判別分析で
特徴で見分ける

判別分析とは？

体操部

番号	身長	体重
1	151	45
2	162	47
3	168	50
4	155	51

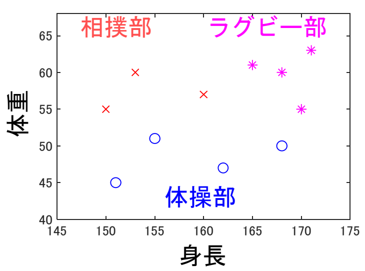
ラグビー部

番号	身長	体重
1	168	60
2	170	55
3	171	63
4	165	61

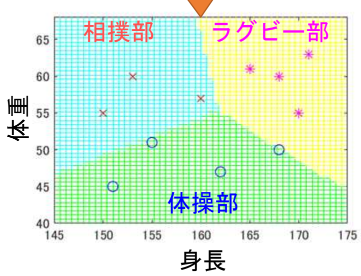
相撲部

番号	身長	体重
1	150	55
2	160	57
3	153	60

境界を決める



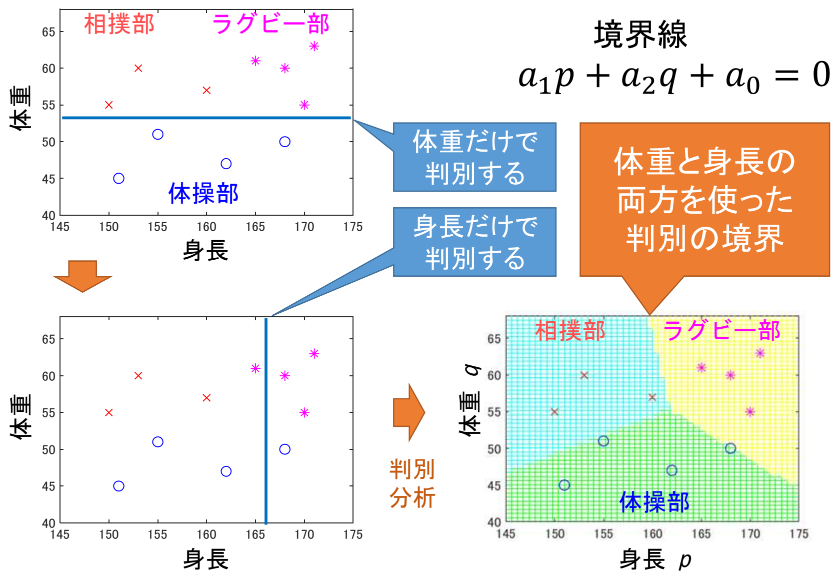
判別
分析

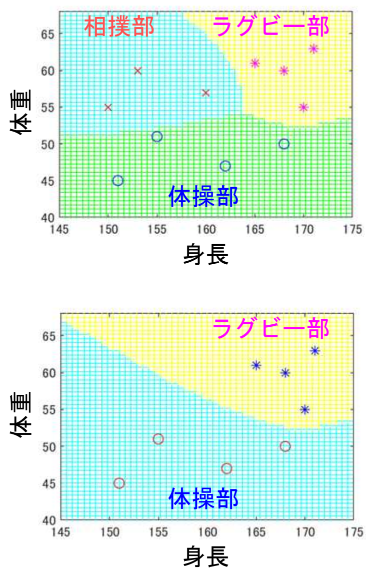
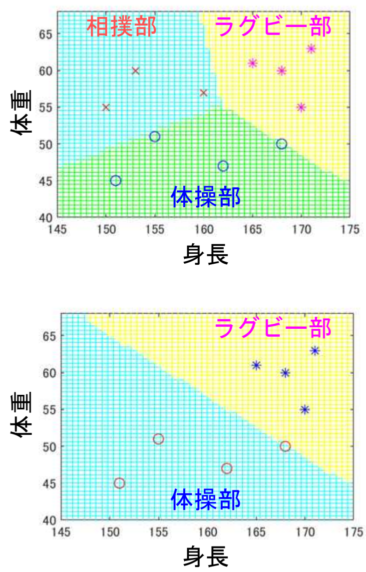


複数の特徴で 分類する

境界線

$$a_1 p + a_2 q + a_0 = 0$$



境界が曲線（2次式）	境界が直線（1次式）
	

最新版

LDA：無し
正規化：無し
不偏分散：無し
２次判曲線：無し

```
Fmin2)+Fmin2; Z2(j)=y;
    for i=1:L; x=(i-1)/L*(Fmax1-
Fmin1)+Fmin1; Z1(i)=x;
                  G1(i,j)=mvnpdf([x
y]',m1',s1);
                  G2(i,j)=mvnpdf([x
y]',m2',s2);
        if(Nc==3) G3(i,j)=mvnpdf([x
y]',m3',s3); end
    end
end
figure('Position',[010 010 300 200]);%
図１
%------------------------- 等高線
Gmax=max(max(G1(:)),max(G2(:)));
if(Nc==3) Gmax=max(Gmax,max(G3(:)));
end
          contour(Z1,Z2,G1'/Gmax);
hold on;
          contour(Z1,Z2,G2'/Gmax);
hold on;
if(Nc==3) contour(Z1,Z2,G3'/Gmax);
hold on; end
colormap('jet');
%------------------------- 入力データ
```

線形判別 をつかう

水俣病のネコ

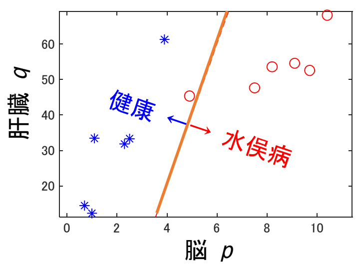
番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

知床海岸のネコの
脳と肝臓の水銀量
（ppm）

石村貞夫「すぐわかる
多変量解析」東京図書



境界線の方程式

$$q = -(a_1 p + a_0)/a_2$$

$$= \ 19.84\,p - 59.04$$

判別理由を 分析する

水俣病のネコ

番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

知床海岸のネコの
脳と肝臓の水銀量
(ppm)

石村貞夫「すぐわかる
多変量解析」東京図書

$$19.84\,\overset{\text{脳}}{p} - \overset{\text{肝臓}}{q} = 59.04$$

差に意味がある

脳が重視される

境界線の方程式

理由を
分析

$$q = -(a_1 p + a_0)/a_2$$

$$= 19.84\,p - 59.04$$

２次判別 をつかう

水俣病のネコ

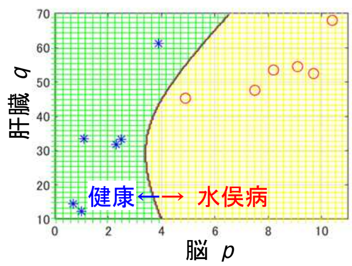
番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

知床海岸のネコの
脳と肝臓の水銀量
(ppm)

石村貞夫「すぐわかる
多変量解析」東京図書



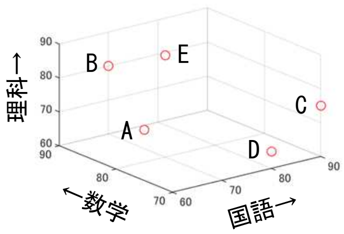
境界線の方程式

２次曲線が得られる

→ 高精度に判別

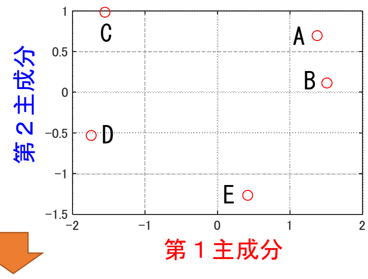
主成分分析 と併用する

3次元のデータ

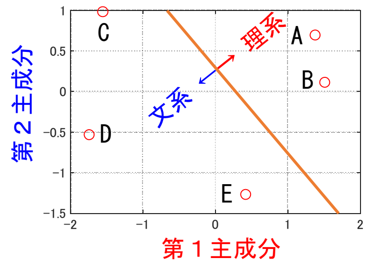


主成分分析
PCA

2次元のデータ



判別分析
LDA

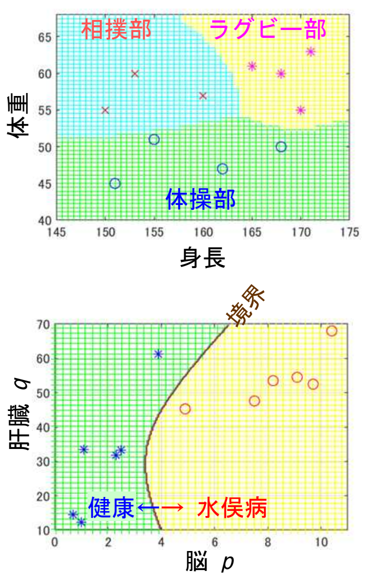


例えば,
文系か理系か
↓
より正確に判別

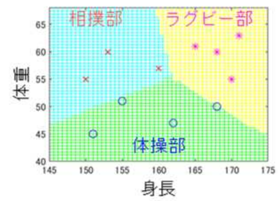
判別分析で
特徴で見分ける

1. 線形判別とは？
2. マハラノビス距離
3. 線形判別の手順
4. マハラノビス距離の計算法
5. 主成分分析（PCA）との併用

2次判別



1次判別



境界は直線（1次式）

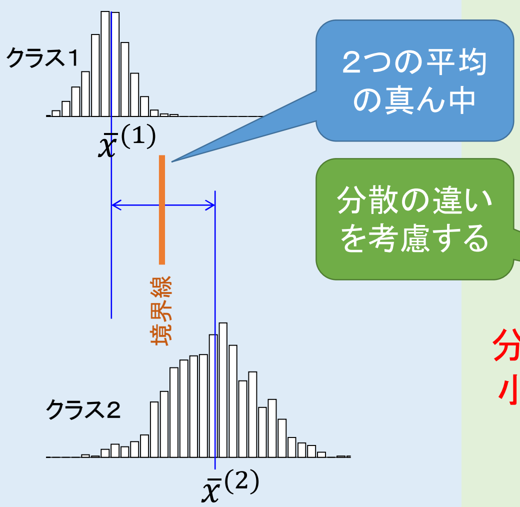
境界は曲線（2次式）

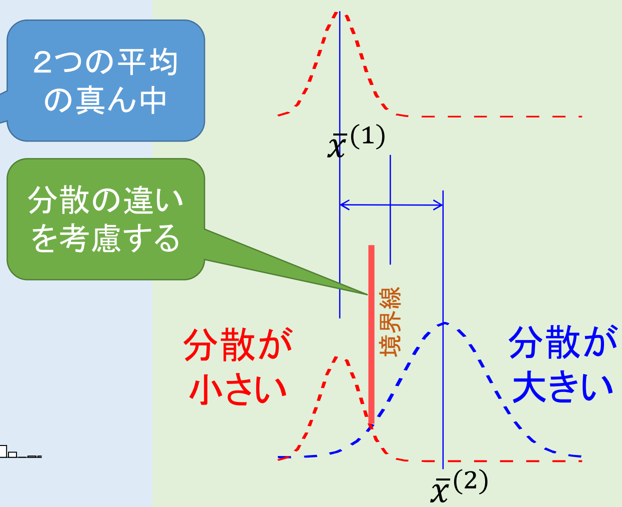
マハラノビス距離を使う

境界 の 定め方（1次元）

2つの平均の中点

マハラノビス距離





２次判曲線：あり

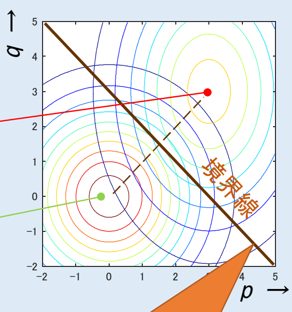
```
Fmin2=40;
Fmax2=68;

L=50;

for j=2:L;      y=(j-1)/L*(Fmax2-Fmin2)+Fmin2; Z2(j)=y;
    for i=2:L; x=(i-1)/L*(Fmax1-Fmin1)+Fmin1; Z1(i)=x;
        G1(j,i)=mvnpdf([x y]',m1',s1);
        G2(j,i)=mvnpdf([x y]',m2',s2);
        G3(j,i)=mvnpdf([x y]',m3',s3);
    end
end

figure('Position',[010 010 300 200]);
contour(Z1,Z2,G1);        hold on;
contour(Z1,Z2,G2);        hold on;
contour(Z1,Z2,G3);        hold on;
axis([Fmin1,Fmax1,Fmin2,Fmax2]);
colormap('jet');

plot(d1(:,1),d1(:,2),'ob'); hold on;
plot(d2(:,1),d2(:,2),'*m'); hold on;
plot(d3(:,1),d3(:,2),'xr'); hold on;
axis([Fmin1,Fmax1,Fmin2,Fmax2]);

%---------------------- マハラノビス境界線
figure('Position',[010 210 300 200]);
for j=1:L;      y=(j-1)/L*(Fmax2-Fmin2)+Fmin2; Z2(j)=y;
    for i=1:L; x=(i-1)/L*(Fmax1-Fmin1)+Fmin1; Z1(i)=x;
        if( (G1(j,i)>G2(j,i)) && (G1(j,i)>G3(j,i)) )
plot(x,y,'+g'); hold on; end
        if( (G2(j,i)>G1(j,i)) && (G2(j,i)>G3(j,i)) )
plot(x,y,'+y'); hold on; end
        if( (G3(j,i)>G1(j,i)) && (G3(j,i)>G2(j,i)) )
plot(x,y,'+c'); hold on; end
    end
end
```

境界 の 定め方（2次元）

重心の中線

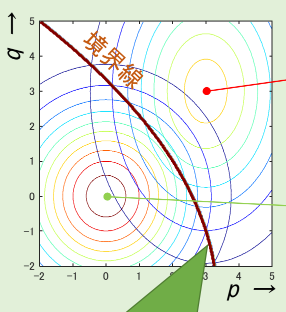
$\overline{x}^{(1)} = [3 \quad 3]^{\mathrm{T}}$

$\overline{x}^{(2)} = [0 \quad 0]^{\mathrm{T}}$



2つのクラスの重心の中間にある境界

マハラノビス距離



$S^{(1)} = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}$

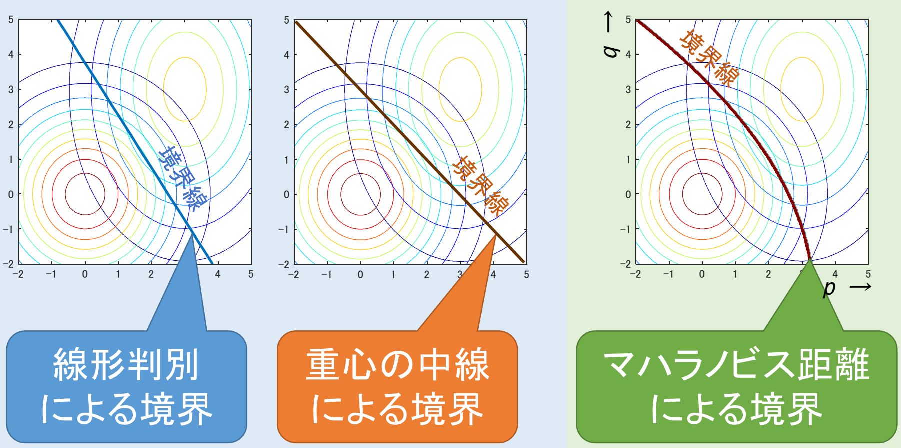
$S^{(2)} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

分散や共分散の違いを考慮した境界

記号 T は転置を表す

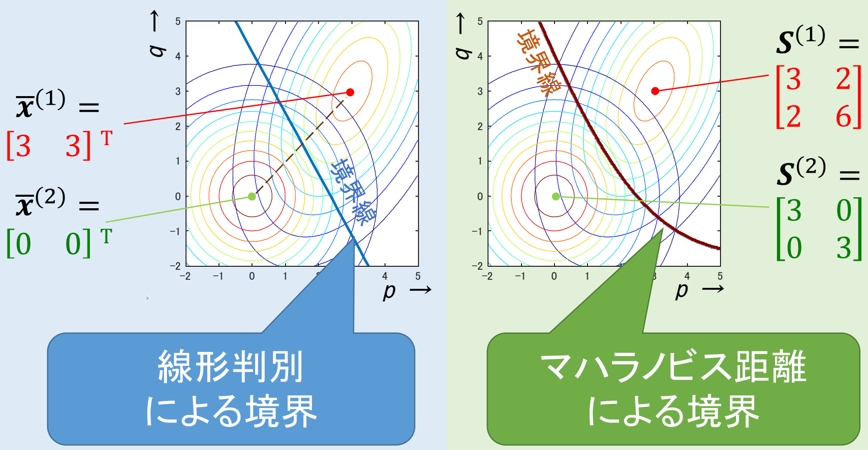
境界 の 定め方（2次元）

1次判別　　2次判別



境界 の 定め方（2次元）

1次判別　　2次判別



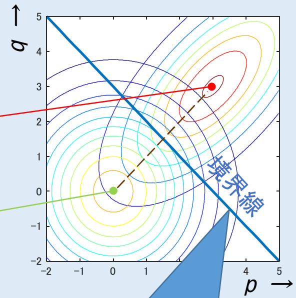
記号 T は転置を表す

境界 の 定め方（2次元）

1次判別

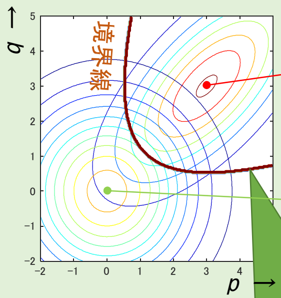
$\overline{\boldsymbol{x}}^{(1)} = [3 \quad 3]^{\mathrm{T}}$

$\overline{\boldsymbol{x}}^{(2)} = [0 \quad 0]^{\mathrm{T}}$



線形判別
による境界

2次判別



$\boldsymbol{S}^{(1)} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$

$\boldsymbol{S}^{(2)} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

マハラノビス距離
による境界

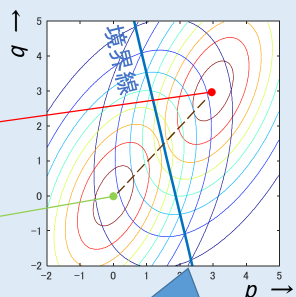
記号 T は転置を表す

境界 の 定め方（2次元）

1次判別

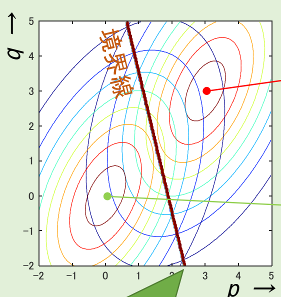
$\overline{\boldsymbol{x}}^{(1)} = [3 \quad 3]^{\mathrm{T}}$

$\overline{\boldsymbol{x}}^{(2)} = [0 \quad 0]^{\mathrm{T}}$



線形判別
による境界

2次判別



$\boldsymbol{S}^{(1)} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

$\boldsymbol{S}^{(2)} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

マハラノビス距離
による境界

$\boldsymbol{S}^{(1)} = \boldsymbol{S}^{(2)}$ ならば境界線は同じ

記号 T は転置を表す

```
    for i=1:L; x=(i-1)/L*(Fmax1-
Fmin1)+Fmin1; Z1(i)=x;
        G1(i,j)=mvnpdf([x
y]',m1',sgm1);
        G2(i,j)=mvnpdf([x
y]',m2',sgm2);
    end
end

figure('Position',[10 10 300 300]);
contour(Z1,Z2,G1');         hold on;
contour(Z1,Z2,G2');         hold on;
colormap('jet');
for j=1:L
    for i=1:L
        if(G1(i,j)>G2(i,j)) G3(i,j)=1;
else G3(i,j)=0; end
    end
end
G3=G3*max(max(G1(:)),max(G2(:)));
contour(Z1,Z2,G3','LineWidth',2);
axis([Fmin1,Fmax1,Fmin2,Fmax2]);

figure('Position',[10 310 300 300]);
contour(Z1,Z2,G1');         hold on;
```

境界 の 関数

$$a_1 p + a_2 q + a_0$$
$$+ a_3 p^2 + a_4 q^2 + a_5 pq$$
$$= 0$$

2次判別は2次式

マハラノビス距離
による境界線

$$a_1 p + a_2 q + a_0$$
$$= 0$$

1次判別は1次式

重心の中線
となる境界線

線形判別
による境界線

境界 の 関数

$$a_1 p + a_2 q + a_0 + a_3 p^2 + a_4 q^2 + a_5 pq = 0$$

2次式

$$\frac{1}{\sqrt{|\boldsymbol{S}|(2\pi)^2}} e^{-\frac{(\boldsymbol{x}-\overline{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{S}^{-1} (\boldsymbol{x}-\overline{\boldsymbol{x}})}{2}}$$

点 $\boldsymbol{x}$ と点 $\overline{\boldsymbol{x}}$ の間の
マハラノビス距離

$$a_1 p + a_2 q + a_0 = 0$$

1次式

$$\frac{a_0 + a_1 p + a_2 q}{\sqrt{a_1^2 + a_2^2}}$$

直線と点 (p, q) の間の
ユークリッド距離

参考

$\boldsymbol{x} = \begin{bmatrix} p \\ q \end{bmatrix}$, $\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{p} \\ \overline{q} \end{bmatrix}$, $\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$ とすると、

但し、$s_{21} = s_{12}$

$$(\boldsymbol{x}-\overline{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{S}^{-1} (\boldsymbol{x}-\overline{\boldsymbol{x}}) = \begin{bmatrix} \underbrace{p \quad q}_{\leftarrow 1次 \rightarrow} & 1 & \underbrace{p^2 \quad q^2 \quad pq}_{\leftarrow 2次 \rightarrow} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_0 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}$$

ここで、

$$\begin{bmatrix} a_1 \\ a_2 \\ a_0 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \frac{1}{s_{11}s_{22} - {s_{12}}^2} \begin{bmatrix} 2(s_{12}\overline{q} - s_{22}\overline{p}) \\ 2(s_{12}\overline{p} - s_{11}\overline{q}) \\ s_{22}\overline{p}^2 + s_{11}\overline{q}^2 - 2s_{12}\overline{p}\overline{q} \\ s_{22} \\ s_{11} \\ -2s_{12} \end{bmatrix}$$

判別分析で
特徴で見分ける

1. 線形判別とは？
2. マハラノビス距離
3. 線形判別の手順
4. マハラノビス距離の計算法
5. 主成分分析(PCA)との併用

手順① データを文字で表現

水俣病のネコ

番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

$\boldsymbol{x}_1^{(1)} = [\ 9.1 \quad 54.5\ ]^{\mathrm{T}}$　　$\boldsymbol{x}_1^{(2)} = [\ 2.3 \quad 31.8\ ]^{\mathrm{T}}$　　Tは転置

$\boldsymbol{x}_2^{(1)} = [10.4 \quad 68.0\ ]^{\mathrm{T}}$　　$\boldsymbol{x}_2^{(2)} = [\ 0.7 \quad 14.5\ ]^{\mathrm{T}}$

$\vdots$　　$\vdots$

$\boldsymbol{x}_6^{(1)} = [\ 4.9 \quad 45.3\ ]^{\mathrm{T}}$　　$\boldsymbol{x}_6^{(2)} = [\ 1.0 \quad 12.3\ ]^{\mathrm{T}}$

手順② クラス毎の平均

$\overline{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}}$ $\quad\quad$ $\overline{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$

$$\overline{\boldsymbol{x}}^{(k)} = \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \boldsymbol{x}_i^{(k)}, \quad \text{クラス } k = 1,2$$

$$\boldsymbol{x}_1^{(1)} - \overline{\boldsymbol{x}}^{(1)} = [\ 0.80 \quad 0.93\ ]^{\mathrm{T}}$$

0.80	0.93
2.10	14.43
−0.10	−0.07
−0.80	−5.97
1.40	−1.07
−3.40	−8.27

$$\boldsymbol{x}_1^{(2)} - \overline{\boldsymbol{x}}^{(2)} = [\ 0.38 \quad 0.72\ ]^{\mathrm{T}}$$

0.38	0.72
−1.22	−16.58
0.58	2.22
−0.82	2.32
1.98	30.12
−0.92	−18.78

手順③ 分散共分散行列

$$\boldsymbol{x}^{(1)} - \overline{\boldsymbol{x}}^{(1)} = \begin{bmatrix} 0.80 & 0.93 \\ 2.10 & 14.43 \\ -0.10 & -0.07 \\ -0.80 & -5.97 \\ 1.40 & -1.07 \\ -3.40 & -8.27 \end{bmatrix}^{\mathrm{T}}, \quad N^{(1)} = 6$$

$$\boldsymbol{x}^{(2)} - \overline{\boldsymbol{x}}^{(2)} = \begin{bmatrix} 0.38 & 0.72 \\ -1.22 & -16.58 \\ 0.58 & 2.22 \\ -0.82 & 2.32 \\ 1.98 & 30.12 \\ -0.92 & -18.78 \end{bmatrix}^{\mathrm{T}}, \quad N^{(2)} = 6$$

$$\boldsymbol{S}^{(k)} = \underbrace{\frac{1}{N^{(k)} - 1}}_{\text{不偏分散}} \left(\boldsymbol{x}^{(k)} - \overline{\boldsymbol{x}}^{(k)}\right)\left(\boldsymbol{x}^{(k)} - \overline{\boldsymbol{x}}^{(k)}\right)^{\mathrm{T}}, \quad k = 1,2$$

$$\boldsymbol{S}^{(1)} = \begin{bmatrix} 3.84 & 12.49 \\ 12.49 & 62.85 \end{bmatrix} \quad\quad \boldsymbol{S}^{(2)} = \begin{bmatrix} 1.48 & 19.36 \\ 19.36 & 309.13 \end{bmatrix}$$

手順④　クラス内分散

$$\boldsymbol{S}_w = \sum_{k=1}^{2} \frac{N^{(k)} - 1}{N^{(1)} + N^{(2)} - 2} \boldsymbol{S}^{(k)}$$

$$\boldsymbol{S}_w = \frac{(6-1)\boldsymbol{S}^{(1)} + (6-1)\boldsymbol{S}^{(2)}}{6+6-2}$$

$$= \begin{bmatrix} 2.66 & 15.93 \\ 15.93 & 185.99 \end{bmatrix}$$

手順⑤　全体の平均

$$\bar{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}} \qquad \bar{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\bar{\boldsymbol{x}} = \sum_{k=1}^{2} \frac{N^{(k)}}{N^{(1)} + N^{(2)}} \bar{\boldsymbol{x}}^{(k)}$$

$$\bar{\boldsymbol{x}} = [5.11 \quad 42.33]^{\mathrm{T}}$$

手順⑥　クラス間分散

$$\bar{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}}$$

$$\bar{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\bar{\boldsymbol{x}} = [5.11 \quad 42.33]^{\mathrm{T}}$$

$$\boldsymbol{S}_b = \sum_{k=1}^{2} \frac{N^{(k)}}{N^{(1)} + N^{(2)}} \left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}^{(k)}\right)\left(\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}^{(k)}\right)^{\mathrm{T}}$$

$$\boldsymbol{S}_b = \begin{bmatrix} 10.19 & 35.88 \\ 35.88 & 126.38 \end{bmatrix}$$

手順⑦　固有値問題

$$\boldsymbol{S}_w = \begin{bmatrix} 2.66 & 15.93 \\ 15.93 & 185.99 \end{bmatrix}$$

クラス内分散

$$\boldsymbol{S}_b = \begin{bmatrix} 10.19 & 35.88 \\ 35.88 & 126.38 \end{bmatrix}$$

クラス間分散

$$\boldsymbol{S}_b \boldsymbol{U} = \boldsymbol{S}_w \boldsymbol{U} \boldsymbol{L}$$

[V, D]=eig(Sb, Sw)
MATLAB

$$\boldsymbol{U} = \begin{bmatrix} 0.34 & -0.81 \\ -0.10 & 0.04 \end{bmatrix} \qquad \boldsymbol{L} = \begin{bmatrix} 0 & 0 \\ 0 & 4.50 \end{bmatrix}$$

原田達也, 「画像認識」, 講談社MLPシリーズ

手順⑦の意味 「最小二乗法」

$$maximize \ \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u}}$$

クラス間分散 → 大きく
クラス内分散 → 小さく

$$maximize \ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u} \quad s.t. \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u} = 1$$

$$J(\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u} + \lambda\left(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u} - 1\right)$$

$$\frac{\partial J(\boldsymbol{u})}{\partial(\boldsymbol{u})} = 2\,\boldsymbol{S}_b\boldsymbol{u} - 2\,\lambda\boldsymbol{S}_w\boldsymbol{u} = 0$$

$$\boxed{\boldsymbol{S}_b\boldsymbol{U} = \boldsymbol{S}_w\boldsymbol{U}\boldsymbol{L}}$$

原田達也「画像認識」講談社

参考

手順⑦'

$$\boldsymbol{S}_w = \begin{bmatrix} 2.66 & 15.93 \\ 15.93 & 185.99 \end{bmatrix}$$

$$\overline{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}}$$
$$\overline{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\boldsymbol{u} = \boldsymbol{S}_w^{-1}\left(\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}\right)$$

A=(inv(Sw)*(m1-m2))
MATLAB

$$[a_0 \quad a_1 \quad a_2] = [-10.21 \quad 3.43 \quad -0.17]$$

$$q = -\frac{a_1}{a_2}p - \frac{a_0}{a_2} \ = \ 19.84p - 59.04$$

石村貞夫「すぐわかる多変量解析」東京図書

手順⑧　クラスの中点

$$\bar{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}}$$

$$\bar{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\boldsymbol{U} = \begin{bmatrix} 0.34 & -0.81 \\ -0.10 & 0.04 \end{bmatrix}$$

$$a_0 = \frac{-[a_1 \quad a_2] \cdot \left(\bar{\boldsymbol{x}}^{(1)} + \bar{\boldsymbol{x}}^{(2)}\right)}{2}$$

$$a_1 = -0.81$$

$$a_2 = 0.04$$

$$a_0 = 2.41$$

$$[a_0 \quad a_1 \quad a_2] = [2.41 \quad -0.81 \quad 0.04]$$

クラス境界 の 計算結果

水保病のネコ

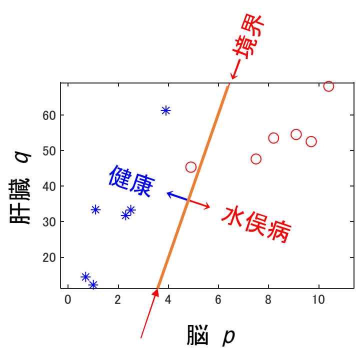
番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

境界線からの
ユークリッド距離

水保病	健康ネコ
−3.37	2.28
−3.99	3.00
−2.53	2.15
−2.12	3.55
−4.07	2.16
0.36	2.59



$$D(p, q) = \frac{a_0 + a_1 p + a_2 q}{\sqrt{a_1^2 + a_2^2}}$$

p, q の 1 次式

$$q = -\frac{a_1}{a_2} p - \frac{a_0}{a_2}$$

$$= 19.84p - 59.04$$

判別理由を 分析する

水俣病のネコ

番号	脳	肝臓
1	9. 1	54. 5
2	10. 4	68. 0
3	8. 2	53. 5
4	7. 5	47. 6
5	9. 7	52. 5
6	4. 9	45. 3

健康なネコ

番号	脳	肝臓
1	2. 3	31. 8
2	0. 7	14. 5
3	2. 5	33. 3
4	1. 1	33. 4
5	3. 9	61. 2
6	1. 0	12. 3

知床海岸のネコの
脳と肝臓の水銀量
(ppm)

石村貞夫「すぐわかる
多変量解析」東京図書

$$19.84\,\underset{}{\overset{\text{脳}}{p}} - \overset{\text{肝臓}}{q} = 59.04$$

差に意味がある

脳が重視される

理由を
分析

境界線の方程式

$$q = -(a_1 p + a_0)/a_2$$

$$= \ 19.84\,p - 59.04$$

参考

手順⑨ クラスの中点(その2)

$$\bar{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}}$$

$$\bar{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\boldsymbol{U} = \begin{bmatrix} 0.34 & -0.81 \\ -0.10 & 0.04 \end{bmatrix}$$

$$b_0 = \frac{-[a_1 \quad a_2]\cdot\left(\bar{\boldsymbol{x}}^{(1)} + \bar{\boldsymbol{x}}^{(2)}\right)}{2}$$

$$b_0 = 2.35$$

$$\begin{aligned} b_1 &= \ \ 0.34 \\ b_2 &= -0.10 \end{aligned}$$

$$[b_0 \quad b_1 \quad b_2] = [2.35 \quad 0.34 \quad -0.10]$$

参考

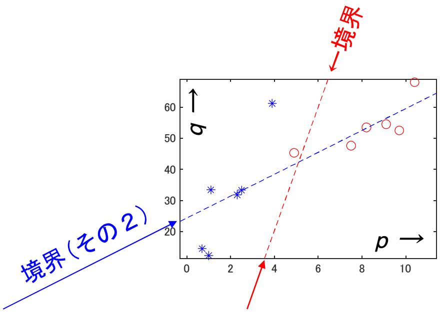
境界（その2）の計算結果

水俣病のネコ

番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3



$$q = -\frac{b_1}{b_2}p - \frac{b_0}{b_2}$$

$$= 3.52p + 24.33$$

$$q = -\frac{a_1}{a_2}p - \frac{a_0}{a_2}$$

$$= 19.84p - 59.04$$

判別分析で特徴で見分ける

1. 線形判別とは？
2. マハラノビス距離
3. 線形判別の手順 MATLAB
4. マハラノビス距離の計算法
5. 主成分分析（PCA）との併用

MATLAB

①　元のデータ

```
clear all; close all;

d1=[9.1000    54.5000;% 水俣病のネコ
    10.4000   68.0000;
     8.2000   53.5000;
     7.5000   47.6000;
     9.7000   52.5000;
     4.9000   45.3000];
d2=[2.3000    31.8000;% 健康なネコ
     0.7000   14.5000;
     2.5000   33.3000;
     1.1000   33.4000;
     3.9000   61.2000;
     1.0000   12.3000];

MM=length(d1(1,:));% 属性の種類
N1=length(d1(:,1));% クラス1のデータ数
N2=length(d2(:,1));% クラス2のデータ数
NN=N1+N2;
```

MATLAB

②　データの線形判別

```
m1=mean(d1) % 平均ベクトル（クラス１）
m2=mean(d2) % 平均ベクトル（クラス２）

x1=d1-m1     % 平均偏差行列（クラス１）
x2=d2-m2     % 平均偏差行列（クラス２）

S1=x1'*x1/(N1-1) % 分散共分散行列（クラス１）
S2=x2'*x2/(N2-1) % 分散共分散行列（クラス２）

mm=(N1*m1+N2*m2)/(N1+N2);
Sb=(N1*(mm-m1)'*(mm-m1)+N2*(mm-m2)'*(mm-m2))/NN; % クラス間分散
Sw=((N1-1)*S1+(N2-1)*S2)/(NN-2);                 % クラス内分散

[V,D]=eig(Sb,Sw) % 固有値問題

A(:)=V(:,MM  );  % 係数ベクトル１
B(:)=V(:,MM-1);  % 係数ベクトル２

A0=-(m1*A'+m2*A')/2; % 中点１
B0=-(m1*B'+m2*B')/2; % 中点２
```

MATLAB

全て

古い

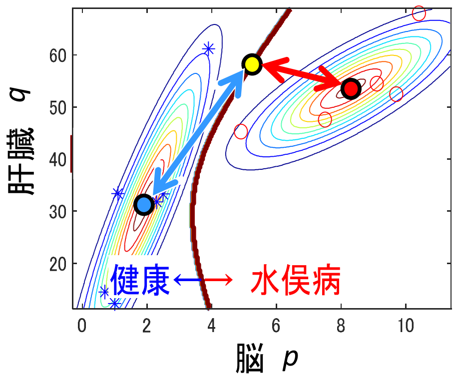
LDA：あり
正規化：あり
不偏分散：あり
2次曲線:なし

```
[V,D]=eig(Sb,Sw); % 固有値問題

A(:)=V(:,MM); B(:)=V(:,MM-1);
A=[-(m1*A'+m2*A')/2 A];
B=[-(m1*B'+m2*B')/2 B];

figure('Position',[10 10 300 200]); %
図示：判別直線
plot(d1(:,1),d1(:,2),'or'); hold on;
plot(d2(:,1),d2(:,2),'*b'); hold on;
xx=linspace(Fmin1,Fmax1,32);
yy=(-A(2)*xx-A(1))/A(3);
plot(xx,yy,'r--');
yy=(-B(2)*xx-B(1))/B(3); hold on;
plot(xx,yy,'b--');
axis([Fmin1,Fmax1,Fmin2,Fmax2]);

%------------- 別解
C=(inv(Sw)*(m1'-m2'))';
C=[-(m1*C'+m2*C')/2 C]; % ３成分の比は
Aと同じ

%--------------- 図示：２次判別の境界
と等高線
L=300;
```

判別分析で特徴で見分ける

1. 線形判別とは？
2. マハラノビス距離
3. 線形判別の手順
4. マハラノビス距離の計算法
5. 主成分分析（PCA）との併用

2種類の距離



水俣病のネコ

番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

健康なネコ

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

ユークリッド距離
→ 異なる

マハラノビス距離
→ 同じ

↔ は●から○までの距離
↔ は●から○までの距離

再掲

手順① データを文字で表現

番号	脳	肝臓
1	9.1	54.5
2	10.4	68.0
3	8.2	53.5
4	7.5	47.6
5	9.7	52.5
6	4.9	45.3

番号	脳	肝臓
1	2.3	31.8
2	0.7	14.5
3	2.5	33.3
4	1.1	33.4
5	3.9	61.2
6	1.0	12.3

$$\boldsymbol{x}_1^{(1)} = [\ 9.1 \quad 54.5\ ]^{\mathrm{T}} \qquad \boldsymbol{x}_1^{(2)} = [\ 2.3 \quad 31.8\ ]^{\mathrm{T}}$$

$$\boldsymbol{x}_2^{(1)} = [10.4 \quad 68.0\ ]^{\mathrm{T}} \qquad \boldsymbol{x}_2^{(2)} = [\ 0.7 \quad 14.5\ ]^{\mathrm{T}}$$

$$\vdots \qquad\qquad \vdots$$

T は転置

$$\boldsymbol{x}_6^{(1)} = [\ 4.9 \quad 45.3\ ]^{\mathrm{T}} \qquad \boldsymbol{x}_6^{(2)} = [\ 1.0 \quad 12.3\ ]^{\mathrm{T}}$$

再掲

手順② クラス毎の平均

$$\overline{\boldsymbol{x}}^{(1)} = [8.30 \quad 53.57]^{\mathrm{T}} \qquad \overline{\boldsymbol{x}}^{(2)} = [1.92 \quad 31.08]^{\mathrm{T}}$$

$$\overline{\boldsymbol{x}}^{(k)} = \frac{1}{N^{(k)}} \sum_{i=1}^{N^{(k)}} \boldsymbol{x}_i^{(k)}, \qquad \text{クラス } k = 1,2$$

$$\boldsymbol{x}_1^{(1)} - \overline{\boldsymbol{x}}^{(1)} = [\,0.80 \quad 0.93\,]^{\mathrm{T}} \qquad \boldsymbol{x}_1^{(2)} - \overline{\boldsymbol{x}}^{(2)} = [\,0.38 \quad 0.72\,]^{\mathrm{T}}$$

$$\begin{matrix} 2.10 & 14.43 \\ -0.10 & -0.07 \\ -0.80 & -5.97 \\ 1.40 & -1.07 \\ -3.40 & -8.27 \end{matrix} \qquad \begin{matrix} -1.22 & -16.58 \\ 0.58 & 2.22 \\ -0.82 & 2.32 \\ 1.98 & 30.12 \\ -0.92 & -18.78 \end{matrix}$$

再掲

手順③ 分散共分散行列

$$\boldsymbol{x}^{(1)} - \overline{\boldsymbol{x}}^{(1)} = \begin{bmatrix} 0.80 & 0.93 \\ 2.10 & 14.43 \\ -0.10 & -0.07 \\ -0.80 & -5.97 \\ 1.40 & -1.07 \\ -3.40 & -8.27 \end{bmatrix}^{\mathrm{T}} \qquad \boldsymbol{x}^{(2)} - \overline{\boldsymbol{x}}^{(2)} = \begin{bmatrix} 0.38 & 0.72 \\ -1.22 & -16.58 \\ 0.58 & 2.22 \\ -0.82 & 2.32 \\ 1.98 & 30.12 \\ -0.92 & -18.78 \end{bmatrix}^{\mathrm{T}}$$

$N^{(1)} = 6 \qquad N^{(2)} = 6$

$$\boldsymbol{S}^{(k)} = \frac{1}{N^{(k)} - 1} \left(\boldsymbol{x}^{(k)} - \overline{\boldsymbol{x}}^{(k)}\right)\left(\boldsymbol{x}^{(k)} - \overline{\boldsymbol{x}}^{(k)}\right)^{\mathrm{T}}, \qquad k = 1,2$$

不偏分散

$$\boldsymbol{S}^{(1)} = \begin{bmatrix} 3.84 & 12.49 \\ 12.49 & 62.85 \end{bmatrix} \qquad \boldsymbol{S}^{(2)} = \begin{bmatrix} 1.48 & 19.36 \\ 19.36 & 309.13 \end{bmatrix}$$

手順④　マハラノビス距離

$\boldsymbol{x}$ と $\overline{\boldsymbol{x}}^{(k)}$ の距離の自乗は

$$D_{(k)}^2(\boldsymbol{x}) = (\boldsymbol{x} - \overline{\boldsymbol{x}}^{(k)})^{\mathrm{T}} (\boldsymbol{S}^{(k)})^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}^{(k)})$$

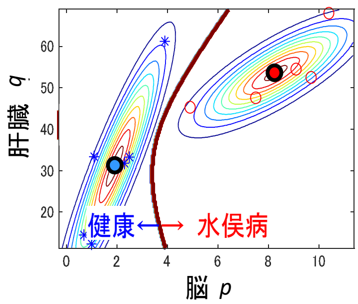
k=1,2

(p, q) を (脳，肝臓) のデータとすると

$$D_{(1)}^2(\boldsymbol{x}) = [p - 8.30 \quad q - 53.57] \begin{bmatrix} 3.84 & 12.49 \\ 12.49 & 62.85 \end{bmatrix}^{-1} \begin{bmatrix} p - 8.30 \\ q - 53.57 \end{bmatrix}$$

$$D_{(2)}^2(\boldsymbol{x}) = [p - 1.92 \quad q - 31.08] \begin{bmatrix} 1.48 & 19.36 \\ 19.36 & 309.13 \end{bmatrix}^{-1} \begin{bmatrix} p - 1.92 \\ q - 31.08 \end{bmatrix}$$

計算の結果



$D_{(1)}$: ● からの距離

$D_{(2)}$: ● からの距離

	番号	脳	肝臓	マハラノビス距離 $D_{(1)}^2$	マハラノビス距離 $D_{(2)}^2$
水俣病のネコ	1	9.1	54.5	0.29	123.20
	2	10.4	68.0	3.75	145.91
	3	8.2	53.5	0.01	90.09
	4	7.5	47.6	0.68	77.77
	5	9.7	52.5	1.92	155.68
	6	4.9	45.3	3.35	16.93
健康なネコ	1	2.3	31.8	9.60	0.43
	2	0.7	14.5	24.30	1.01
	3	2.5	33.3	8.84	0.75
	4	1.1	33.4	13.95	3.45
	5	3.9	61.2	26.63	2.97
	6	1.0	12.3	27.69	1.39

大小で判別する

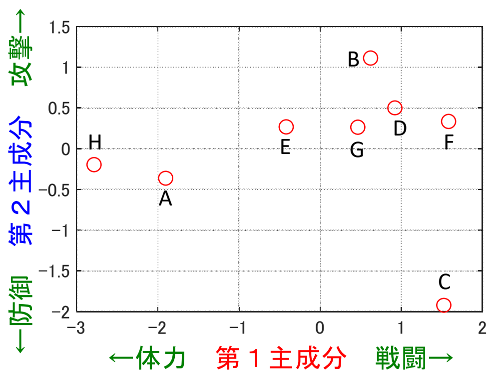
判別分析で
特徴で見分ける

1. 線形判別とは？
2. マハラノビス距離
3. 線形判別の手順
4. マハラノビス距離の計算法
5. 主成分分析（PCA）との併用

PCAで特徴を集約

グループ1		体力	攻撃	防御
A	ハピナス	496	129	169
H	ラッキー	487	60	128
E	カビゴン	330	190	169

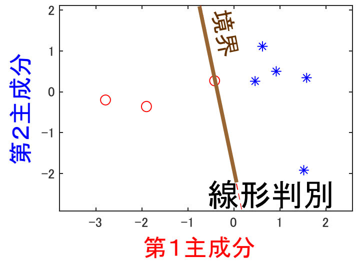
グループ2		体力	攻撃	防御
B	ケッキング	284	290	166
C	ルギア	235	193	310
D	ガブリアス	239	261	193
F	パルキア	189	280	215
G	メルメタル	264	226	190

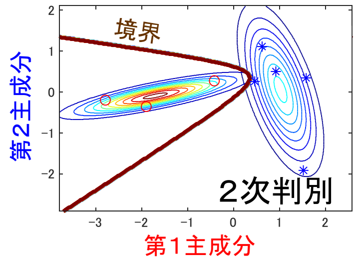


３種類の特徴を　PCA　２種類に集約する

少ない特徴で境界を決める

	第１主成分	第２主成分
グループ1		
A. ハピナス	−1.90	−0.36
H. ラッキー	−2.79	−0.20
E. カビゴン	−0.42	0.27
グループ2		
B. ケッキング	0.62	1.11
C. ルギア	1.52	−1.92
D. ガブリアス	0.92	0.50
F. パルキア	1.58	0.34
G. メルメタル	0.46	0.26





判別理由を 分析する

グループ1	
A	ハピナス
H	ラッキー
E	カビゴン

グループ2	
B	ケッキング
C	ルギア
D	ガブリアス
F	パルキア
G	メルメタル

主成分分析

第１主成分（p）
戦闘的（体力は弱い）

第２主成分（q）
攻撃的（防御は弱い）

線形判別分析

グループの境界

$$q = -(a_1 p + a_0)/a_2$$
$$= -5.61\,p - 1.93$$

理由を分析

戦闘性　攻撃性

$$5.61\,p + q = -1.93$$

和に意味がある

戦闘性を重視

参考

線形判別の固有値問題

$$maximize \ \frac{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u}}$$

クラス間分散 → 大きく

クラス内分散 → 小さく

$$maximize \ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u} \quad s.t. \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u} = 1$$

$$J(\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{u} + \lambda\left(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{u} - 1\right)$$

$$\frac{\partial J(\boldsymbol{u})}{\partial(\boldsymbol{u})} = 2\,\boldsymbol{S}_b\boldsymbol{u} - 2\,\lambda\boldsymbol{S}_w\boldsymbol{u} = 0$$

$$\boxed{\boldsymbol{S}_b\boldsymbol{U} = \boldsymbol{S}_w\boldsymbol{U}\boldsymbol{L}}$$

原田達也「画像認識」講談社

参考

主成分分析の固有値問題

$$minimize \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}_{xx}\boldsymbol{u} \quad s.t. \quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}_{xx}\boldsymbol{u} = 1$$

$$J(\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}_{xx}\boldsymbol{u} - \lambda(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} - 1)$$

$$\frac{\partial J(\boldsymbol{u})}{\partial \boldsymbol{u}} = 2\boldsymbol{C}_{xx}\boldsymbol{u} - 2\lambda\boldsymbol{u} = 0$$

$$\boldsymbol{C}_{xx}\boldsymbol{u} = \lambda\boldsymbol{u}$$

$$\boxed{\boldsymbol{C}_{xx}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{L}} \qquad \begin{cases} \boldsymbol{L} = \mathrm{diag}(\lambda_1, \cdots, \lambda_m) \\ \boldsymbol{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m) \end{cases}$$

原田達也「画像認識」講談社

判別分析
【演習問題】

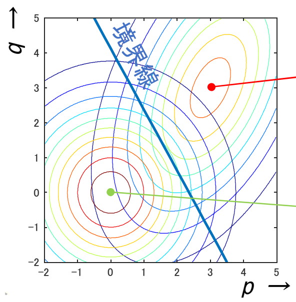
【問題】 境界線の方程式を求めよ

但し，以下とする。

$\overline{\boldsymbol{x}}^{(1)} = [p_1 \quad q_1]$

$\overline{\boldsymbol{x}}^{(2)} = [0 \quad 0]$

$$\boldsymbol{S}_w \approx \frac{\boldsymbol{S}^{(1)} + \boldsymbol{S}^{(2)}}{2} = \begin{bmatrix} s_1 & r \\ r & s_2 \end{bmatrix}$$



$\overline{\boldsymbol{x}}^{(1)} = [3 \quad 3]$

$\boldsymbol{S}^{(1)} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

$\overline{\boldsymbol{x}}^{(2)} = [0 \quad 0]$

$\boldsymbol{S}^{(2)} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$

場合１　$p_1 = q_1$ のとき
場合２　$p_1 = q_1$，$s_1 = s_2$ のとき
場合３　$p_1 = q_1$，$s_1 = s_2$,$r = 0$のとき

境界線の傾き

境界線の方程式は $q = -\frac{a_1}{a_2}p - \frac{a_0}{a_2}$

$$\left.\begin{aligned} \overline{\boldsymbol{x}}^{(1)} &= [p_1 \quad q_1]^{\mathrm{T}} \\ \overline{\boldsymbol{x}}^{(2)} &= [0 \quad 0]^{\mathrm{T}} \\ \boldsymbol{S}_w &= \begin{bmatrix} s_1 & r \\ r & s_2 \end{bmatrix} \end{aligned}\right\} \quad \boldsymbol{u} = \boldsymbol{S}_w^{-1}\left(\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}\right)$$

$$\boldsymbol{u} = \frac{1}{s_1 s_2 - r^2}\begin{bmatrix} s_2 p_1 - r q_1 \\ -r p_1 + s_1 q_1 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

以上より，傾きは $-\frac{a_1}{a_2} = -\frac{s_2 p_1 - r q_1}{s_1 q_1 - r p_1}$

境界線の縦軸との交点

$$a_0 = \frac{-[a_1 \quad a_2]\cdot\left(\overline{\boldsymbol{x}}^{(1)} + \overline{\boldsymbol{x}}^{(2)}\right)}{2} = -\frac{a_1 p_1 + a_2 q_1}{2}$$

$$\begin{aligned} a_0 &= -\frac{(s_2 p_1 - r q_1)p_1 + (s_1 q_1 - r p_1)q_1}{2(s_1 s_2 - r^2)} \\ &= -\frac{s_2 p_1^2 + s_1 q_1^2 - 2 r p_1 q_1}{2(s_1 s_2 - r^2)} \end{aligned}$$

$$-\frac{a_0}{a_2} = \frac{s_2 p_1^2 + s_1 q_1^2 - 2 r p_1 q_1}{2(s_1 q_1 - r p_1)}$$

境界線の方程式

$$q = -\frac{s_2 p_1 - r q_1}{s_1 q_1 - r p_1} p + \frac{s_2 p_1^2 + s_1 q_1^2 - 2 r p_1 q_1}{2(s_1 q_1 - r p_1)}$$

場合１ $p_1 = q_1$ のとき

$$q = -\frac{s_2 - r}{s_1 - r} p + \frac{s_2 + s_1 - 2r}{2(s_1 - r)} p_1$$

場合２ $p_1 = q_1,\ s_1 = s_2$ のとき

$$q = -\frac{s_2}{s_1} p + \frac{s_2 + s_1}{2 s_1} p_1$$

場合３ $p_1 = q_1,\ s_1 = s_2\ , r = 0$のとき

$$q = -p + p_1$$