データ分析基礎
回帰分析 補足スライド

京都大学 国際高等教育院 附属データ科学イノベーション教育研究センター

せきど ひろと
關戸 啓人

sekido.hiroto.7a@kyoto-u.ac.jp

多重共線性と主成分回帰

多重共線性

- ★ 重回帰モデル $B = aT + bK + c + \varepsilon$
  - ★ B：ビールの売上
  - ★ T：東京の気温
  - ★ K：京都の気温
- ★ このように，説明変数間で相関がある場合，多重共線性という問題が起こり，最小二乗推定量が不安定になる

- ★ より正確には，データ行列の条件数が大きいと不安定になる
  - ★ $\mathrm{cond}(A) = \sigma_{\max}(A)/\sigma_{\min}(A)$
  - ★ 条件数が大きいと，ちょっとした摂動で連立一次方程式の解が大きく変わる

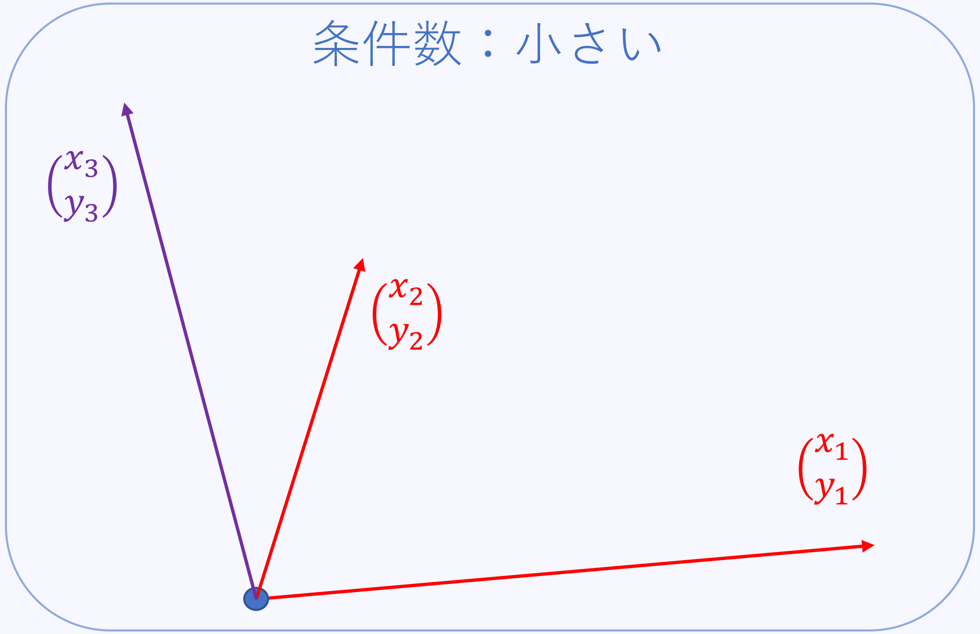
多重共線性

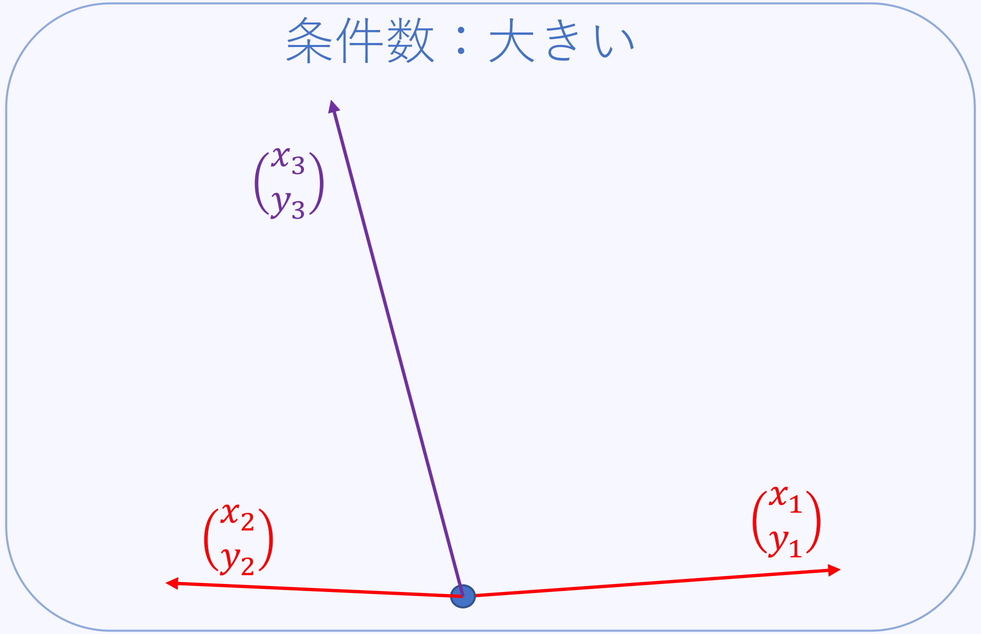
★ 連立一次方程式を次のように解釈する

$$\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} \iff a \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + b \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$$

★ 2つのベクトルをどのように足せば，ベクトル $(x_3 \ \ y_3)^{\mathrm{T}}$ を作れるか？

★ このとき，条件数が大きいというのは，2つのベクトルがほぼ線形従属であることを意味する





多重共線性

★ 回帰分析に於いても，強い相関のあるデータ（似たようなデータ，ほぼ同じ方向を向いたデータ）を用いると結果が不安定になる

★ 重回帰モデルを次のように書き換える

$$\begin{aligned} B &= aT + bK + c + \varepsilon \\ &= a'(T+K) + b'(T-K) + c + \varepsilon \end{aligned}$$

- ★ ただし，$a' + b' = 2a,\ a' - b' = 2b$

★ T, K はほぼ同じ傾向を示すから，$T - K$ は 0 に近い値ばかり取る

- ★ その状況で，$T - K$ の情報を使って「データの近くを通ろうと」すると $T - K$ の係数 b' の推定結果は絶対値が大きく不安定になる
- ★ それが伝搬して，結果的に a, b の推定結果も不安定になる

多重共線性

- ★ 多重共線性の回避方法
  - ★ 不安定になっているように見えたら説明変数を減らす
  - ★ 説明変数を無相関にする（連立一次方程式に於いてベクトルが直交するようにする）

- ★ 説明変数を以下に取り替えて，a', b', c を推定することにする
  - ★ $T+K$：全国的な気温
  - ★ $T-K$：関東と開催での気温の差
- ★ こうすると，2つの説明変数 $T+K$ と $T-K$ はあまり相関は強くない
  - ★ 互いに影響を及ぼし，全体的に推定結果が不安定になることはない
  - ★ $T-K$ は全体的に値が小さく，ビールの売上をうまく説明できないであろうため，この係数に対してはうまく行かない
    - ★ $T-K$ は説明変数として不要

多重共線性

★ 説明変数の無相関化

- ★ 説明変数を無相関にする方法として主成分分析を行う方法がある
- ★ 主成分分析した結果を用いて回帰分析を行うことを主成分回帰と呼ぶ

多重共線性と主成分回帰

★ $A = UDV^{\mathrm{T}}$ と特異値分解されるとする

★ $U^{\mathrm{T}}U = I,\ \ V^{\mathrm{T}}V = VV^{\mathrm{T}} = I$ で D は対角行列で対角成分は全て正

★ 最小二乗推定量は $A^{\mathrm{T}}A\beta = A^{\mathrm{T}}y$ を満たす β のことだから，

$$
\begin{aligned}
&A^{\mathrm{T}}A\beta = A^{\mathrm{T}}y \\
&(UDV^{\mathrm{T}})^{\mathrm{T}}(UDV^{\mathrm{T}})\beta = (UDV^{\mathrm{T}})^{\mathrm{T}}y \\
&VDU^{\mathrm{T}}UDV^{\mathrm{T}}\beta = VDU^{\mathrm{T}}y \\
&VD^2V^{\mathrm{T}}\beta = VDU^{\mathrm{T}}y \\
&\beta = VD^{-1}U^{\mathrm{T}}y
\end{aligned}
$$

★ **ここで，D^{-1} の要素が大きい部分が不安定になる**

多重共線性と主成分回帰

★ $A = UDV^{\mathrm{T}}$ と特異値分解されるとする

★ $U^{\mathrm{T}}U = I,\ \ V^{\mathrm{T}}V = VV^{\mathrm{T}} = I$ で D は対角行列で対角成分は全て正

★ **説明変数を主成分 $AV = UD$ とする**

★ 最小二乗推定量は $(AV)^{\mathrm{T}}(AV)\beta = (AV)^{\mathrm{T}}y$ を満たす β のことだから,

$$
\begin{aligned}
&A^{\mathrm{T}}A\beta = A^{\mathrm{T}}y \\
&(UD)^{\mathrm{T}}(UD)\beta = (UD)^{\mathrm{T}}y \\
&DU^{\mathrm{T}}UD\beta = DU^{\mathrm{T}}y \\
&D^2\beta = DU^{\mathrm{T}}y \\
&\beta = D^{-1}U^{\mathrm{T}}y
\end{aligned}
$$

★ **ここで，D^{-1} の要素が大きい部分，下位の主成分の係数が不安定になる**

正則化付き最小二乗法

Ridge回帰

★ 通常の回帰分析（最小二乗法）

★ $\sum_{k=1}^{n}(y_k - f(x_k;\beta))^2 \to \text{minimize}$

★ Ridge回帰

★ $\sum_{k=1}^{n}(y_k - f(x_k;\beta))^2 + \lambda\|\beta\|_2^2 \to \text{minimize}$

★ $\sum_{k=1}^{n}(y_k - f(x_k;\beta))^2 + \lambda(\beta_1^2 + \beta_2^2 + \ldots + \beta_m^2) \to \text{minimize}$

★ 解くべき連立一次方程式 $(A^{\mathrm{T}}A + \lambda I)\beta = A^{\mathrm{T}}y$

★ 通常は切片にはペナルティを課さないことが多い

★ $\lambda > 0$ はパラメータ

Lasso 回帰

★ 通常の回帰分析（最小二乗法）

- ★ $\sum_{k=1}^{n} (y_k - f(x_k;\beta))^2 \rightarrow \text{minimize}$

★ Lasso 回帰

- ★ $\sum_{k=1}^{n} (y_k - f(x_k;\beta))^2 + \lambda \|\beta\|_1 \rightarrow \text{minimize}$
- ★ $\sum_{k=1}^{n} (y_k - f(x_k;\beta))^2 + \lambda(|\beta_1| + |\beta_2| + \ldots + |\beta_m|) \rightarrow \text{minimize}$
  - ★ スパース推定
    - ★ 係数の推定結果が 0 になりやすい
    - ★ 0 になった場合，その説明変数は不要だったのではないかと考えられる

モデル選択

モデル選択

- ★ 変数指定法
  - ★ 理論的に正しいモデルが証明できる
- ★ 総当たり法
  - ★ 候補になるモデルを全て試し，何らかの基準で最も良いと思われるものを採用
    - ★ AIC，C_p 基準，自由度調整済み決定係数，クロスバリデーション
- ★ 逐次変数選択法
  - ★ 一定の規則に従い，変数を追加したり削除する
  - ★ 総当たり法における局所探索にあたる場合もある
- ★ …

赤池情報量基準 AIC

- ★ 回帰分析の場合 (定数の差は気にしないことにして)
  - ★ $\mathrm{AIC} = n \log(S/n) + 2p$
    - ★ n: データ数
    - ★ S: 残差の二乗和
    - ★ p: パラメータの数
  - ★ この値が小さい方が「良い」モデル

- ★ KL 情報量と最尤推定の理論から出てくるもの
- ★ 通常の線形回帰分析にのみ利用可能
  - ★ WAIC（広く使える情報量規準）

クロスバリデーション（交差検証）

★ 訓練用のデータと，テスト用のデータに分けて，訓練用のデータで学習（係数を推定）し，テスト用のデータを用いてどの程度良いかを評価する

★ 訓練用のデータと，テスト用のデータを変更し，それを繰り返す

★ leave-one-out 交差検証（LOOCV）

- ★ データ数を n とし，訓練用データ $n-1$ 個，テスト用データ 1 個とする
- ★ テスト用データを取り替えながら n 回繰り返す

★ K-fold 交差検証（K 分割交差検証）

- ★ データを K 個のグループに分け，訓練用データ $K-1$ グループ，テスト用データ 1 グループのデータとする
- ★ テスト用データを取り替えながら K 回繰り返す

逐次変数選択法

- ★ 変数増加法
  - ★ 最初，説明変数はなし
  - ★ 説明変数を1つ加えたとき，その係数に対する検定の P 値が最も小さいものを調べる
  - ★ その P 値の値が，一定値より小さければ加えて，同じ手順を繰り返す．一定値より大きければ終了する
- ★ 変数減少法
  - ★ 最初，全ての説明変数を使ったモデルから考える
  - ★ 説明変数の係数に対する検定の P 値が最も大きいものを調べる
  - ★ その P 値の値が，一定値より大きければ説明変数から外し，同じ手順を繰り返す．一定値より小さければ終了する
- ★ 変数増減法