判別分析 - 考え方

数理科学続論J

(Press `?` for help, `n` and `p` for next and previous slide)

村田 昇

2019.11.29

講義の予定

- **第1日: 判別分析の考え方**
- 第2日: 分析の評価

2 . 1

判別分析の考え方

3.1

判別分析 (discriminant analysis)

- 個体の特徴量から その個体の属するクラスを予測する関係式を構成
- 関係式: **判別関数** (discriminant function)
  - 特徴量: $X = (X_1, \ldots, X_q)$
  - クラス: Y ($K(\geq 2)$ 個のラベル)
- 判別関数による分類:
  - 1次式の場合: **線形判別分析** (linear discriminant analysis)
  - 2次式の場合: **2次判別分析** (quadratic discriminant analysis)

判別分析の例

- 検査結果から患者が病気を罹患しているか判定する
  - $X=$ 検査結果
  - $Y=$ 病気・健康
- 今日の経済指標から明日株価が上昇するか予測する
  - $X=$ 今日の経済指標
  - $Y=$ 明日株価の上昇・下降
- 今日の大気の状態から, 明日の天気を予測する
  - $X=$ 今日の大気の状態
  - $Y=$ 晴・くもり・雨・雪

判別分析の考え方

- 確率による定式化
  - $X = \boldsymbol{x}$ の下で $Y = k$ となる **条件付確率** を計算

$$p_k(\boldsymbol{x}) := P(Y = k | X = \boldsymbol{x})$$

  - 所属する確率が最も高いクラスに個体を分類
- 観測データ: n 個のデータ $(Y, X_1, \ldots, X_q)$

$$\{(y_i, x_{i1}, \ldots, x_{iq})\}_{i=1}^{n}$$

から $p_k(\boldsymbol{x})$ を構成

条件付確率

- 以下では X は離散型の q 次元確率変数で説明
- 事象 $X = \boldsymbol{x}$ が起きたという条件の下で 事象 $Y = k$ が起きる条件付確率

$$p_k(\boldsymbol{x}) = P(Y = k | X = \boldsymbol{x}) = \frac{P(Y = k, X = \boldsymbol{x})}{P(X = \boldsymbol{x})}$$

- (以下の議論は連続な確率変数でも成立)

条件付確率の表現

- $p_k(\boldsymbol{x})$ のモデル化の方針:
  - $p_k(\boldsymbol{x})$ を直接モデル化する (例:ロジスティック回帰)
  - $Y = k$ の下での X の条件付き確率質量関数

$$f_k(\boldsymbol{x}) = P(X = \boldsymbol{x} | Y = k) = \frac{P(X = \boldsymbol{x}, Y = k)}{P(Y = k)}$$

    のモデル化を通じて $p_k(\boldsymbol{x})$ をモデル化する
- 本講義では後者について説明

3.6

事後確率による判別

4.1

ベイズの公式 (Bayes' formula)

- $f_k(\boldsymbol{x})$ から $p_k(\boldsymbol{x})$ を得る数学的原理
- 「原因 $X = \boldsymbol{x}$ から結果 $Y = k$ が生じる確率」を「結果 $Y = k$ が生じたとき, 原因が $X = \boldsymbol{x}$ だった確率」から計算する方法
- ベイズの公式

$$P(Y = k|X = \boldsymbol{x}) = \frac{f_k(\boldsymbol{x})P(Y = k)}{\sum_{l=1}^{K} f_l(\boldsymbol{x})P(Y = l)}$$

ベイズの公式の略証

- 定義より

$$f_k(\boldsymbol{x}) = P(X = \boldsymbol{x} | Y = k) = \frac{P(X = \boldsymbol{x}, Y = k)}{P(Y = k)}$$

- 求める条件付確率:

$$P(Y = k | X = \boldsymbol{x}) = \frac{f_k(\boldsymbol{x}) P(Y = k)}{P(X = \boldsymbol{x})}$$

ベイズの公式の略証 (続)

- 分母の展開:

$$
\begin{aligned}
P(X = \boldsymbol{x}) &= \sum_{l=1}^{K} P(X = \boldsymbol{x}, Y = l) \\
&= \sum_{l=1}^{K} f_l(\boldsymbol{x}) P(Y = l)
\end{aligned}
$$

4.4

事前確率と事後確率

- **事前確率** (prior probability): $\pi_k = P(Y = k)$
  - $X = \boldsymbol{x}$ が与えられる前に予測されるクラス
- **事後確率** (posterior probability): $p_k(\boldsymbol{x})$
  - $X = \boldsymbol{x}$ が与えられた後に予測されるクラス
- ベイズの公式の書き換え:

$$p_k(\boldsymbol{x}) = \frac{f_k(\boldsymbol{x})\pi_k}{\sum_{l=1}^{K} f_l(\boldsymbol{x})\pi_l}$$

(事前確率が特徴量の確率の重みで変更される)

事前確率の決め方

- 事前に特別な情報がない場合:
  データから自然に決まる確率

$$\pi_k = \frac{Y = k\text{であるサンプル数}}{\text{全サンプル数}}$$

- 事前に情報がある場合:
  - 例: 身長や体重などの特徴量から喫煙者か否かを判別
    - 喫煙者の身長や体重などの特徴量を収集
    - 非喫煙者の身長や体重などの特徴量を収集
    - 事前確率は(別の調査の)日本人の喫煙者の割合を利用

4.6

線形判別分析

5.1

判別関数

- 判別の手続き
  - 特徴量 $X = \boldsymbol{x}$ の取得
  - 事後確率 $p_k(\boldsymbol{x})$ の計算
  - 事後確率最大のクラスにデータを分類
- **判別関数**: $\delta_k(\boldsymbol{x})\ (k = 1, \dots, K)$

$$p_k(\boldsymbol{x}) < p_l(\boldsymbol{x}) \Leftrightarrow \delta_k(\boldsymbol{x}) < \delta_l(\boldsymbol{x})$$

  事後確率の順序を保存する計算しやすい関数
- 判別関数 $\delta_k(\boldsymbol{x})$ を最大化するようなクラス k に分類

線形判別

- $f_k(\boldsymbol{x})$ の仮定:
  - q 変量正規分布の密度関数
  - 平均ベクトル $\boldsymbol{\mu}_k$: クラスごとに異なる
  - 共分散行列 Σ: すべてのクラスで共通

$$f_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{q/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right)$$

- 線形判別関数: $\boldsymbol{x}$ の1次式

$$\delta_k(\boldsymbol{x}) = \boldsymbol{x}^\top\Sigma^{-1}\boldsymbol{\mu}_k - \frac{1}{2}\boldsymbol{\mu}_k^\top\Sigma^{-1}\boldsymbol{\mu}_k + \log\pi_k$$

5.3

同値性の確認

- 事後確率と判別関数の関係

$$
\begin{aligned}
& p_k(\boldsymbol{x}) < p_k(\boldsymbol{x}) \\
& \Leftrightarrow f_k(\boldsymbol{x})\pi_k < f_l(\boldsymbol{x})\pi_l \\
& \Leftrightarrow \log f_k(\boldsymbol{x}) + \log \pi_k < \log f_l(\boldsymbol{x}) + \log \pi_l \\
& \Leftrightarrow -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) + \log \pi_k \\
& \qquad < -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_l)^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_l) + \log \pi_l \\
& \Leftrightarrow \delta_k(\boldsymbol{x}) < \delta_l(\boldsymbol{x})
\end{aligned}
$$

5.4

平均・分散の推定

- 平均の推定 (クラスごとに行う)

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n_k} \sum_{i:y_i=k} \boldsymbol{x}_i$$

- 分散の推定 (まとめて行う)

$$\hat{\Sigma} = \frac{1}{n-k} \sum_{k=1}^{K} \sum_{i:y_i=k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top$$

- ただし n_k は $y_i = k$ であるようなデータの総数

R: 線形判別関数 `lda()`

- 分析を行う関数: `MASS::lda()`
- データフレームに対する分析:
  - データフレーム `mydata`: 必要な変数を含むデータフレーム
  - 列名: x1の変数名, …, xpの変数名, yの変数名
- 書式は `lm()` とほぼ同じ

```
## データフレームを分析
require(MASS) # または library(MASS)
est <- lda(yの変数名 ~ x1の変数名 + ... + xpの変数名, data = mydata)
## 各クラスの判別関数値を図示
plot(est)
```

演習: 人工データによる線形判別

- 09-binary.r を確認してみよう

5.7

演習: 実データによる例

- 09-weather.r を確認してみよう

2次判別分析

6.1

2次判別

- $f_k(\boldsymbol{x})$ の仮定:
  - q 変量正規分布の密度関数
  - 平均ベクトル $\boldsymbol{\mu}_k$: クラスごとに異なる
  - 共分散行列 Σ_k: **クラスごとに異なる**

$$f_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{q/2}\sqrt{\det \Sigma_k}} \exp\Bigl(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}$$

- 2次判別関数: $\boldsymbol{x}$ の2次式

$$\delta_k(\boldsymbol{x}) = -\frac{1}{2}\det \Sigma_k - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) + \log$$

同値性の確認

- 事後確率と判別関数の関係

$$
\begin{aligned}
& p_k(\boldsymbol{x}) < p_l(\boldsymbol{x}) \\
& \Leftrightarrow f_k(\boldsymbol{x})\pi_k < f_l(\boldsymbol{x})\pi_l \\
& \Leftrightarrow \log f_k(\boldsymbol{x}) + \log \pi_k < \log f_l(\boldsymbol{x}) + \log \pi_l \\
& \Leftrightarrow -\frac{1}{2} \det \Sigma_k - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) + \log \pi_k \\
& \qquad\quad < -\frac{1}{2} \det \Sigma_l - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_l)^\top \Sigma_l^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_l) + \log \pi \\
& \Leftrightarrow \delta_k(\boldsymbol{x}) < \delta_l(\boldsymbol{x})
\end{aligned}
$$

6.3

平均・分散の推定

- 平均の推定 (クラスごとに行う)

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{n_k} \sum_{i:y_i=k} \boldsymbol{x}_i$$

- 分散の推定 (クラスごとに行う)

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} \sum_{i:y_i=k} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top$$

- ただし n_k は $y_i = k$ であるようなデータの総数

R: 2次判別関数 `qda()`

- 分析を行う関数: `MASS::qda()`
- データフレームに対する分析:
  - データフレーム `mydata`: 必要な変数を含むデータフレーム
  - 列名: x1の変数名, …, xpの変数名, yの変数名
- 書式は `lda()` と同じ

```
## データフレームを分析
require(MASS)
est <- qda(yの変数名 ~ x1の変数名 + ... + xpの変数名, data = mydata)
```

演習: 人工データによる2次判別

- 09-quad.r を確認してみよう

6.6

多値判別

7.1

多値判別の考え方

- 判別関数の比較 (判別関数 δ_k を比較)
- 2値判別の統合 (組み合わせ数 ${}_nC_2$)
- $k-1$ 個の特徴量への変換 (Rの線形判別)

7.2

変動の分解

- 3種類の変動の関係
  - 全変動　: $A = \sum_{i=1}^{n} (\boldsymbol{x}_i - \mu)(\boldsymbol{x}_i - \mu)^\top$
  - 群内変動: $W = \sum_{i=1}^{n} (\boldsymbol{x}_i - \mu_{y_i})(\boldsymbol{x}_i - \mu_{y_i})^\top$
  - 群間変動: $B = \sum_{k=1}^{K} n_k(\mu_k - \mu)(\mu_k - \mu)^\top$
    (n_k はクラス k のデータ数)

全変動A = 群内変動W + 群間変動B

Fisherの線形判別

- 特徴量 $Z = \alpha^\top X$
- 良い Z の基準:
  - クラス内では集まっているほど良い
  - クラス間では離れているほど良い
- Fisherの基準:

$$\text{maximize} \quad \alpha^\top B\alpha \quad \text{s.t.} \quad \alpha^\top W\alpha = \text{const.}$$

7.4

Fisherの線形判別の解

- α は $W^{-1}B$ の最大固有値 (主成分分析の導出と同様)
  - $K = 2$ の場合:

$$\alpha \propto W^{-1}(\mu_1 - \mu_2)$$

    (線形判別と一致する)
  - 一般の K の場合: 第1から第 $K-1$ 固有値を用いる
- 判別方法: 特徴量の距離を用いる
  - $d_k = \sum_{l=1}^{K-1}(\alpha_l^\top \boldsymbol{x} - \alpha_l^\top \mu_k)^2$ を計算
  - 最小の d_k となるクラス k に判別

演習: 3値判別の例

- 09-triple.r を確認してみよう

7.6

演習: 多値判別の例

- 09-multi.r を確認してみよう

演習: 実データによる例

- 以下のデータについて判別分析を行ってみよう
  - MASS::biopsy
  - MASS::crabs
  - rattle::wine

7.8