - \begin{pmatrix} \bar{d}_{k1} \\ \bar{d}_{k2} \end{pmatrix}$$

$$\bar{\mathbf{d}}_1 - \bar{\mathbf{d}} = \begin{pmatrix} 7 \\ 6 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \bar{\mathbf{d}}_2 - \bar{\mathbf{d}} = \begin{pmatrix} 5 \\ 4 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{d}_{11} - \bar{\mathbf{d}}_1 = \begin{pmatrix} 5 \\ 8 \end{pmatrix} - \begin{pmatrix} 7 \\ 6 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \mathbf{d}_{12} - \bar{\mathbf{d}}_1 = \begin{pmatrix} 7 \\ 4 \end{pmatrix} - \begin{pmatrix} 7 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \mathbf{d}_{13} - \bar{\mathbf{d}}_1 = \begin{pmatrix} 8 \\ 5 \end{pmatrix} - \begin{pmatrix} 7 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$\mathbf{d}_{14} - \bar{\mathbf{d}}_1 = \begin{pmatrix} 8 \\ 7 \end{pmatrix} - \begin{pmatrix} 7 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\mathbf{d}_{21} - \bar{\mathbf{d}}_2 = \begin{pmatrix} 5 \\ 5 \end{pmatrix} - \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{d}_{22} - \bar{\mathbf{d}}_2 = \begin{pmatrix} 7 \\ 2 \end{pmatrix} - \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \mathbf{d}_{23} - \bar{\mathbf{d}}_2 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{d}_{24} - \bar{\mathbf{d}}_2 = \begin{pmatrix} 4 \\ 6 \end{pmatrix} - \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

部分集団間の分散共分散行列

$$\begin{aligned} \mathbf{M} &= (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}} \quad \bar{\mathbf{d}}_2 - \bar{\mathbf{d}})(\bar{\mathbf{d}}_1 - \bar{\mathbf{d}} \quad \bar{\mathbf{d}}_2 - \bar{\mathbf{d}})^T \\ &= \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \end{aligned}$$

残差の分散共分散行列

$$\begin{aligned} \mathbf{V} &= \begin{pmatrix} -2 & 0 & 1 & 1 & 0 & 2 & -1 & -1 \\ 2 & -2 & -1 & 1 & 1 & -2 & -1 & 2 \end{pmatrix} \begin{pmatrix} -2 & 2 \\ 0 & -2 \\ 1 & -1 \\ 1 & 1 \\ 0 & 1 \\ 2 & -2 \\ -1 & -1 \\ -1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 4+0+1+1+0+4+1+1 & -4+0-1+1+0-4+1-2 \\ -4+0-1+1+0-4+1-2 & 4+4+1+1+1+4+1+4 \end{pmatrix} = \begin{pmatrix} 12 & -9 \\ -9 & 20 \end{pmatrix} \\ \mathbf{V} &= SS_{\text{residual}} = \begin{pmatrix} 12 & -9 \\ -9 & 20 \end{pmatrix} \end{aligned}$$

$\frac{(\mathbf{A}^T \mathbf{M} \mathbf{A})}{(\mathbf{A}^T \mathbf{V} \mathbf{A})}$ の最大化を考えます。

$$\begin{aligned} g(\mathbf{A}) &= \mathbf{A}^T \mathbf{M} \mathbf{M}^T \mathbf{A} = 2\mathbf{A}^t \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \mathbf{A} \\ &= 2(a_1 \quad a_2) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 2(a_1 + a_2 \quad a_1 + a_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 2(a_1^2 + 2a_1a_2 + a_2^2) \end{aligned}$$

$$\begin{aligned} h(\mathbf{A}) &= \mathbf{A}^T \begin{pmatrix} 12 & -9 \\ -9 & 20 \end{pmatrix} \mathbf{A} \\ &= (a_1 \quad a_2) \begin{pmatrix} 12 & -9 \\ -9 & 20 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\ &= (12a_1 - 9a_2 \quad -9a_1 + 20a_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\ &= 12a_1^2 - 9a_1a_2 - 9a_1a_2 + 20a_2^2 \\ &= 12a_1^2 - 18a_1a_2 + 20a_2^2 \\ &= 2(6a_1^2 - 9a_1a_2 + 10a_2^2) \\ f(\mathbf{A}) &= \frac{2(a_1^2 + 2a_1a_2 + a_2^2)}{2(6a_1^2 - 9a_1a_2 + 10a_2^2)} = \frac{a_1^2 + 2a_1a_2 + a_2^2}{6a_1^2 - 9a_1a_2 + 10a_2^2} \\ \frac{\partial f(\mathbf{A})}{\partial a_1} &= \frac{\partial \left(\frac{g(a_1)}{h(a_1)} \right)}{\partial a_1} = \frac{\frac{\partial g(a_1)}{\partial a_1} h(a_1) - g(a_1) \frac{\partial h(a_1)}{\partial a_1}}{h(a_1)^2} \\ g(a_1) &= a_1^2 + 2a_1a_2 + a_2^2 \end{aligned}$$

$$h(a_1) = 6a_1^2 - 9a_1a_2 + 10a_2^2$$

$$\frac{\partial g(a_1)}{\partial a_1} = 2a_1 + 2a_2$$

$$\frac{\partial h(a_1)}{\partial a_1} = 12a_1 - 9a_2$$

$$\begin{aligned} \frac{\partial f(\mathbf{A})}{\partial a_1} &= \frac{(2a_1 + 2a_2)(6a_1^2 - 9a_1a_2 + 10a_2^2) - (a_1^2 + 2a_1a_2 + a_2^2)(12a_1 - 9a_2)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{(12a_1^3 - 18a_1^2a_2 + 20a_1a_2^2 + 12a_1^2a_2 - 18a_1a_2^2 + 20a_2^3) - (12a_1^3 + 24a_1^2a_2 + 12a_1a_2^2 - 9a_1^2a_2 - 18a_1a_2^2 - 9a_2^3)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{(12a_1^3 - 6a_1^2a_2 + 2a_1a_2^2 + 20a_2^3) - (12a_1^3 + 15a_1^2a_2 - 6a_1a_2^2 - 9a_2^3)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{-21a_1^2a_2 + 8a_1a_2^2 + 29a_2^3}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \end{aligned}$$

$$\frac{\partial f(\mathbf{A})}{\partial a_2} = \frac{\partial \left(\frac{g(a_2)}{h(a_2)} \right)}{\partial a_1} = \frac{\frac{\partial g(a_2)}{\partial a_2} h(a_2) - g(a_2) \frac{\partial h(a_2)}{\partial a_2}}{h(a_1)^2}$$

$$\frac{\partial g(a_2)}{\partial a_2} = 2a_1 + 2a_2$$

$$\frac{\partial h(a_2)}{\partial a_2} = -9a_1 + 20a_2$$

$$\begin{aligned} \frac{\partial f(\mathbf{A})}{\partial a_2} &= \frac{(2a_1 + 2a_2)(6a_1^2 - 9a_1a_2 + 10a_2^2) - (a_1^2 + 2a_1a_2 + a_2^2)(-9a_1 + 20a_2)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{(12a_1^3 - 18a_1^2a_2 + 20a_1a_2^2 + 12a_1^2a_2 - 18a_1a_2^2 + 20a_2^3) - (-9a_1^3 - 18a_1^2a_2 - 9a_1a_2^2 + 20a_1^2a_2 + 40a_1a_2^2 + 20a_2^2)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{(12a_1^3 - 6a_1^2a_2 + 2a_1a_2^2 + 20a_2^3) - (-9a_1^3 + 2a_1^2a_2 + 31a_1a_2^2 + 20a_2^2)}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \\ &= \frac{21a_1^3 - 8a_1^2a_2 - 29a_1a_2^2}{(6a_1^2 - 9a_1a_2 + 10a_2^2)^2} \end{aligned}$$

極値の条件は以下の通りです。

$$\frac{\partial f(\mathbf{A})}{\partial a_1} = \frac{\partial f(\mathbf{A})}{\partial a_2} = 0$$

$$(6a_1^2 - 9a_1a_2 + 10a_2^2)^2 > 0$$

次の連立方程式を解きます。

$$-21a_1^2a_2 + 8a_1a_2^2 + 29a_2^3 = 0$$

$$21a_1^3 - 8a_1^2a_2 - 29a_1a_2^2 = 0$$

$$a_1 \neq 0$$

両辺を a_1^3 で割ります。

$$-21 \frac{a_2}{a_1} + 8 \left(\frac{a_2}{a_1} \right)^2 + 29 \left(\frac{a_2}{a_1} \right)^3 = 0$$

$$21 - 8 \frac{a_2}{a_1} - 29 \left(\frac{a_2}{a_1} \right)^2 = 0$$

2つの式は同じ式です。

$$\frac{a_2}{a_1} = t$$

$$-21t + 8t^2 + 29t^3 = 0$$

$$21 - 8t - 29t^2 = 0$$

$$29t^2 + 8t - 21 = 0$$

$$(29t - 21)(t + 1) = 0$$

$$t = \frac{21}{29}, \quad t = -1$$

$t = \frac{21}{29}$ から $a_1 = 29, a_2 = 21$

$t = -1$ から $a_1 = 1, a_2 = -1$

$$f(\mathbf{A}) = \frac{a_1^2 + 2a_1a_2 + a_2^2}{6a_1^2 - 9a_1a_2 + 10a_2^2}$$

$$f(t) = \frac{1 + 2t + t^2}{6 - 9t + 10t^2} = \frac{\frac{1}{t^2} + 2\frac{1}{t} + 1}{6\frac{1}{t^2} - 9\frac{1}{t} + 10}$$

t	$-\infty$	-1	$\frac{21}{29}$	$-\infty$
$\frac{\partial V}{\partial t}$		-	0	+
V	0.1	0	0.6289	0.1

結論。次のベクトル \mathbf{A} が、 $f(\mathbf{A})$ を最大化し、F比を最大化します。

$$f(\mathbf{A}) = \frac{a_1^2 + 2a_1a_2 + a_2^2}{6a_1^2 - 9a_1a_2 + 10a_2^2}$$

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 29 \\ 21 \end{pmatrix}$$

\mathbf{A} は基準超平面の法線ベクトルです。ここで、部分集団1と2の中心間の midpoint を含む、基準超平面に平行な超平面を境界超平面として選ぶと、判別スコアの境界値は以下の通りです。

$$DS_{threshold} = \begin{pmatrix} 29 \\ 21 \end{pmatrix}^T \frac{1}{2} \left(\begin{pmatrix} 7 \\ 6 \end{pmatrix} + \begin{pmatrix} 5 \\ 4 \end{pmatrix} \right) = (29 \ 21) \begin{pmatrix} 6 \\ 5 \end{pmatrix} = 29 \times 6 + 21 \times 5 = 289$$

VI-1-3-4. 線形代数学的な説明

ラグランジェの未定乗数法は制約条件のもとで、極値の条件と極値を見つける方法です。この方法は、関数の最大化や最小化にしばしば使われます。その基本的な考え方は超空間中で二つの軌跡が接するという事です。つまり、曲面と曲面、あるいは、曲面と平面が接するという事です。この時二つの平面は接面とそれに直交する法線を共有します。この二つの平面のうちの一つが最大化または最小化する関数で、もう一つが制約条件です。制約条件の軌跡は固定されています。最大化または最小化する関数の軌跡は制約条件の軌跡が示す領域の内側に存在したとします。ここで、最大化または最小化する関数の軌跡を広げていきます。最小に二つの軌跡が接した点が、関数が最大化または最小化する点です。さらに最大化または最小化する軌跡を広げていったときに、最後に二つの軌跡が接した点が、最小化または最大化する軌跡が最大の点です。これについては、V-2-6-3のラグランジェの未定乗数法で詳しく説明しています。この項で線形代数学的解説に使う論理も同じ論理を使いますが、関数が示す領域を拡大するのではなくて、回転します。ここでは超平面が制約条件なのですが、まだ固定されていません。むしろ、回転することによって、基準超平面と並行する平面を探すという感じです。

図70は楕円の回転と原点からのベクトルへの写像の長さの関係を示しています。楕円の短径がベクトルの方向と平行になったときに、写像の長さが最短となり、楕円の長径がベクトルと一平行になった時に最長となることは自明です。二つの部分集団があったとき、射影するためのベクトルは、二つの集団の中心を結ぶ線になるはずですが、部分集団の楕円の形と大きさは等分散を仮定しているのと同じですが、その楕円を回転させたときに、楕円の写像の長さと、2つの中心を結ぶ線の比が最小になるのは、二つの中心を結ぶ線と超楕円の最短径

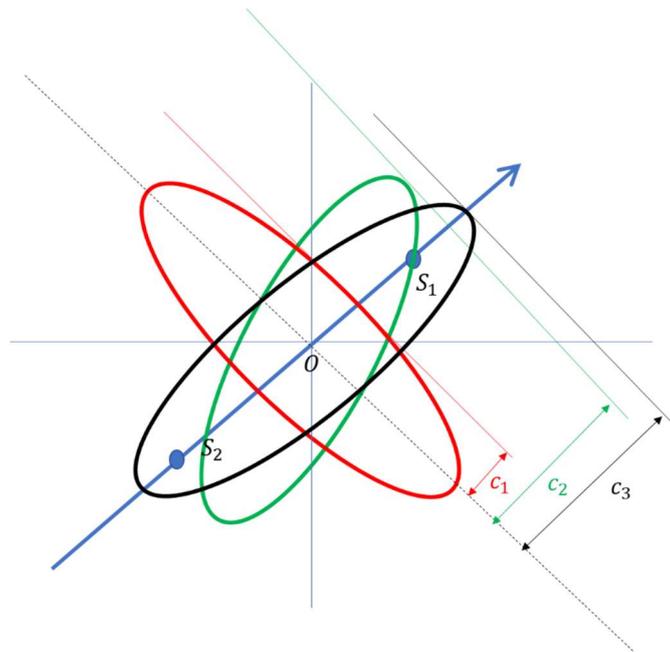


図70 楕円の回転とベクトルへの投影

が一致した時です。その時、もっとも二つの部分集団の違いが強調されることとなります。
これが基本的な考え方です。

ここでコーシー・シュワルツの不等式は、V-2-6-2で説明しました

$$(\mathbf{a}^t \mathbf{a})(\mathbf{b}^t \mathbf{b}) \geq (\mathbf{a}^t \mathbf{b})^2$$

そこで、基本の不等式w次のように変形できることを紹介しました。

$$\mathbf{a} = \mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha}$$

$$\mathbf{b} = \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta}$$

(\mathbf{B} は対称行列)

$$\left(\left(\mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} \right)^T \left(\mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} \right) \right) \left(\left(\mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right)^T \left(\mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right) \right) \geq \left(\left(\mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} \right)^T \left(\mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right) \right)^2$$

$$\left(\boldsymbol{\alpha}^t \mathbf{B}^{\frac{1}{2}T} \mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} \right) \left(\boldsymbol{\beta}^t \mathbf{B}^{-\frac{1}{2}T} \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right) \geq \left(\boldsymbol{\alpha}^t \mathbf{B}^{\frac{1}{2}T} \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right)^2$$

$$\left(\boldsymbol{\alpha}^T \mathbf{B}^{\frac{1}{2}} \mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} \right) \left(\boldsymbol{\beta}^T \mathbf{B}^{-\frac{1}{2}} \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right) \geq \left(\boldsymbol{\alpha}^T \mathbf{B}^{\frac{1}{2}} \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta} \right)^2$$

$$(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})(\boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha}) \geq (\boldsymbol{\alpha}^T \boldsymbol{\beta})^2$$

統合が成立つ条件は

$$\mathbf{a} = \mathbf{c} \mathbf{b}$$

です。 \mathbf{a} と \mathbf{b} のベクトルが、逆方向も含めて同じ方向を向いているときに、 $(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})(\boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha})$ が最小値 $(\boldsymbol{\alpha}^T \boldsymbol{\beta})^2$ になるという意味です。

さらに変形します。

$$\mathbf{B}^{\frac{1}{2}} \boldsymbol{\alpha} = \mathbf{c} \mathbf{B}^{-\frac{1}{2}} \boldsymbol{\beta}$$

$$\boldsymbol{\alpha} = \mathbf{c} \mathbf{B}^{-1} \boldsymbol{\beta}$$

両辺を $(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})$ で割ります。

$$\frac{(\boldsymbol{\alpha}^T \boldsymbol{\beta})^2}{\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha}} \leq \boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha}$$

$$(\because \boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha} > 0)$$

$$\frac{\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha}}{(\boldsymbol{\alpha}^T \boldsymbol{\beta})^2} \geq \boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha}$$

$$\left(\because \frac{(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})}{(\boldsymbol{\alpha}^T \boldsymbol{\beta})^2} > 0, \boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha} > 0, \right)$$

$$\frac{(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})}{(\boldsymbol{\alpha}^T \boldsymbol{\beta})(\boldsymbol{\alpha}^T \boldsymbol{\beta})} \geq \boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha}$$

$$\frac{(\boldsymbol{\alpha}^T \mathbf{B} \boldsymbol{\alpha})}{\boldsymbol{\alpha}^T \boldsymbol{\beta} \boldsymbol{\beta}^T \boldsymbol{\alpha}} \geq \boldsymbol{\alpha}^T \mathbf{B}^{-1} \boldsymbol{\alpha}$$

$$(\because \boldsymbol{\alpha}^T \boldsymbol{\beta} = \boldsymbol{\beta}^T \boldsymbol{\alpha})$$

ここで、 α を基準超平面の法線ベクトル、 β を部分集団の中心を結ぶベクトル、 $B^{\frac{1}{2}}$ を基準超平面を回転させる変形だと考えて、次の様に書き換えます。

$$\alpha = A, \beta = g, B = U,$$

$$\frac{A^T U A}{A^T g g^T A} \geq A^T U^{-1} A$$

等号条件は明らかで、 $A^T g, g^T A$ は内積ですから、ベクトルの方向が一致するときに最大値になります。わざわざ次のように計算する必要もないでしょう。

$$\alpha = c B^{-1} \beta$$

$$A = c U^{-1} g$$

$$U A = c g$$

したがって、この2つのベクトルの方向が一致したときに、 $\frac{A^T U A}{A^T g g^T A}$ が最大値です。

実際に、これを線形判別分析に当てはめるには、実際のデータから、 g と U をつくるらなければなりません。部分集団が2つならば、 g は2つの部分集団を結ぶベクトルに決まっています。 U については、次の残渣分散の分散共分散行列(V)を使うという考えかたもありますが

$$V = \sum_{k=1}^m \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)^T$$

しかし、この場合は、全分散を決めている分散は、部分集団間の分散と残渣分散しかないのだから、

$$SS_{total} = SS_{residual} + SS_{subpopulation}$$

この両辺を残渣分散で割って

$$\frac{SS_{total}}{SS_{subpopulation}} = \frac{SS_{residual}}{SS_{subpopulation}} + 1$$

このように変形して、 $\frac{SS_{total}}{SS_{subpopulation}}$ が最小になれば、 $\frac{SS_{residual}}{SS_{subpopulation}}$ が最大化すると考えて、

SS_{total} が示す楕円の法線ベクトルへの投影図の長さを基準超平面を回転させて、法線を回転させることによって最小化する という計算にします。 $V = SS_{residual}$ を計算する必要はありません。これが、超楕円の回転による、線形判別分析の説明です。

練習

以前に使ったデータセットで、回転による線形判別分析を練習してみます。

データセット

データ

部分集団	標本番号	d_1	d_2
1	1	5	8
1	2	7	4
1	3	8	5
1	4	8	7
2	1	5	5
2	2	7	2
2	3	4	3
2	4	4	6

全平均

$$\bar{\bar{d}}_1 = \frac{5 + 7 + 8 + 8 + 5 + 7 + 4 + 4}{8} = \frac{48}{8} = 6$$

$$\bar{\bar{d}}_2 = \frac{8 + 4 + 5 + 7 + 5 + 2 + 3 + 6}{8} = \frac{40}{8} = 5$$

$$\bar{\bar{\mathbf{d}}} = \begin{pmatrix} 6 \\ 5 \end{pmatrix}$$

$$\bar{d}_{11} = \frac{5 + 7 + 8 + 8}{4} = \frac{28}{4} = 7$$

$$\bar{d}_{12} = \frac{5 + 7 + 4 + 4}{4} = \frac{20}{4} = 5$$

$$\bar{d}_{21} = \frac{8 + 4 + 5 + 7}{4} = \frac{24}{4} = 6$$

$$\bar{d}_{22} = \frac{5 + 2 + 3 + 6}{4} = \frac{16}{4} = 4$$

$$\bar{\mathbf{d}}_1 = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

$$\bar{\mathbf{d}}_2 = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

部分集団の中心を結ぶベクトル

$$\mathbf{g} = c(\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2) = c \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

各データの全平均からの偏差

$$\mathbf{x}_{11} = \begin{pmatrix} 5 \\ 8 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \mathbf{x}_{12} = \begin{pmatrix} 7 \\ 4 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{x}_{13} = \begin{pmatrix} 8 \\ 5 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \mathbf{m}_{14} = \begin{pmatrix} 8 \\ 7 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$\mathbf{x}_{21} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{x}_{22} = \begin{pmatrix} 7 \\ 2 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}, \mathbf{x}_{23} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \mathbf{m}_{24} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} - \begin{pmatrix} 6 \\ 5 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$\mathbf{u}^T = \begin{pmatrix} -1 & 1 & 2 & 2 & -1 & 1 & -2 & -2 \\ 3 & -1 & 0 & 2 & 0 & 3 & -2 & 1 \end{pmatrix}$$

$$\mathbf{U} = \mathbf{u}\mathbf{u}^T = \begin{pmatrix} 20 & -1 \\ -1 & 28 \end{pmatrix}$$

$$\mathbf{U}^{-1} = \frac{1}{559} \begin{pmatrix} 28 & 1 \\ 1 & 20 \end{pmatrix}$$

$$\mathbf{A} = c\mathbf{U}^{-1}\mathbf{g}$$

$$\mathbf{A} = c \frac{1}{559} \begin{pmatrix} 28 & 1 \\ 1 & 20 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = c \frac{1}{559} \begin{pmatrix} 29 \\ 21 \end{pmatrix}$$

$c = 559$ を選択

$$\mathbf{A} = \begin{pmatrix} 29 \\ 21 \end{pmatrix}$$

これが結論です。分散分析で行う線形判別分析と線形代数的に行う線形判別分析が同じ結果になることが確かめられました。線形代数的な方法のメリットは計算過程が単純だということです。分散を分離する必要がありません。弱点は、線形代数学を知らないと、その原理の理解が難しいことかもしれません。しかし、要領を覚えると、その理論も単純で難しくありません。著者は、背景の理解のために、V-2-6. 最大・最小のラグランジェの未定乗数法を再読することをお勧めします。

問題になるのは、ベクトル \mathbf{g} の選択です。練習に使った例では、部分集団が2つだから、その中点を結ぶ直線が選べます。その直線の長さが、二つの集団の距離を表しているからです。部分集団が3つ以上あった時にどうするかを考えておかななくてはなりません。一つは、いくつかの部分集団の中心間に回帰直線を引いて、それを用いることです。しかし、変数の数が集団の数よりも多い時にはそれが出来ません。ただし、筆者はそんなことが起こったケースを知りません。判別分析は2部分集団間で行うことが多いのです。そもそも、変数間の関係、集団間の関係が複雑な時に、線形判別分析を使うべきでないでしょう。主成分分析、因子分析、多次元尺度構成法、クラスター分析等々、代替の方法はありまし、コンピュータの計算速度に頼れば線形でない判別分析も可能です。線形判別分析の利点は、判別スコアが単純に基準超平面からの距離なので、境界値の選択に私たちの政策的判断や考え方が持ち込めることです。

VI-1-3-5. 判別分析の境界値

線形判別分析で基準超平面の傾きを最適化する方法を学びました。一つの例として、全平均を通る平面を判別平面として選んだ例を図70に示します。データセットは線形判別分析の練習に使ったものです。何故、全体の平均通る平面を選択したのかという理由は、ほかに情報がないからです。全体のデータの分布を見渡して、とりあえず2集団間の等分散性を仮定しました。情報が限られていますから、その判断の妥当性に自信があるわけではありません。

この場合、2 集団のサンプル数が同じだから、集団間の中点と全体の平均は一致します。実際に、図 71 は、その方法で得られた境界値で集団を分けると、既往のデータが正しく分けられたことを示しています。

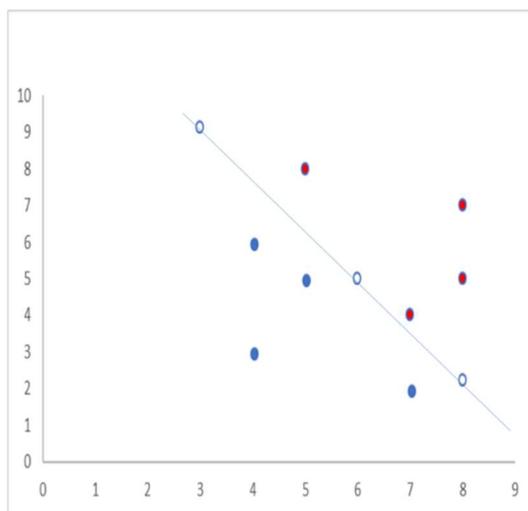


図 71. 境界値 $Z = 29d_1 + 21d_2 = 279$ でデータを判別した結果
赤丸：部分集団 1. 青丸：部分集団 2. 青線：境界線

境界値による判別

部分手段	d_1	d_2	判別スコア	境界値(279)との差
1	5	8	313	34
1	7	4	287	8
1	8	5	337	58
1	8	7	379	100
2	5	5	250	-29
2	7	2	245	-34
2	4	3	179	-100
2	4	6	242	-37

VI-1-3-6. 判別分析の応用に関する議論.

各部分集団の分布を正規分布と仮定すれば、部分集団の中心からの距離を偏差で標準化すれば、その点にデータがある確率を求めることが出来ます。上にあげた例では、境界線は中心から等距離のところにあるから、境界値上のデータを部分集団のどちらかに属するかを

判定した場合、その誤りのリスクは、どちらに判定しても同じです。しかし、いつでもそのように境界平面を決めなくてはならないわけではありません。多くの場合、正しくはBのものをAと判定する事と、正しくはAのものをBと判定することの意味は違います。たとえば、病気の簡易判定の場合、病人を健康であると診断する間違いと、健康な人を病人だと判断することの深刻さは違います。目的は多くの病気の人を発見することで、健康な人を発見することではありません。病気の人を見逃してはいけません。こういう場合、病気の人を健康だと判断してしまう可能性を最小化するように境界値を設定します。つまり、図 72 に示したように、境界線（境界平面）を健常者の平均に近づけるのです。

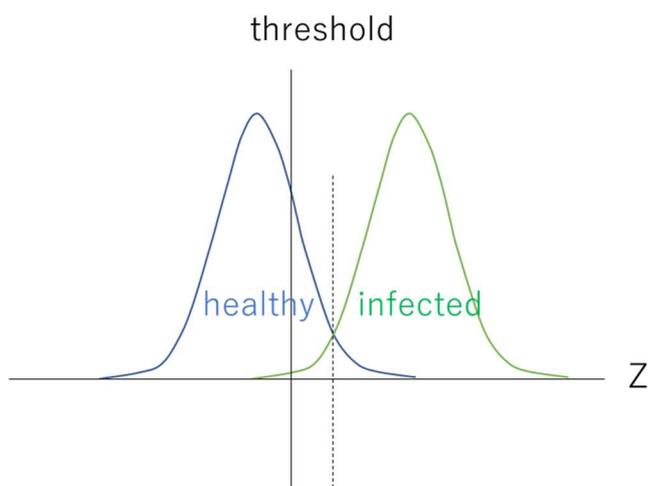


図 72. 判別スコアの境界値の最適化.

私たちは、分散分析を知らなくても、微分を知らなくても、線形判別分析ができることを学びました。そのことは重回帰と同じです。また、変数の数の適正化も重回帰と同様に重要です。たとえば、患者に不必要な負担をかけないために、測定項目を減らすことは、診断学では重要です。その他の応用場面でも同じような必要があります。多すぎる変数は、非効率だし、余計な混乱を招きます。この解説では、部分集団間の等分散性を仮定しました。これは、場合によっては不自然です。その一方で、変数間の等分散性は無視しています。重回帰と同様に、場合によっては、元データを偏差で割って標準化した方が良いかもしれません。そうすると、データ分布が超球体のように同心円状に分布するので、線形判別分析に意味がなくなって、重み付き平均と大して変わらないかもしれません。もし、私たちが、変数間の直交性を意識するのであれば、初めに主成分分析を行って、主成分得点を使うという方法も考えられるかもしれません。しかし、主成分の解釈について不必要な混乱を招くかもしれません。おそらく、あまり効率的ではないでしょう。すでに述べたように、コンピュータの計算速度に頼れば、等分散を前提としない判別分析もいくつかあります。これについては別の章で解説します。判別分析については多くの議論があり、私たちは目的・状況・データの性質に応じて適正な判別方法を選ぶ必要があります。