

ベイズ判別 非線形ロジスティック判別

照屋 佑喜仁

April 10, 2025

① 非線形ロジスティック判別

非線形ロジスティック判別

- 前回，線形ロジスティック判別法により2群判別を行ったが，両群の構造が複雑だと線形判別法でうまく判別できないことがある．
- **非線形関数**による判別法の構成が必要
- 以前学んだロジスティック回帰モデルの非線形化と同様にして線形ロジスティック判別の非線形化を行う．

非線形ロジスティック判別

Review

G_1 , G_2 いずれかの群から p 次元データ \mathbf{x} が観測されたとする. このデータが群 G_1 に属すると仮定した場合の事後確率 $P(G_1|\mathbf{x})$ と, 群 G_2 に属すると仮定した場合の事後確率 $P(G_2|\mathbf{x})$ との比の対数を変数の線形結合で表されると仮定したのが線形ロジスティック判別であった. (前回)

$$\log \frac{P(G_1|\mathbf{x})}{P(G_2|\mathbf{x})} = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}$$

非線形ロジスティック判別

非線形ロジスティック判別では、変数の線形結合ではなく変数の非線形関数で表すことによって行う。

この非線形関数を基底関数 $b_0(\mathbf{x}) \equiv 1, b_1(\mathbf{x}), \dots, b_m(\mathbf{x})$ の線形結合で表す。

$$\log \frac{P(G_1|\mathbf{x})}{P(G_2|\mathbf{x})} = \sum_{j=0}^m w_j b_j(\mathbf{x}) = \mathbf{w}^T \mathbf{b}(\mathbf{x}) \quad (7.38)$$

ただし $\mathbf{w} = (w_0, w_1, \dots, w_m)^T$ はパラメータベクトル (重みベクトル), $\mathbf{b}(\mathbf{x}) = (b_0(\mathbf{x}), \dots, b_m(\mathbf{x}))^T$ は基底関数ベクトルとする。

非線形ロジスティック判別

2 群の事後確率について, $P(G_2|\mathbf{x}) = 1 - P(G_1|\mathbf{x})$ なので, (7.38) 式は

$$\begin{aligned}\log \frac{P(G_1|\mathbf{x})}{1 - P(G_1|\mathbf{x})} &= \mathbf{w}^T \mathbf{b}(\mathbf{x}) \\ \therefore P(G_1|\mathbf{x}) &= \frac{\exp \{ \mathbf{w}^T \mathbf{b}(\mathbf{x}) \}}{1 + \exp \{ \mathbf{w}^T \mathbf{b}(\mathbf{x}) \}}\end{aligned}$$

この事後確率を反応確率と考えると, 以前学習した非線形ロジスティックモデルに対応していることがわかる.

Review

ロジスティックモデルとは一般に, 任意の実数値 x (刺激レベル) に対して $(0, 1)$ の数値 y (反応率) を出力する関数を出力するようなものであった. さらに, 反応したかどうかを表す確率変数 Y を導入することで, 要因 x と反応確率 $\pi (= P(Y = 1|x))$ をロジスティックモデルでモデル化するのであった.

非線形ロジスティック判別

非線形ロジスティック回帰では多項式モデルや B-スプラインを用いることができるのであった。今回は具体的な基底関数としてガウス型基底関数に基づく非線形ロジスティック判別を試みる。

Review (2 次元ガウス関数)

$$\begin{aligned} b_j(\mathbf{x}) &= \phi_j(x_1, x_2) \\ &= \exp \left\{ -\frac{(x_1 - \mu_{j1}^2) + (x_2 - \mu_{j2}^2)}{2h_j^2} \right\}, \quad j = 1, 2, \dots, m \end{aligned}$$

m は基底関数の個数, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ は基底関数の位置を定める中心ベクトル, h_j^2 は関数の広がりを表す量

非線形ロジスティック判別

図の2群を判別するような判別関数 $h(x_1, x_2)$ は、最尤法によって推定した係数ベクトルを重みとする次の式で表す。

$$h(x_1, x_2) = \hat{w}_0 + \sum_{j=1}^m \hat{w}_j \phi_j(x_1, x_2) = 0$$
$$\therefore \log \frac{P(G_1|\mathbf{x})}{1 - P(G_1|\mathbf{x})} = \hat{\mathbf{w}}\mathbf{b}(\mathbf{x}) \begin{cases} \leq 0 \Rightarrow G_1 \\ > 0 \Rightarrow G_2 \end{cases}$$

非線形ロジスティック判別

- 非線形ロジスティック判別法は文字認識や音声認識，画像認識などパターン認識の分析法として有効.
- 高次元データに対して最尤法がモデルの推定に有効に機能しない場合がある.
- そのような場合は正則化最尤法が有効. 適切に正則化パラメータを選び非線形判別関数を構成.

多群の非線形判別

- 3群以上の複数の群から観測された学習データから非線形判別法を構成する.
- L 個の群 G_1, G_2, \dots, G_L があるとする.
- これらから, n 個の p 次元データ $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ が学習データとして観測されたとする.
- 新たなデータ \mathbf{x} が L 個の群のいずれかから観測されたとき, それが第 k 群からのものであるという事後確率を次で定める.

$$P(G_k|\mathbf{x}), K = 1, 2, \dots, L$$

多群の非線形判別

- 2 群の場合を拡張して非線形判別法を構成する.
- 群 G_k ($k = 1, 2, \dots, L-1$) と第 L 番目の群 G_L との間の $L-1$ 個の事後確率の比を考えると

$$\log \left\{ \frac{P(G_k|\mathbf{x})}{P(G_L|\mathbf{x})} \right\} = \sum_{j=0}^m w_{kj} b_j(\mathbf{x}) = \mathbf{w}_k^T \mathbf{b}(\mathbf{x}), \quad k = 1, 2, \dots, L-1$$

$$\text{ただし } \mathbf{w}_k = (w_{k0}, w_{k1}, \dots, w_{km})^T$$

変形すると

$$P(G_k|\mathbf{x}) = \exp \{ \mathbf{w}_k^T \mathbf{b}(\mathbf{x}) \} P(G_L|\mathbf{x}) \quad (\text{A})$$

多群の非線形判別

ここで,

$$P(G_L|\mathbf{x}) = 1 - \sum_{g=1}^{L-1} P(G_g|\mathbf{x}) = 1 - \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}) \} P(G_L|\mathbf{x})$$

$$\therefore P(G_L|\mathbf{x}) = \frac{1}{1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}) \}}$$

A に代入して

$$P(G_k|\mathbf{x}) = \frac{\exp \{ \mathbf{w}_k^T \mathbf{b}(\mathbf{x}) \}}{1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}) \}}$$

多群の非線形判別

さらに、各データ \mathbf{x}_i に対して群の所属を表すラベル変数ベクトル \mathbf{y}_i を次のように決める.

$$\begin{aligned}\mathbf{y}_i &= (y_{i1}, y_{i2}, \dots, y_{i(L-1)})^T \\ &= \begin{cases} (0, \dots, 0, \overset{k}{1}, 0, \dots, 0) & \text{if } \mathbf{x}_i \in G_k, \ k = 1, \dots, L-1 \\ (0, \dots, 0) & \text{if } \mathbf{x}_i \in G_L \end{cases}\end{aligned}$$

i 番目のデータが第 k 群 ($k \neq L$) から観測されたなら第 k 成分のみ 1 でその他の成分は 0 の $(L-1)$ 次元ベクトル, 第 L 群からならすべての成分が 0 の $(L-1)$ 次元ベクトルに対応させる.

多群の非線形判別

2スライド前の結果から、事後確率は基底関数の係数ベクトル（重みベクトル）に依存する。

学習するデータから推定するパラメータを $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{L-1}\}$ として $P(G_k|\mathbf{x}_i) = \pi_k(\mathbf{x}_i; \mathbf{w})$ と表す。

2群のラベル変数はベルヌーイ分布に従ったのであった。

多群の $(L-1)$ 次元ラベル変数ベクトル \mathbf{y}_i は多項分布に従う。

$$f(\mathbf{y}_i|\mathbf{x}_i; \mathbf{w}) = \prod_{k=1}^{L-1} \pi_k(\mathbf{x}_i; \mathbf{w})^{y_{ik}} \pi_L(\mathbf{x}_i; \mathbf{w})^{1 - \sum_{g=1}^{L-1} y_{ig}}$$

多群の非線形判別

例えば $x_i \in G_j$, $j \neq L$ のとき, $\mathbf{y}_i = (0, \dots, 0, \overset{j}{1}, 0, \dots, 0)$
 よって

$$f(\mathbf{y}_i | \mathbf{x}_i; \mathbf{w}) = \pi_j(\mathbf{x}_i; \mathbf{w})$$

$1 - \sum_{g=1}^{L-1} y_{ig}$ が 0 になるのが効いている.

$x_i \in G_L$ なら

$$f(\mathbf{y}_i | \mathbf{x}_i; \mathbf{w}) = \pi_L(\mathbf{x}_i; \mathbf{w})$$

y_{ik} が 0 になるのが効いている.

確かに \mathbf{y}_i は前スライドで示した多項分布に従う.

多群の非線形判別

対数尤度関数は

$$\begin{aligned}
 \ell(\mathbf{w}) &= \log \left[\prod_{i=2}^n \prod_{k=1}^{L-1} \pi_k(\mathbf{x}_i; \mathbf{w})^{y_{ik}} \pi_L(\mathbf{x}_i; \mathbf{w})^{1 - \sum_{g=1}^{L-1} y_{ig}} \right] \\
 &= \sum_{i=1}^n \log \left[\prod_{k=1}^{L-1} \pi_k(\mathbf{x}_i; \mathbf{w})^{y_{ik}} \pi_L(\mathbf{x}_i; \mathbf{w})^{1 - \sum_{g=1}^{L-1} y_{ig}} \right] \\
 &= \sum_{i=1}^n \left[\sum_{k=1}^{L-1} y_{ik} \log \pi_k(\mathbf{x}_i; \mathbf{w}) + \left(1 - \sum_{g=1}^{L-1} y_{ig}\right) \log \pi_L(\mathbf{x}_i; \mathbf{w}) \right]
 \end{aligned}$$

(後半 k は効いていない)

多群の非線形判別

ただし,

$$\pi_k(\mathbf{x}_i; \mathbf{w}) = \frac{\exp \{ \mathbf{w}_k^T \mathbf{b}(\mathbf{x}_i) \}}{1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \}}, \quad k = 1, 2, \dots, L-1$$

$$\pi_L(\mathbf{x}_i; \mathbf{w}) = \frac{1}{1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \}}$$

多群の非線形判別

$\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{L-1}\}$ の最尤推定値は対数尤度関数の最大化によって得られるが、解析的に陽に求めることは難しいため、ニュートン・ラフソン法などの数値最適化法を用いて推定する。

$$\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}} = \left(\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_1}, \dots, \frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_{L-1}} \right)^T = \mathbf{0}$$

多群の非線形判別

(偏微分を計算するために次の式を考える) ここで,

$$\begin{aligned}\log \pi_k(\mathbf{x}_i; \mathbf{w}) &= \mathbf{w}_k^T \mathbf{b}(\mathbf{x}_i) - \log\left(1 + \sum_{g=1}^{L-1} \exp\{\mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i)\}\right) \\ \frac{\partial}{\partial \mathbf{w}_p} \log \pi_k(\mathbf{x}_i; \mathbf{w}) &= \frac{\partial}{\partial \mathbf{w}_p} \mathbf{w}_k^T \mathbf{b}(\mathbf{x}_i) - \frac{1}{1 + \sum_{g=1}^{L-1} \exp\{\mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i)\}} \\ &\quad \frac{\partial}{\partial \mathbf{w}_p} \left\{ 1 + \sum_{g=1}^{L-1} \exp\{\mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i)\} \right\}\end{aligned}\tag{B}$$

多群の非線形判別

右辺第1項について

$$\frac{\partial}{\partial \mathbf{w}_p} \mathbf{w}_k^T \mathbf{b}(\mathbf{x}_i) = \begin{cases} \mathbf{b}(\mathbf{x}_i) & \text{if } k = p \\ \mathbf{0} & \text{if } k \neq p \end{cases}$$

右辺第2項について, もし $g \neq p$ なら

$$\frac{\partial}{\partial \mathbf{w}_p} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} = 0$$

$g = p$ なら

$$\frac{\partial}{\partial \mathbf{w}_p} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} = \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i)$$

多群の非線形判別

よって

$$(B) = \begin{cases} \{1 - \pi_k(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i) & \text{if } k = q \\ -\pi_k(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i) & \text{if } k \neq q \end{cases}$$

前スライド右辺第2項の議論と同様にすれば

$$\frac{\partial}{\partial \mathbf{w}_p} \log \pi_L(\mathbf{x}_i; \mathbf{w}) = -\pi_p(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i)$$

$\ell(\mathbf{w})$ を \mathbf{w}_p で偏微分した時の [] の中を計算する. i を固定して

$$\begin{aligned} & y_{ip} \mathbf{b}(\mathbf{x}_i) + \sum_{k=1}^{L-1} y_{ik} \{-\pi_k(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i) + (1 - \sum_{g=1}^{L-1} y_{ig}) \{-\pi_k(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i)\} \\ &= \{y_{ip} - \pi_k(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i) \end{aligned}$$

多群の非線形判別

以上より

$$\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_p} = \sum_{i=1}^n \{y_{ip} - \pi_p(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i)$$

(教科書では G_k と総和の添字が競合している？大変でした)
ニュートン・ラフソン法など数値的最適化を用いるために必要な2階微分も計算してみる。

多群の非線形判別

$$\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_p \partial \mathbf{w}_q^T} = - \sum_{i=1}^n \frac{\partial}{\partial \mathbf{w}_q^T} \pi_p(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i)$$

ここで

$$\frac{\partial}{\partial \mathbf{w}_q^T} \pi_p(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i) = \frac{\partial}{\partial \mathbf{w}_q^T} \frac{\exp \{ \mathbf{w}_p^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i)}{1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \}}$$

多群の非線形判別

$q = p$ ならば

$$\frac{\exp \{ \mathbf{w}_p^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T \left[1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} \right]}{\left[1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} \right]^2} - \frac{\exp \{ \mathbf{w}_p^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i) \exp \{ \mathbf{w}_p^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i)^T}{\left[1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} \right]^2}$$

多群の非線形判別

$$\begin{aligned}
 &= \pi_p(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T - \pi_p(\mathbf{x}_i; \mathbf{w})^2 \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T \\
 &= \pi_p(\mathbf{x}_i; \mathbf{w}) \{1 - \pi_p(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T
 \end{aligned}$$

$q \neq p$ ならば

$$\begin{aligned}
 &- \frac{\exp \{ \mathbf{w}_p^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i) \exp \{ \mathbf{w}_q^T \mathbf{b}(\mathbf{x}_i) \} \mathbf{b}(\mathbf{x}_i)^T}{\left[1 + \sum_{g=1}^{L-1} \exp \{ \mathbf{w}_g^T \mathbf{b}(\mathbf{x}_i) \} \right]^2} \\
 &= -\pi_p(\mathbf{x}_i; \mathbf{w}) \pi_q(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T
 \end{aligned}$$

多群の非線形判別

以上より

$$\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}_p \partial \mathbf{w}_q^T} = \begin{cases} - \sum_{i=1}^n \pi_p(\mathbf{x}_i; \mathbf{w}) \{1 - \pi_p(\mathbf{x}_i; \mathbf{w})\} \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T & \text{if } p = q \\ \sum_{i=1}^n \pi_p(\mathbf{x}_i; \mathbf{w}) \pi_q(\mathbf{x}_i; \mathbf{w}) \mathbf{b}(\mathbf{x}_i) \mathbf{b}(\mathbf{x}_i)^T & \text{if } p \neq q \end{cases}$$

2階微分を用いて数値的最適化によって基底関数の係数ベクトルを推定し、データ \mathbf{x} の各群へ属する事後確率を計算する。 L 個の事後確率を比較して最大となる群に属すると判断する。