判別分析

基本的な考え方

村田 昇

講義概要

- **第 1 日: 判別分析の考え方**
- 第 2 日: 分析の評価

判別分析の考え方

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する方法
- 関係式: **判別関数** (discriminant function)
  - 説明変数: $X = (X_1, \dots, X_q)$
  - 目的変数: Y ($K(\geq 2)$ 個のクラスラベル)
- 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
  1. $X = \boldsymbol{x}$ の下で $Y = k$ となる **条件付確率** を計算

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

2. 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 $(Y, X_1, \dots, X_q)$ の組

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

- 観測データから Y の条件付確率 $p_k(\boldsymbol{x})$ を構成

条件付確率

- 以下では X は離散型の q 次元確率変数として説明
- 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

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

- 連続な確率変数の場合は確率密度関数を用いる

条件付確率の表現

- Y の条件付確率 $p_k(\boldsymbol{x})$ 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

事後確率による判別

Bayes の公式

- $f_k(\boldsymbol{x})$ 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 $Y = k$ が生じた原因が $X = \boldsymbol{x}$ である確率**から計算する方法

- Bayes の公式 (Bayes' formula)

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

Bayes 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:

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

- 分母の展開:

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

事前確率と事後確率

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

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

  - 事前確率が説明変数の条件付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

$$\pi_k = \frac{Y = k \text{ のサンプル数}}{\text{全サンプル数}}$$

- 事前に情報がある場合 (例):
  食事・運動・飲酒・ストレスなどの生活の特徴から生活習慣病か否かを判別
  - 健常者の食事・運動・飲酒・ストレスなどの特徴量を収集
  - 罹患者の食事・運動・飲酒・ストレスなどの特徴量を収集
  - 事前確率は **別の調査の日本人の罹患率** を利用

線形判別分析

判別関数

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

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

  - 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

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

同値性の確認

- 事後確率と判別関数の関係

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

平均・分散の推定

- 平均の推定 (クラスごとに行う)

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

  - ただし n_k は $y_i = k$ であるようなデータの総数

- 分散の推定 (まとめて行う)

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

演習

R: 線形判別関数 MASS::lda()

- データフレームに対する分析:

```
library(MASS) # または require(MASS)
lda(formula = y の変数名 ~ x1 の変数名 + ... + xp の変数名,
    data = データフレーム)
## formula: 目的変数名 ~ 説明変数名
## data: 目的変数，説明変数を含むデータフレーム
## 書式は lm() とほぼ同じ
```

- 判別関数値の図示:

```
est <- lda(formula = y の変数名 ~ x1 の変数名 + ... + xp の変数名,
           data = データフレーム)
plot(est)
```

練習問題

- 東京の気候データを用いて以下の分析を行いなさい
  - 10 月と 11 月の気温と湿度のデータを抽出する

```
TW.data <- read.csv("data/tokyo_weather.csv")
TW.subset  <- subset(TW.data,
                     subset= month %in% c(10,11),
                     select=c(temp,humid,month))
```

  - 半分のデータを用いて線形判別関数を構成し，残りのデータを用いて判別を行う

```
library(MASS)
idx <- seq(2,60,by = 2)
TW.train <- TW.subset[ idx,] # 訓練データ
TW.test  <- TW.subset[-idx,] # 試験データ
TW.lda <- lda(month ~ temp + humid, data=TW.train) # 線形判別関数の構成
TW.est <- predict(TW.lda) # 判別関数によるクラス分類結果の取得
TW.pred <- predict(TW.lda, newdata=TW.test) # 新しいデータの予測
```

2 次判別分析

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**

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

- 2 次判別関数: $\boldsymbol{x}$ の 2 次式

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

同値性の確認

- 事後確率と判別関数の関係

$$
\begin{aligned}
& p_k(\boldsymbol{x}) < p_l(\boldsymbol{x}) \\
& \Leftrightarrow f_k(\boldsymbol{x})\pi_k < f_l(\boldsymbol{x})\pi_l \\
& \qquad \text{(分母は共通)} \\
& \Leftrightarrow \log f_k(\boldsymbol{x}) + \log \pi_k < \log f_l(\boldsymbol{x}) + \log \pi_l \\
& \Leftrightarrow -\frac{1}{2}\det \Sigma_k - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) + \log \pi_k \\
& \qquad < -\frac{1}{2}\det \Sigma_l - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_l)^\top \Sigma_l^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_l) + \log \pi_l \\
& \Leftrightarrow \delta_k(\boldsymbol{x}) < \delta_l(\boldsymbol{x}) \\
& \qquad \text{(2 次の項は右辺と左辺で共通)}
\end{aligned}
$$

平均・分散の推定

- 平均の推定 (クラスごとに行う)

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

  - だたし n_k は $y_i = k$ であるようなデータの総数

- 分散の推定 (クラスごとに行う)

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

演習

問題

- 以下の問に答えなさい
  - X の条件付確率 $f_k(\boldsymbol{x})$ に関する仮定:
    * 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

    のもとで事後確率と 2 次判別関数の同値性

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

    を示しなさい

解答例

- 同値関係を順に確認すればよい

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

演習

R: 2 次判別関数 `MASS::qda()`

- データフレームに対する分析:

```
library(MASS) # または require(MASS)
qda(formula = y の変数名 ~ x1 の変数名 + ... + xp の変数名,
    data = データフレーム)
## formula: 目的変数名 ~ 説明変数名
## data: 目的変数, 説明変数を含むデータフレーム
```

練習問題

- 東京の気候データを用いて以下の分析を行いなさい
  - 前問と同様な設定で 2 次判別を行いなさい

    ```
    TW.qda <- qda(month ~ temp + humid, data=TW.train) # 2 次判別関数の構成
    TW.est <- predict(TW.qda) # 判別関数によるクラス分類結果の取得
    TW.pred <- predict(TW.qda, newdata=TW.test) # 新しいデータの予測
    ```

  - 別の月や変数を用いて判別分析を行いなさい

多値判別

多値判別の構成方法

- 判別関数の比較
  - 判別関数 δ_k を比較
  - 正規分布を仮定する場合は一般には 2 次判別
- 2 値判別の統合
  - 2 クラスでの比較: 最大の組合せ数 ${}_K C_2$
  - グループでの比較: 最大の組合せ数 $2^K - 2$
- $K-1$ 個の特徴量への変換
  - 説明変数の線形結合による特徴量の構成
  - **Fisher の線形判別**

変動の分解

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

$$
\begin{aligned}
&(\text{全変動}) = (\text{群内変動}) + (\text{群間変動})\\
&A = W + B
\end{aligned}
$$

演習

問題

- 以下の問に答えなさい
  - 全変動が群内・群間変動に分解されることを示しなさい
  - 説明変数の線形結合で新たな特徴量を構成する

$$
Z = \boldsymbol{\alpha}^\top X
$$

    このとき Z の群内変動と群間変動を求めなさい

解答例

- 定義どおりに計算する

$$
\begin{aligned}
A &= \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top\\
&= \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i} + \boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i} + \boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})^\top\\
&= \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})^\top + \sum_{i=1}^{n}(\boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})(\boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})^\top\\
&\quad + \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})(\boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})^\top + \sum_{i=1}^{n}(\boldsymbol{\mu}_{y_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})^\top
\end{aligned}
$$

- 添字の扱いに注意する

$$
\begin{aligned}
&= \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})^\top + \sum_{k=1}^{K}\sum_{i:y_i=k}(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top\\
&\quad + \sum_{k=1}^{K}\sum_{i:y_i=k}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top + \sum_{k=1}^{K}\sum_{i:y_i=k}(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top\\
&= \sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})(\boldsymbol{x}_i - \boldsymbol{\mu}_{y_i})^\top + \sum_{k=1}^{K} n_k(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^\top\\
&= W + B
\end{aligned}
$$

- 定義どおりに計算する

$$
\begin{aligned}
\sum_{i=1}^{n}(z_i-\mu_{y_i})^2 &= \sum_{i=1}^{n}(z_i-\mu_{y_i})(z_i-\mu_{y_i}) \\
&= \sum_{i=1}^{n}(\boldsymbol{\alpha}^\top \boldsymbol{x}_i - \boldsymbol{\alpha}^\top \boldsymbol{\mu}_{y_i})(\boldsymbol{\alpha}^\top \boldsymbol{x}_i - \boldsymbol{\alpha}^\top \boldsymbol{\mu}_{y_i})^\top \\
&= \boldsymbol{\alpha}^\top \sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_{y_i})(\boldsymbol{x}_i-\boldsymbol{\mu}_{y_i})^\top \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top W \boldsymbol{\alpha}
\end{aligned}
$$

$$
\sum_{k=1}^{K} n_k(\mu_k-\mu)^2 = \boldsymbol{\alpha}^\top B \boldsymbol{\alpha}
$$

Fisher の判別分析

Fisher の線形判別

- 判別のための特徴量 $Z = \boldsymbol{\alpha}^\top X$
- 良い Z の基準:
  - クラス内では集まっているほど良い ($\boldsymbol{\alpha}^\top W \boldsymbol{\alpha}$ は小)
  - クラス間では離れているほど良い ($\boldsymbol{\alpha}^\top B \boldsymbol{\alpha}$ は大)
- Fisher の基準:

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

  クラス内変動を一定にしてクラス間変動を最大化する

Fisher の線形判別の解

- $\boldsymbol{\alpha}$ は $W^{-1}B$ の固有値 (主成分分析の導出と同様)
  - $K=2$ の場合: 最大固有値を用いる (線形判別と一致)

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

  - 一般の K の場合: 第 1 から第 $K-1$ 固有値を用いる
- 判別の手続き:
  特徴量とクラスの中心までの距離を用いる
  1. $d_k = \sum_{l=1}^{K-1}(\boldsymbol{\alpha}_l^\top \boldsymbol{x} - \boldsymbol{\alpha}_l^\top \boldsymbol{\mu}_k)^2$ を計算
  2. 最小の d_k となるクラス k に判別

演習

練習問題

- 東京の気候データを用いて以下の分析を行いなさい
  - 9 月，10 月，11 月の気温と湿度のデータを用いて判別関数を作成しなさい.

```
TW.subset  <- subset(TW.data,
                     subset= month %in% c(9,10,11),
                     select=c(temp,humid,month))
TW.lda <- lda(month ~ temp + humid, data=TW.subset)
```

– 別の月や変数を用いて判別分析を行いなさい

```
## 雨の有無を識別する例
TW.mydata <- transform(TW.data,
                       rain=factor(rain>0), # 雨の有無でラベル化する
                       month=factor(month)) # 月ごとの気候の違いの補正のため
TW.mylda <- lda(rain ~ temp + solar + wind + month,
                data=TW.mydata)
```

次週の予定

- 第 1 日: 判別分析の考え方
- **第 2 日: 分析の評価**