尤度比に基づくベイズ分類

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

多変量解析☆演習　L07(2026-11-05 Thu)

最終更新: Time-stamp: "2026-11-05 Thu 18:57 JST hig"

今日の目標

- 尤度比に基づくベイズ分類を説明できる
- 混合ガウス分布の場合に尤度比に基づくベイズ分類を適用して判別境界を決定できる

ここまで来たよ

7. 尤度比に基づくベイズ分類
   - 復習
   - 尤度比に基づくベイズ分類
   - 2 次元の混合ガウス分布の尤度比に基づくベイズ分類

ROC=Curve=Receiver Operating Characteristic

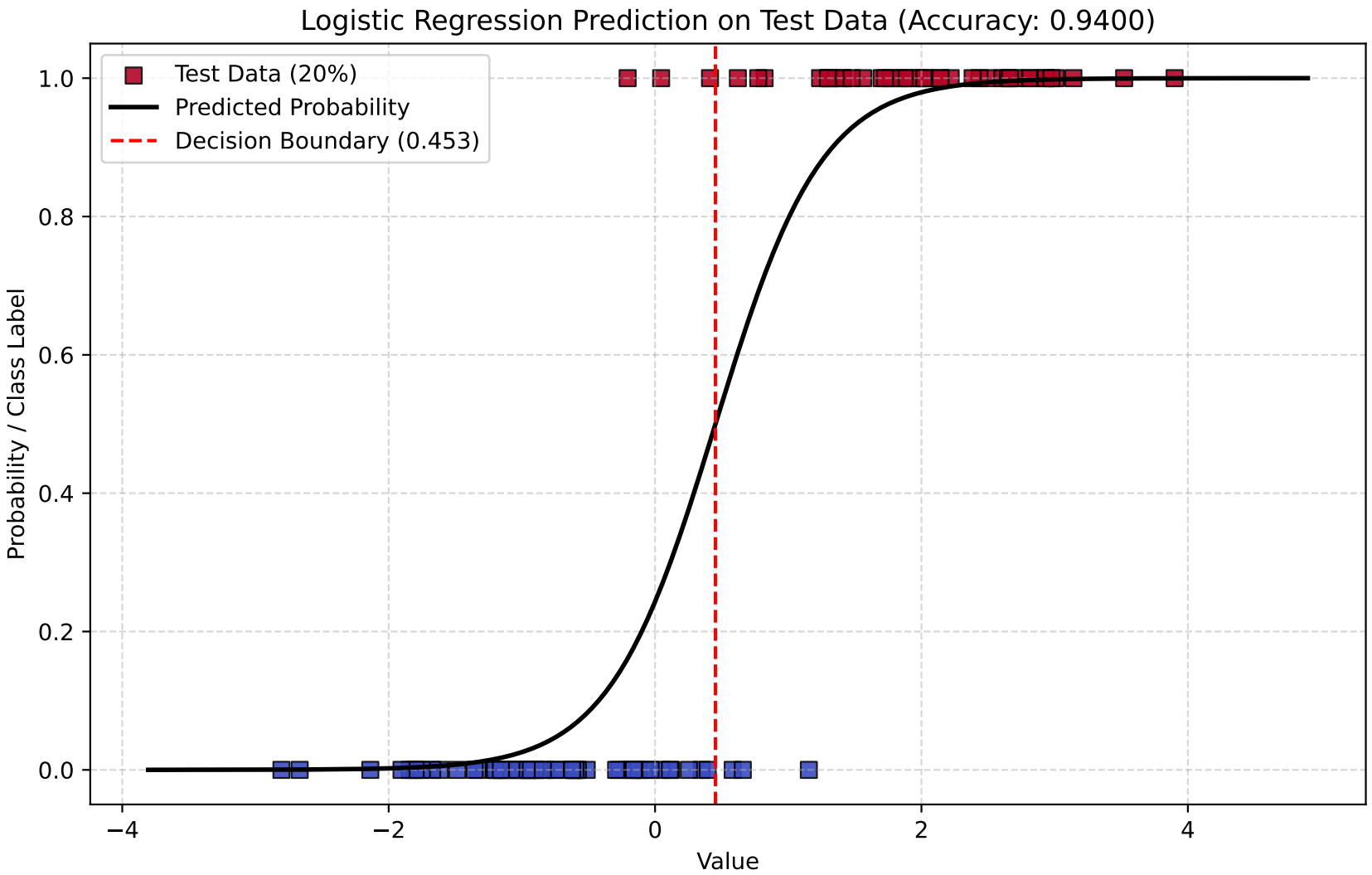
- 縦軸 recall= 真陽性率 $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
- 横軸 1−specificity= 偽陽性率 $= \frac{\mathrm{FP}}{\mathrm{TN+FP}}$
- 左上ほどよい

閾値 c を動かした時の振る舞い全体を表現するのが ROC という図.
c をパラメタとして動かして得られる曲線

- $c = 1$ で右上 (真陽性, 偽陽性)= $(0, 0)$
- $c = 0$ で 左下 (真陽性, 偽陽性)= $(1, 1)$ を通る

途中で，左上に再接近するところの c を採用したい.

ロジスティック回帰


Logistic Regression Prediction on Test Data (Accuracy: 0.9400)
Test Data (20%)
Predicted Probability
Decision Boundary (0.453)
Probability / Class Label
Value
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4

ロジスティック重回帰モデルの判別境界

定義 (ロジスティック単回帰モデル)

$$P(Y=1) = S(\boldsymbol{\beta}^\top \boldsymbol{x}) = \frac{1}{1+\mathrm{e}^{-\boldsymbol{\beta}^\top \boldsymbol{x}}}, \quad \text{つまり} \quad Y \sim Bin(1, S(\boldsymbol{\beta}^\top \boldsymbol{x}))$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} 1 \\ x_1 \end{pmatrix}, \quad \boldsymbol{\beta}^\top \boldsymbol{x} = \beta_0 + \beta_1 \cdot x_1$$

$P(Y=1) \geq c$ (閾値) の時に，$\hat{y}=1$ と予想するなら，$\hat{y}=0,1$ の判別境界は
$S(\boldsymbol{\beta}^\top \boldsymbol{x}) =$ 定数.
説明変数の個数 p によって，

$$\boldsymbol{\beta}^\top \boldsymbol{x} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = \text{定数}$$

1. $p=1$: x_1 直線上の 1 点
2. $p=2$: (x_1, x_2) 平面上の直線
3. $p=3$: (x_1, x_2, x_3) 空間内の平面

識別モデルと生成モデル

説明変数 x から，Y の確率が決まるモデルでは，いったん X を確率変数に格上げして，この確率を，条件付き確率 $P(Y=y|X=x)=f_{\mathrm{Y|X}}(y|x)$ と考える．

分類の識別モデル 例 ロジスティック回帰．条件付き確率 $P(Y=y|X=x)$ だけを考え，x の周辺分布 $P(X=x)$ は仮定しない．x はドーナツ製造機のダイヤル．

分類の生成モデル 例 今からやる，尤度比に基づくベイズ分類．$P(Y=y|X=x)$ に加え，x の分布 $P(X=x)$ を考える．したがって，同時分布 $P(X=x,Y=y)=P(Y=y|X=x)P(X=x)$ を考える．

> 何もないところから，事前分布にしたがって x 続いて条件付き分布にしたがって y を生み出せるから，この名前

ここまで来たよ

7. 尤度比に基づくベイズ分類
   - 復習
   - 尤度比に基づくベイズ分類
   - 2 次元の混合ガウス分布の尤度比に基づくベイズ分類

尤度比に基づくベイズ分類の手順

X, Y はともに確率変数で，同時分布 $f(x, y)$ があると思っている．母数はすべて既知とする．
x が 1 個与えられた時に $y = 0, 1$ を予測したい．そのあと評価．

ベイズ分類

与えられるもの

- 既知の，または想定した事前分布（周辺分布）$f_{\mathrm{Y}}(y)$.
- 条件付き確率 $f_{\mathrm{X}|\mathrm{Y}}(x|y)$. $y = 0, 1$ それぞれの x の分布．x が 1 個与えられれば尤度．

入力

- 説明変数 x

出力

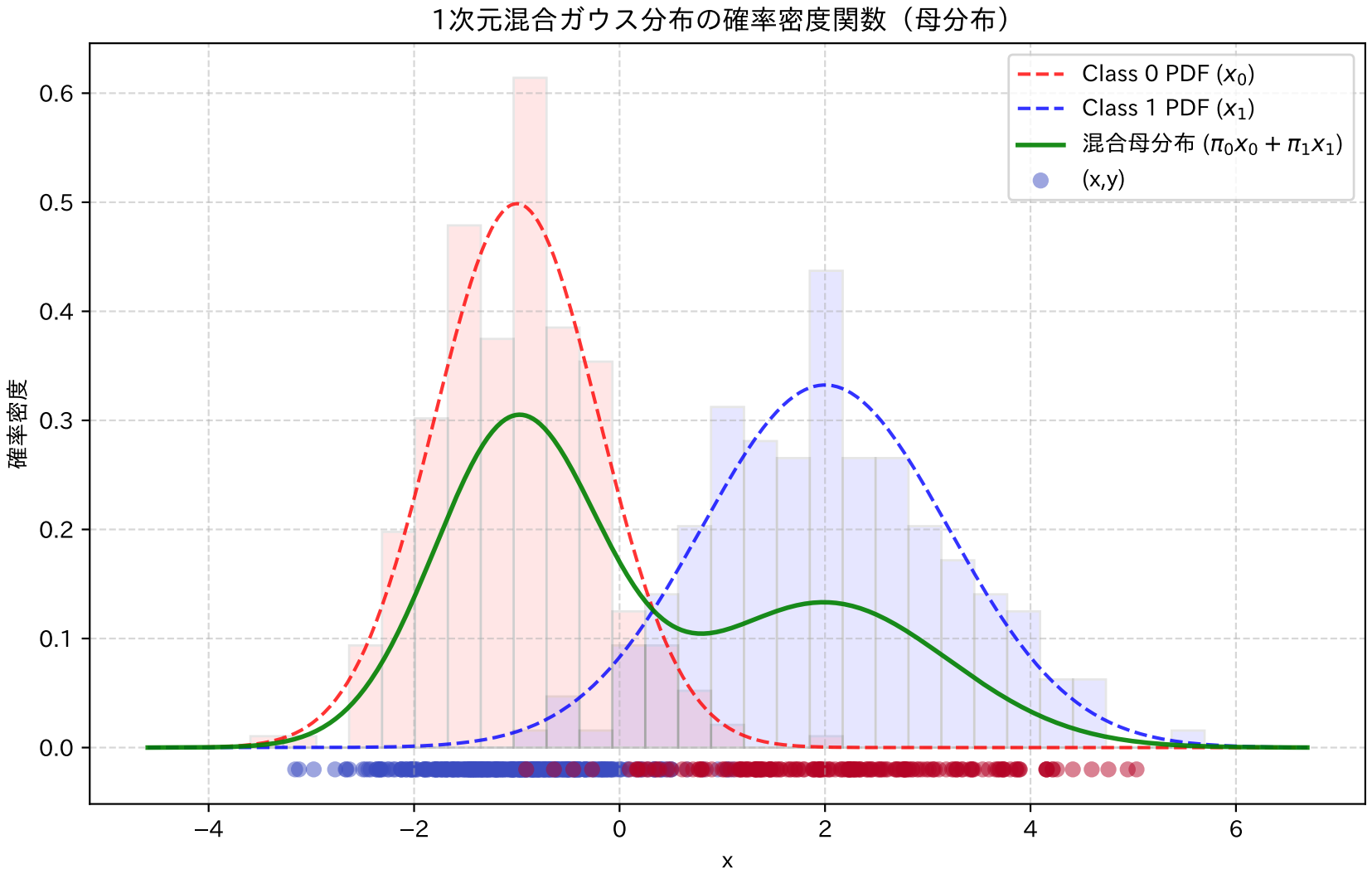
- 事後分布（条件付き分布）

$$f_{\mathrm{Y}|\mathrm{X}}(y|x) = \frac{f_{\mathrm{X}|\mathrm{Y}}(x|y) f_{\mathrm{Y}}(y)}{\sum_y f_{\mathrm{X}|\mathrm{Y}}(x|y) f_{\mathrm{Y}}(y)}$$

判定

- $f_{\mathrm{Y}|\mathrm{X}}(1|x) \geq \frac{1}{2}$ ならば $\hat{y} = 1$.
- または，$\geq c$ (閾値）で判定してもよい．

混合ガウス分布のベイズ分類



1 次元混合ガウス分布の場合のベイズ分類

$f_{\mathrm{X}}(x)$ が混合ガウス分布になる場合, 同時分布は $f(x,y) = \pi_y \cdot f(x|\mu_y, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \mathrm{e}^{-\frac{1}{2\sigma_y^2}(x-\mu_y)^2}$.

ベイズ推定

与えられるもの

- 既知の，または想定した事前分布（周辺分布）$f_{\mathrm{Y}}(y) = \pi_y$.
- 条件付き確率 $f_{\mathrm{X}|\mathrm{Y}}(x|y,\mu,\sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \mathrm{e}^{-\frac{1}{2\sigma_y^2}(x-\mu_y)^2}$ $y = 0, 1$ それぞれの x の分布．x が 1 個与えられれば尤度．π_y, μ_y, σ_y^2 も与えられる．

入力

説明変数 x

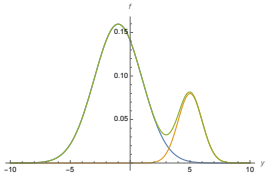
出力

事後分布（条件付き分布）

$$f_{\mathrm{Y}|\mathrm{X}}(y|x) = \frac{\pi_y f(x|y,\mu_y,\sigma_y^2)}{\pi_0 f(x|y,\mu_0,\sigma_0^2) + \pi_1 f(x|y,\mu_1,\sigma_1^2)}$$

判定

- $f_{\mathrm{Y}|\mathrm{X}}(1|x) \geq \frac{1}{2}$ ならば $\hat{y} = 1$.
- または，$\geq c$ (閾値）で判定してもよい．



与えられなかった時の母数，事前確率の推定

決まっているかのように説明した μ_y, σ_y^2, π_y を，訓練データ $(x, y)_{i=1,\ldots,n}$ からどう決めるか？

- 訓練データから周辺分布 $\pi_y = f_{\mathrm{Y}}(y)$ を決める．
  - 比率の推定．$\hat{\pi}_1 = (y = 1\text{の個数})/n$.
- 訓練データから 母数 μ_y, σ_y^2 を，y ごとに推定する．
  - y ごとの標本平均値，不偏標本分散で．

ベイズ統計学では，母数 π_y, μ_y, σ_y^2 の分布（！）も考えるのだが，ここではそう考えていない．(X, Y) に対してベイズの定理を適用しているだけ

L07-Q1

Quiz(混合ガウス分布)

周辺分布 $f_{\mathrm{X}}(x)$ が混合ガウス分布 $(\pi_{\mathrm{Y}}, \mu_{\mathrm{Y}}, \sigma_{\mathrm{Y}})$ になる同時分布 $f(x, y)$ $(y = 0, 1)$ を考える.
パラメタを $(\pi_0 = 3/10, \pi_1 = 7/10, \mu_0 = 2, \mu_1 = 6, \sigma_0 = 2, \sigma_1 = 1/2)$ とする.
$X = 5$ という条件のもとで $Y = 0$ である条件付き確率 $P(Y = 0|X = 5)$ を，標準正規分布の確率密度関数 $\phi(\cdot)$ で表そう.

ここまで来たよ

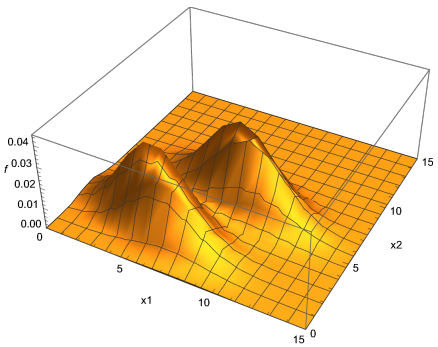
7. 尤度比に基づくベイズ分類
   - 復習
   - 尤度比に基づくベイズ分類
   - 2 次元の混合ガウス分布の尤度比に基づくベイズ分類

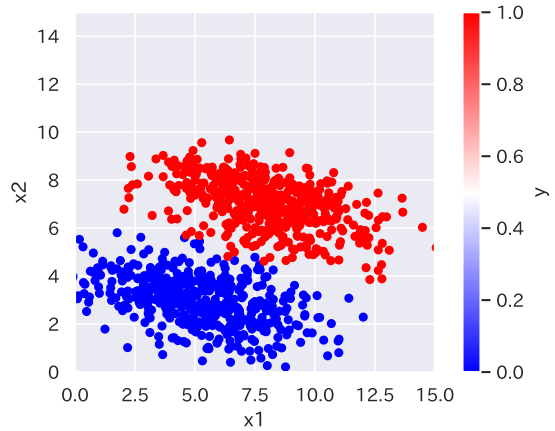
2 次元の混合ガウス分布

2 次元の混合ガウス分布

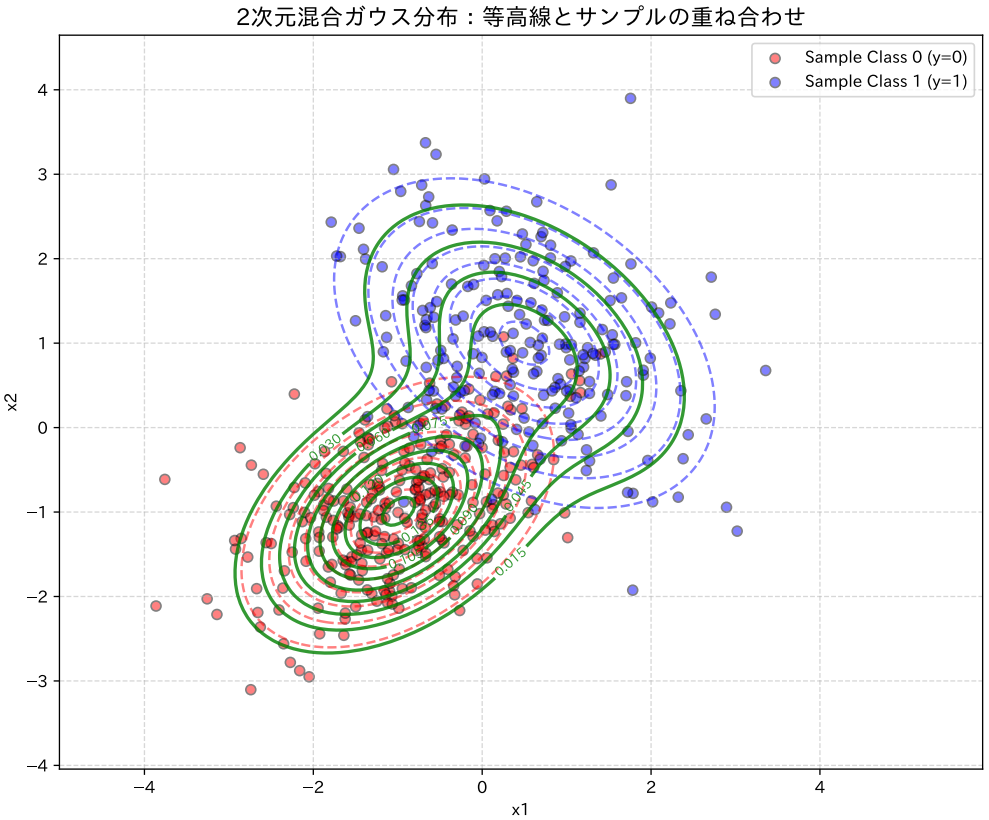
x が 2 次元ベクトル $\boldsymbol{x} = (x_1, x_2)^\top$ になっただけ.

$$f(\boldsymbol{x}, y) = \begin{cases} \dfrac{1}{(2\pi)^{2/2}(\det \Sigma_0)^{1/2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^\top \Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)} \cdot \pi_0 & (y=0) \\ \dfrac{1}{(2\pi)^{2/2}(\det \Sigma_1)^{1/2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)} \cdot \pi_1 & (y=1) \\ 0 & (y\text{が他}) \end{cases}$$





2 次元混合ガウス分布のベイズ分類



2 次元混合ガウス分布の時の判別境界

$$\frac{1}{f_{\mathrm{X}}(x)}\pi_1\frac{1}{\sqrt{(2\pi)^2\det\Sigma_1}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)} \geq c$$

$$\frac{1}{f_{\mathrm{X}}(x)}\pi_0\frac{1}{\sqrt{(2\pi)^2\det\Sigma_1}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^\top\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)} \leq 1-c$$

各辺の比をとって左辺を尤度の比にする

$$\frac{\sqrt{\det\Sigma_0}}{\sqrt{\det\Sigma_1}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)}\mathrm{e}^{+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^\top\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)} \geq \frac{c}{1-c}\frac{\pi_0}{\pi_1}$$

両辺の対数をとって左辺を数尤度の差にする

$$\log\det\Sigma_0 - \log\det\Sigma_1 - (\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + (\boldsymbol{x}-\boldsymbol{\mu}_0)^\top\Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0) \geq 2\log\frac{c}{1-c}\frac{\pi_0}{\pi_1}$$

等号が，判別境界の式（2 次曲線の式）.

$\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$ の間のマハラノビスの距離

$$D_\Sigma(\boldsymbol{x}, \boldsymbol{x}') = ((\boldsymbol{x} - \boldsymbol{x}')^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{x}'))^{1/2}$$

距離空間, 位相

簡単化：QDA(2 次判別分析)

一部の項を無視して，

$$-(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top \Sigma_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)+(\boldsymbol{x}-\boldsymbol{\mu}_0)^\top \Sigma_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\geq 0$$

で考えると，マハラノビスの距離の大小で分類できる．

$$D_{\Sigma_1}(\boldsymbol{x},\boldsymbol{\mu}_1)\leq D_{\Sigma_0}(\boldsymbol{x},\boldsymbol{\mu}_0)$$

ならば $\hat{y}=1$.
判別境界は（尤度に基づくベイズ分類からずれた）2 次曲線．

超簡単化：LDA(線形判別分析)

混合ガウス分布のとき，QDA 同様，一部の項を無視して，かつ，
$\Sigma_0 = \Sigma_1$ と仮定.
判別境界は，$\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1$ に「直交」する直線.

Fisher の線形判別関数 ($z < 0$ なら $y = 1$ と答える)

$$z = \frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_0}{\sigma}\right)^2\right] \overset{\text{整理}}{=} \frac{\mu_0 - \mu_1}{\sigma} \cdot \frac{x - \frac{\mu_0+\mu_1}{2}}{\sigma}$$