線形回帰モデルの変数選択法とそれによる解析例

1、データの説明：三浦

2、線形回帰モデルの説明：大崎

3、MallowsのCp規準：黒木，岡田

4、変数選択法と解析例：浅野，辻

アメリカ合衆国の主要60都市における①
大気汚染の健康に対するデータの一部

City	RA	ED	…	$\log_{10}SO_2$	MORT
OH	36	11.4	…	1.77	921.9
NY	35	11.0	…	1.59	997.9
PA	44	9.8	…	1.52	962.4

線形回帰モデル

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

$$i = 1, 2, \cdots, n$$

$$y = X \beta + \varepsilon$$

$n \times 1$　　$n \times (p+1)$　$(p+1) \times 1$　　$n \times 1$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + x_{11}\beta_1 + \cdots + x_{1p}\beta_p \\ \beta_0 + x_{21}\beta_1 + \cdots + x_{2p}\beta_p \\ \vdots \\ \beta_0 + x_{n1}\beta_1 + \cdots + x_{np}\beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$= \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & & \vdots \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

代表的なモデル選択

Akaikeの AIC 規準 → 小さいほど良いモデル

$$AIC = -2\ell(\hat{\beta}) + 2(p+1)$$

$$= -2(\text{最大対数尤度}) + 2(\text{モデルに含まれる独立なパラメータの数})$$

Mallowsの C_p 規準 → 小さいほど良いモデル

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} + 2(p+1) - n$$

$$= \frac{RSS_p}{\hat{\sigma}^2} + 2 \times (\text{モデルに含まれる独立なパラメータの数}) - n$$

真のモデル

$$
\begin{cases}
E(y) = 0 \\
VAR(y) = \tau^2 I_n
\end{cases}
$$

$y = X\beta + \varepsilon$ の回帰モデル

$$
\begin{cases}
E(y) = X\beta \\
VAR(y) = \sigma^2 I_n
\end{cases}
$$

$$\hat{\beta} = ({}^tXX)^{-1}{}^tXy$$

$$\hat{y} = \hat{\theta} = \underbrace{X({}^tXX)^{-1}{}^tX}_{H:\text{ハット行列}}y = Hy$$

真の世界で評価すると $E(\hat{\theta}) = H\theta \neq \theta$

推定値の良さを評価する尺度として

$$\hat{\theta} - \theta = (\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)$$

$$\Delta p = E^t(\hat{\theta} - \theta)(\hat{\theta} - \theta)$$

$$= \underbrace{(p+1)\tau^2}_{\text{ペナルティ}} + \underbrace{{}^t\theta(I-H)\theta}_{\text{バイアス}}$$

残差平方和 $RSS_p = {}^t(y-\hat{\theta})(y-\hat{\theta})$ ⑦

真の世界では

$$E(RSS_p) = (n-p-1)\tau^2 + {}^t\theta(I-H)\theta$$

Δ_p の不偏推定量は

$$RSS_p + \{2(p+1) - n\}\hat{\tau}^2$$

Mallows の C_p 規準

$$C_p = \frac{RSS_p}{\hat{\tau}^2} + 2(p+1) - n = \left(\begin{array}{c}\text{モデルに含まれる}\\ \text{独立なパラメータ数}\end{array}\right) - n$$

変数選択法

フルモデルのサイズ…p，サブモデルを合わせた数…2^p-1

●総当たり法

2^p-1通り（$p=6$のとき63通り）

すべての組み合わせの中から検討する方法

…とても大変

●逐次選択法

逐次的に変数を選択する方法

・変数増加法　・変数減少法

変数増加法のアルゴリズム

Step1: ・$i \leftarrow 1$（i：選択された変数の個数）

・MORTとの相関係数が最大となる説明変数をえらぶ。

・C_iを求める。

Step2: ・$i \leftarrow i+1$

・最大の寄与率$\left(\frac{RSS_i}{\sum(y_i-\bar{y}_i)^2}\right)$となる説明変数をえらぶ。

・C_iを求める。

Step3: if $C_{i-1} > C_i$: Step2へ戻る

else $C_{i-1} \leq C_i$: 選択を終了する

変数減少法のアルゴリズム

Step 1 : $i \leftarrow p$　（i：選択された変数の個数）

寄与率とC_pを計算する。

Step 2 : $i \leftarrow i-1$

フルモデルの寄与率からの減少量が最小となる変数を選ぶ。

Step 3 : $C_i \leq C_{i+1}$　YES ⟶ Step 2 へ戻る

NO ---→ 終了。