Nospare

統計検定 (R) 準 1 級対策講座

回帰分析編

菅澤 翔之助

2023 年 7 月 26 日

本日の内容

目標問題

Nospare

本セミナーの目標: 回帰モデルに関する理解を深め，回帰モデルに関連する問題を解けるようになる。

以下の話題について解説する。

- 最小二乗法
- 不偏推定量
- 最小分散性
- 分散の不偏推定量
- 決定係数
- 赤池情報量規準 (AIC)
- 正則化
- 残差プロット
- Q-Q プロット
- 質的回帰

問 1

Nospare

次のモデルを考える。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_k\beta_k + \epsilon$$

ただし, $\epsilon \sim N(0, \sigma^2)$ であるとする。ここで, $(y_i, x_{i,1}, x_{i,2}, \ldots, x_{i,k})$ $(i = 1, \ldots, n)$ の n 個の観測データが与えられたとき, 最小二乗法を用いてベクトル β の推定量 $\hat{\beta}$ を求めよ。

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,2} & \cdots & x_{1,k} \\ \vdots & \vdots & \ddots & & & \vdots \\ 1 & x_{i,1} & & x_{i,i} & & x_{i,k} \\ \vdots & \vdots & & & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,i} & \cdots & x_{n,k} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

ただし, $X^T X$ は正定値行列であると仮定する。

問 2

最小二乗推定量 $\hat{\beta} = (X^T X)^{-1} X^T Y$ が β の不偏推定量となっていることを示せ。

問 3

Nospare

最小二乗推定量 $\hat{\beta} = (X^T X)^{-1} X^T Y$ について, $\mathrm{Var}(\hat{\beta})$ を求めよ。

問 4

回帰モデルの誤差分散 σ^2 の不偏推定量を求めよ。

問 5

推定値 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ を用いて予測値を

$$\hat{y}_i = \hat{\beta}_0 + x_{i,1}\hat{\beta}_1 + \ldots + x_{i,k}\hat{\beta}_k$$

と定義する。このとき, モデルの適合度を表す決定係数 R^2 は次のように与えられる。

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}$$

ただし, $\bar{y}$ は y_i の標本平均である。

この式を修正することで, モデルのパラメータ数を考慮した自由度調整済み決定係数 R_k^2 を与えよ。

問 6

Nospare

自由度調整済み決定係数以外に変数選択に用いられる指標として赤池情報量規準 (AIC) がある。AIC は一般的なモデル選択で使われているが，回帰モデルにおいては以下のように与えられる。

$$AIC = n\left(\log RSS + \log\frac{2\pi}{n} + 1\right) + 2(k+2)$$

ただし，RSS はモデルを使って得た予測値と実際の値の差の 2 乗の合計値 (残差平方和) である。変数選択の際には，AIC が最も小さくなるような変数の組み合わせが良いとされる。

今，20 個の観測データに対して以下のモデルを適用したところ $RSS = 10$ となった。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + \epsilon$$

このとき，AIC を計算せよ。

問 7

最小二乗法は $(Y - X\beta)^T(Y - X\beta)$ を最小化する β を求めるが, その代わりに

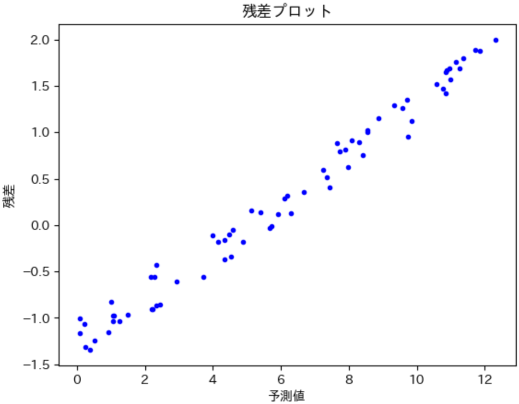
$$(Y - X\beta)^T(Y - X\beta) + P(\beta)$$

を最小化する β を求めることによって, 過学習 (過剰適合) を抑えることが可能になる。ここで, $P(\beta)$ は正則化項である。

Elastic net で用いられる正則化項 $P(\beta)$ の形を答えよ。

問 8

以下の残差プロットから読み取れる問題点を挙げよ。



問 9

以下の (a) に当てはまる関数形を答えよ。

残差を標準化 (平均 0, 分散 1 に変換) した後, 縦軸の値に標準化残差 (r) を小さい方から並べ, 横軸の値に累積分布関数の分位点 (z) を小さい順に取っていったものをプロットする。もし誤差に正規分布を仮定した場合, その仮定が適切であれば標準化残差は標準正規分布から発生していると考えられるため, プロットした結果は関数 (a) におおよそ乗ると考えられる。

問 10

次の (a),(b) に当てはまる語句を答えよ。

$Y \in \{0, 1\}$ を被説明変数としたとき, 説明変数 $x = (x_1, x_2, \ldots, x_k)$ を用いて $Y = 1$ となる確率 $\pi(x)$ を説明するモデルを考える。

以下のモデルは (a) と呼ばれる。

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

また, 以下のモデルは (b) と呼ばれる。

$$\pi(x) = \Phi(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k), \quad \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$$

本章

回帰分析に関する重要事項や各種性質などについて解説する。

最終的に上で挙げた目標問題を理解して解けるようになることを目指す。

回帰モデルの最小二乗法

Nospare

- 次の (重) 回帰モデルを考える。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_k\beta_k + \epsilon$$

ただし, $\epsilon \sim N(0, \sigma^2)$ とする。

- 以下のように $(y_i, x_{i,1}, x_{i,2}, ...x_{i,k})$ $(i = 1, \ldots, n)$ の n 個の観測データが与えられたとき, 最小二乗法を用いてベクトル β の推定量を求める。(このとき, X^TX が正定値行列であることを仮定する。)

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,2} & \cdots & x_{1,k} \\ \vdots & \vdots & \ddots & & & \vdots \\ 1 & x_{i,1} & & x_{i,i} & & x_{i,k} \\ \vdots & \vdots & & & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,i} & \cdots & x_{n,k} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

回帰モデルの最小二乗法

Nospare

X^TX が正定値行列であるため

$$\begin{aligned}
&\|Y - X\beta\|^2 \\
&= (Y - X\beta)^T(Y - X\beta) \\
&= [(X^TX)^{-1}X^TY - \beta]^T(X^TX)[(X^TX)^{-1}X^TY - \beta] \\
&\quad - Y^TX(X^TX)^{-1}X^TY + Y^TY
\end{aligned}$$

と変形できる。その結果，上式を最小化する β は

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

となる。 $\Rightarrow$ $\hat{\beta}$ を最小二乗推定量という。

回帰係数の不偏推定量

Nospare

- 最小二乗法によって求めた推定量 $\hat{\beta} = (X^TX)^{-1}X^TY$ が β の不偏推定量になっていることを確認する。

  (確認) パラメータ θ に対する推定量 $\hat{\theta}$ が不偏推定量であるとは $E[\hat{\theta}] = \theta$ が成り立つこと。

- Y は $Y = X\beta + \epsilon$ であり，ϵ が確率変数のため Y も確率変数になる。

  $\Rightarrow \quad Y \sim N(X\beta, \sigma^2 I_n)$ (I_n は $n \times n$ の単位行列)

- $E[Y] = X\beta \quad \mathrm{Var}(Y) = \sigma^2 I_n$

回帰係数の不偏推定量

Nospare

- 以下が成り立つ。

$$
\begin{aligned}
E[\hat{\beta}] &= (X^TX)^{-1}X^TE[Y] \\
&= (X^TX)^{-1}X^TX\beta = \beta
\end{aligned}
$$

- 最小二乗推定量 $\hat{\beta}$ は β の不偏推定量になっている。

最小二乗推定量の最小分散性

Nospare

- 以下が成り立つ。

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= (X^TX)^{-1}X^T\mathrm{Var}(Y)X(X^TX)^{-1}\\ &= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}\\ &= \sigma^2(X^TX)^{-1}\end{aligned}$$

- 上式の導出には以下の性質を用いている。

$$\mathrm{Var}(AX) = A\mathrm{Var}(X)A^T, \quad (AB)^T = B^TA^T, \quad (A^T)^{-1} = (A^{-1})^T$$

X: 確率ベクトル　A, B: 定数行列

- 実はこの分散共分散行列がクラメールラオ不等式の下限を達成していて最小の分散共分散行列となっているが示されている。

回帰モデルの誤差分散の不偏推定量

Nospare

- 回帰モデルにおける誤差分散 σ^2 の不偏推定量を求める。
- $P = X(X^TX)^{-1}X^T$ と定義すると, 予測値 $X\hat{\beta}$ は PY で表せる.

  $\Rightarrow$ 残差: $e = Y - X\hat{\beta} = (I_n - P)Y$
- P は以下の性質 (べき等性) を満たす「べき等行列」になっている。

$$P^T = P, \quad P^2 = P$$

- その他の性質: $PX = X \quad (I_n - P)X = 0$

  $\Rightarrow$ $e = (I_n - P)Y = (I_n - P)\epsilon$ と表せる。

回帰モデルの誤差分散の不偏推定量

- 残差二乗和 $e^T e$ の期待値を求める。　注: $\epsilon \sim N(0, \sigma^2 I_n)$

$$
\begin{aligned}
\mathrm{E}[e^T e] &= \mathrm{E}[\epsilon^T (I_n - P)^T (I_n - P)\epsilon] \\
&= \mathrm{E}[\epsilon^T (I_n - P)\epsilon] \\
&= n\sigma^2 - \mathrm{E}[\epsilon^T P\epsilon]
\end{aligned}
$$

- トレースの性質: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$$
\mathrm{E}[\epsilon^T P\epsilon] = \mathrm{tr}(PE[\epsilon\epsilon^T]) = \sigma^2 \mathrm{tr}(P)
$$

回帰モデルの誤差分散の不偏推定量

- $\mathrm{tr}(P)$ の評価

$$\begin{aligned}\mathrm{tr}(P) = \mathrm{tr}\{X(X^TX)^{-1}X^T\} &= \mathrm{tr}\{(X^TX)^{-1}(X^TX)\} \\ &= \mathrm{tr}(I_{k+1}) = k+1\end{aligned}$$

$$\Rightarrow \quad E[e^Te] = (n-k-1)\sigma^2$$

- σ^2 の不偏推定量

$$\hat{\sigma}^2 = \frac{e^Te}{n-k-1}$$

前半のまとめ

- 回帰係数は最小二乗法で推定することができる
- 最小二乗推定量は不偏推定量になっており最小分散性も持つ。
- 残差を用いることで誤差分散の不偏推定量が得られる。

後半の内容

- 変数選択の各種手法
- 回帰モデルの過学習 (過剰適合) を抑えるための方法
- 回帰モデルに仮定した誤差分布が適切であったかの確認
- 質的回帰の代表的な 2 つの手法の概要

決定係数

Nospare

- 最小二乗推定量を用いて予測値 $\hat{y}_1, \ldots, \hat{y}_n$ が得られる
- 変動の分解: (全変動) = (残差変動) + (回帰変動)

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$$

- 決定係数: モデルの適合度を表す指標

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$\bar{y}$: y_i の標本平均

- 決定係数は 0 から 1 の値を取り, 1 に近いほど適合度が高い

自由度調整済み決定係数

Nospare

- 決定係数の性質: 説明変数が増えるほど (k が大きくなるほど) R^2 は大きくなる

$$E\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right] = (n-k-1)\sigma^2$$

$$E\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right] = (n-1)\sigma^2$$

- 決定係数は変数選択には使えない (あるだけ説明変数を入れたモデルが良いという結論になってしまう)。
- 分母分子で “自由度” が異なるのが原因

$\Rightarrow$ 自由度調整を行なった決定係数が必要

自由度調整済み決定係数

- 自由度調整済み決定係数

$$
\begin{aligned}
R_k^2 &= 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-k-1)}{\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)} \\
&= 1 - \frac{n-1}{n-k-1}(1-R^2)
\end{aligned}
$$

- 決定係数とは異なり, 説明変数が増える (k が大きくなる) につれて必ずしも R_k^2 は増加しないため, 変数選択に用いることができる

赤池情報量規準 (AIC)

Nospare

- 変数選択に用いられる他の指標として赤池情報量規準 (AIC) がある。
- 回帰モデルにおける AIC は以下の式のように与えられる。

$$AIC = n\left(\log RSS + \log\frac{2\pi}{n} + 1\right) + 2(k+1)$$

ただし, RSS(残差平方和) はモデルによる予測値と観測値の差の 2 乗和である。

- AIC の値が最も小さくなるような変数の組み合わせを選ぶ。

赤池情報量規準 (AIC)

- 一般的な AIC の定義

  a を最大対数尤度, b をモデルのパラメータ数としたとき, AIC は以下のように与えられる。

$$AIC = -2a + 2b$$

- a: 大きいほど ($-a$ が小さいほど) モデルのデータに対する当てはまりが良い
- b: 小さいほどモデルがシンプル

  $\Rightarrow$ AIC が小さいほど良いモデル

- 回帰モデル $y \sim N(X\beta, \sigma^2 I_n)$ の最大対数尤度は以下で与えられる。

$$n\left(\log RSS + \log\frac{2\pi}{n} + 1\right)$$

正則化

Nospare

- 推定後に変数選択をして過学習を防ぐアプローチではなく，推定段階で過学習を抑えるような方法を考える。
- 最小二乗法は $(Y - X\beta)^T(Y - X\beta)$ を最小化する β を求めるが，その代わりに
$$(Y - X\beta)^T(Y - X\beta) + \lambda\|\beta\|_q^q$$
を最小化する β を求めることによって，過学習 (過剰適合) を抑えることが可能になる。($\|\beta\|_q$ は L_q ノルムで $\|\beta\|_q^q = \sum_{j=1}^k |\beta_k|^q$ である。)
- このような推定方法は正則化と呼ばれ，追加的な項 $\lambda\|\beta\|_q^q$ を正則化項と呼ぶ。q の値としては $q = 1$ (L1 正則化または Lasso) や $q = 2$ (L2 正則化または Ridge) がよく用いられる。
- λ は正則化項の影響を調整する値である。

正則化

- $L1$ 正則化の特徴

  変数選択ができる (対応する回帰係数を 0 に推定できる) ため, 過学習を効率的に抑えられるという長所があるが, 相関が強い変数は片方しか選ばれにくいという問題点がある。

- L_1 正則化に L_2 正則化を組み合わせることで, 上記の問題点が抑えられることが知られている。 $\Rightarrow$ Elastic net

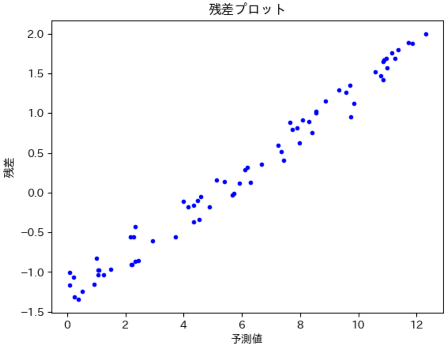
  Elastic net の正則化項

$$\lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right), \quad \alpha \in (0, 1)$$

残差プロット

Nospare

- 誤差 ϵ に関する仮定の妥当性を残差を用いて視覚的に確かめる
- 回帰モデルを推定し, 横軸に予測値, 縦軸に残差をとると以下のような図 (残差プロット) が得られた。



残差プロット

Nospare

- モデルの仮定が正しい場合: 残差に系統的な構造が残らない。(ランダムな量になる。)
- 系統的な構造が残っている場合: 必要な説明変数を取り入れていない, 誤差項の独立性の仮定が誤っているなどの問題があり得る。

  先の図は実際に独立な正規分布を誤差項に仮定した回帰モデルを使った残差プロットであるため, このモデルの特定化には何らかの誤りがある可能性が考えられる。
- 残差プロットは誤差項における等分散性の仮定や独立性の仮定の正当性に加えて, 外れ値の検出にも有用である。

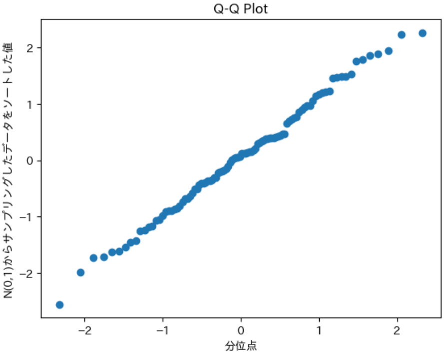
Q-Q プロット

Nospare

- 残差を標準化 (平均 0, 分散 1 に変換) した後, 縦軸の値に標準化残差 (r) を小さい方から並べ, 横軸の値に累積分布関数の分位点 (z) を小さい順に取っていったものを Q-Q プロットと呼ぶ。
- もし誤差に正規分布を仮定した場合, その仮定が適切であれば標準化残差は標準正規分布から発生していると考えられるため, プロットした結果は関数 $r = z$ (45 度線) におおよそ乗ると考えられる。
- もし 45 度線から明らかに逸脱している場合, 誤差項に対する正規分布の仮定を見直す必要がある。

Q-Q プロット

- 例: 正規分布 $N(0,1)$ から独立に生成したデータについての Q-Q プロット (実際におおよそ 45 度線上に乗っている。)



質的回帰

Nospare

- $Y \in \{0, 1\}$ を被説明変数としたとき, 説明変数 $x = (x_1, x_2, \ldots, x_k)$ を用いて $Y = 1$ となる確率 $\pi(x)$ を説明するモデルを考える。
- ロジスティック回帰モデル

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

- プロビット回帰モデル

$$\pi(x) = \Phi(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k), \quad \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$$

質的回帰

Nospare

- Y と x の組が観測データとして与えられれたとき, 基本的には最尤推定を用いてパラメータ $\beta_0, \beta_1, \ldots, \beta_k$ を推定する。
- (重) 回帰モデルとは異なり, 最尤推定量は解析的に得られないためコンピュータを用いた数値計算を用いる必要がある。

問 1

次のモデルを考える。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_k\beta_k + \epsilon$$

ただし, $\epsilon \sim N(0, \sigma^2)$ であるとする。ここで, $(y_i, x_{i,1}, x_{i,2}, \ldots, x_{i,k})$ $(i = 1, \ldots, n)$ の n 個の観測データが与えられたとき, 最小二乗法を用いてベクトル β の推定量 $\hat{\beta}$ を求めよ。

$$X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,2} & \cdots & x_{1,k} \\ \vdots & \vdots & \ddots & & & \vdots \\ 1 & x_{i,1} & & x_{i,i} & & x_{i,k} \\ \vdots & \vdots & & & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,i} & \cdots & x_{n,k} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

ただし, $X^T X$ は正定値行列であると仮定する。

問 1 解答

Nospare

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

問 2

最小二乗推定量 $\hat{\beta} = (X^T X)^{-1} X^T Y$ が β の不偏推定量となっていることを示せ。

問 2 解答

Nospare

$$
\begin{aligned}
E[\hat{\beta}] &= (X^TX)^{-1}X^TE[Y] \\
&= (X^TX)^{-1}X^TX\beta = \beta
\end{aligned}
$$

問 3

Nospare

最小二乗推定量 $\hat{\beta} = (X^T X)^{-1} X^T Y$ について, $\mathrm{Var}(\hat{\beta})$ を求めよ。

問 3 解答

Nospare

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}) &= (X^T X)^{-1} X^T \mathrm{Var}(Y) X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

問 4

回帰モデルの誤差分散 σ^2 の不偏推定量を求めよ。

問 4 解答

$$\hat{\sigma}^2 = \frac{e^T e}{n - k - 1}$$

問 5

Nospare

推定値 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ を用いて予測値を

$$\hat{y}_i = \hat{\beta}_0 + x_{i,1}\hat{\beta}_1 + \ldots + x_{i,k}\hat{\beta}_k$$

と定義する。このとき, モデルの適合度を表す決定係数 R^2 は次のように与えられる。

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}$$

ただし, $\bar{y}$ は y_i の標本平均である。

この式を修正することで, モデルのパラメータ数を考慮した自由度調整済み決定係数 R_k^2 を与えよ。

問 5 解答

$$R_k^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

問 6

自由度調整済み決定係数以外に変数選択に用いられる指標として赤池情報量規準 (AIC) がある。AIC は一般的なモデル選択で使われているが，回帰モデルにおいては以下のように与えられる。

$$AIC = n\left(\log RSS + \log\frac{2\pi}{n} + 1\right) + 2(k+2)$$

ただし，RSS はモデルを使って得た予測値と実際の値の差の 2 乗の合計値 (残差平方和) である。変数選択の際には，AIC が最も小さくなるような変数の組み合わせが良いとされる。

今，20 個の観測データに対して以下のモデルを適用したところ $RSS = 10$ となった。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + \epsilon$$

このとき，AIC を計算せよ。

問 6 解答

Nospare

モデルにおいて説明変数の個数は 4 つであり，与えられた式に $k = 4$, $n = 20$, $RSS = 10$ を代入し計算することで $AIC \approx 54.89$ となる。

注: AIC の値自体に特に意味はない (複数モデルで比べることに意味がある)

問 7

Nospare

最小二乗法は $(Y - X\beta)^T(Y - X\beta)$ を最小化する β を求めるが, その代わりに

$$(Y - X\beta)^T(Y - X\beta) + P(\beta)$$

を最小化する β を求めることによって, 過学習 (過剰適合) を抑えることが可能になる。ここで, $P(\beta)$ は正則化項である。

Elastic net で用いられる正則化項 $P(\beta)$ の形を答えよ。

問 7 解答

$$P(\beta) = \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right), \quad \alpha \in (0, 1)$$

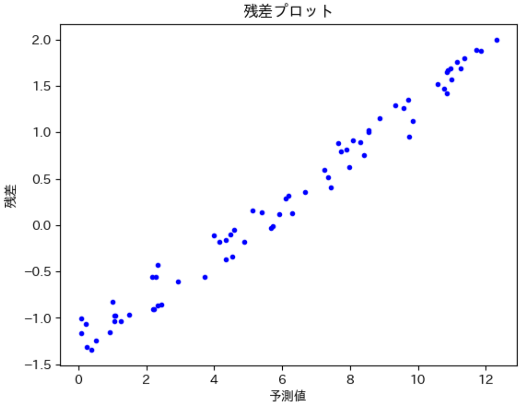
または

$$P(\beta) = \lambda \sum_{j=1}^{k} \left(\alpha |\beta_j| + \frac{1-\alpha}{2} \beta_j^2 \right), \quad \alpha \in (0, 1)$$

問 8

Nospare

以下の残差プロットから読み取れる問題点を挙げよ。



問 8 解答

Nospare

残差には互いに独立な正規分布を用いているのにも関わらず, 明らかに独立性を満たしていない残差になっている。そのため残差に互いに独立な正規分布を用いることを考え直すべきである。

問 9

Nospare

以下の (a) に当てはまる関数形を答えよ。

残差を標準化 (平均 0, 分散 1 に変換) した後, 縦軸の値に標準化残差 (r) を小さい方から並べ, 横軸の値に累積分布関数の分位点 (z) を小さい順に取っていったものをプロットする。もし誤差に正規分布を仮定した場合, その仮定が適切であれば標準化残差は標準正規分布から発生していると考えられるため, プロットした結果は関数 (a) におおよそ乗ると考えられる。

問 9 解答

Nospare

答: $r = z$

問 10

次の (a),(b) に当てはまる語句を答えよ。

$Y \in \{0,1\}$ を被説明変数としたとき, 説明変数 $x = (x_1, x_2, \ldots, x_k)$ を用いて $Y = 1$ となる確率 $\pi(x)$ を説明するモデルを考える。

以下のモデルは (a) と呼ばれる。

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

また, 以下のモデルは (b) と呼ばれる。

$$\pi(x) = \Phi(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k), \quad \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$$

問 10 解答

Nospare

(a): ロジスティック回帰モデル

(b): プロビット回帰モデル

演習問題

Nospare

父親の身長, 髪の長さから子供の身長の長さを推定する回帰モデルを考える。

以下は 4 組の親子についてのデータである。

A 家: 父 身長 180cm 髪の長さ 5cm 子供 身長 178cm

B 家: 父 身長 160cm 髪の長さ 11cm 子供 身長 160cm

C 家: 父 身長 170cm 髪の長さ 11cm 子供 身長 170cm

D 家: 父 身長 150cm 髪の長さ 0cm 子供 身長 152cm

問 1

Nospare

x_1 を父親の身長, x_2 を父親の髪の長さ, $\epsilon \sim N(0,1)$ とした時,
次の 2 つのモデルを考える。

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \epsilon \tag{1}$$

$$y = \beta_3 + x_1\beta_4 + \epsilon \tag{2}$$

このとき最小二乗法を用いて β_0, β_1, β_2, β_3, β_4 を決定せよ。

問 1 解答

行列計算により

$$\beta_0 = \frac{4791}{245},\ \beta_1 = \frac{2161}{2450},\ \beta_2 = -\frac{2}{147},\ \beta_3 = \frac{99}{5},\ \beta_4 = \frac{22}{25}$$

となる。

問2

Nospare

問１で得られた２つのモデルについて決定係数の大小関係を求め，どちらのモデルがより適合性があるか判断せよ。

問 2 解答

Nospare

- (1) のモデルの決定係数 ≈ 0.9980
- (2) のモデルの決定係数 ≈ 0.9979

となる。

$\Rightarrow$ (1) のモデルの方が適合度が高い。

問 3

問１で得られた２つのモデルについて自由度調整済み決定係数の大小関係を求め，どちらのモデルがより良いか判断せよ。

問 3 解答

Nospare

- (1) のモデルの自由度調整済み決定係数 ≈ 0.01
- (2) のモデルの自由度調整済み決定係数 ≈ 0.99

となる。

$\Rightarrow$ (2) の方が良いモデル

(決定係数と自由度調整済み決定係数で大小関係が逆転する。)