第 7 回：重回帰分析
【教科書第 6 章第 1 節～第 3 節】

北村　友宏

2025 年 11 月 11 日

本日の内容

1. 重回帰モデル
2. OLS 推定量の性質
3. 実証分析例

重回帰

- 定数項以外に説明変数が複数ある回帰モデルを重回帰モデル（multiple regression model）という．

定数項以外に説明変数が k 個ある場合，

$$
\begin{aligned}
y_i =& \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \\
& + \beta_k x_{ik} + u_i, \\
E(u_i \mid x_{i1}, x_{i2}, \cdots, x_{ik}) =& 0, \\
E(u_i u_j \mid x_{i1}, x_{i2}, \cdots, x_{ik}) =& 0 \quad (i \neq j), \\
V(u_i \mid x_{i1}, x_{i2}, \cdots, x_{ik}) =& \sigma^2, \\
i =& 1, 2, \cdots, n.
\end{aligned}
$$

各観測値の式を並べると，

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + u_1,\\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + u_2,\\
&\vdots\\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + u_n.
\end{aligned}
$$

ベクトル・行列を用いて表示すると，

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.
$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},$$

$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ とすると，重回帰モデルは次のように簡潔に表すことができる．

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\beta} + \boldsymbol{u}, \\ E(\boldsymbol{u} \mid \boldsymbol{X}) &= \boldsymbol{0}, \\ V(\boldsymbol{u} \mid \boldsymbol{X}) &= \sigma^2 \boldsymbol{I}_n. \end{aligned}$$

モデルを

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e},$$

と書き換え，

$$\sum_{i=1}^{n} e_i^2 = \boldsymbol{e}'\boldsymbol{e} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)' \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right),$$

が最小になるように OLS 推定量を求めると，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

- $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$

（導出方法は付録参照）

OLS 推定における仮定（重回帰の場合）

- 「説明変数を所与とした誤差項の条件付き期待値」と，「誤差項の条件なし期待値」が等しく，その値はゼロ．
  - $E(\boldsymbol{u} \mid \boldsymbol{X}) = E(\boldsymbol{u}) = \boldsymbol{0}$.

  $\Rightarrow$ 説明変数と誤差項は平均独立で，誤差項の期待値はゼロ．
- 説明変数を所与として，誤差項の分散は一定で，異なる個体の誤差項同士は無相関．
  - $V(\boldsymbol{u} \mid \boldsymbol{X}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \boldsymbol{I}_n$.
- 説明変数を所与として，誤差項は正規分布に従う．
  - $\boldsymbol{u} \mid \boldsymbol{X} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n\right)$.

偏回帰係数

- 重回帰モデルの回帰係数を偏回帰係数（partial regression coefficient）という．
- 重回帰モデル

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i,$$

の偏回帰係数 β_j $(j = 1, 2, \cdots, k)$ は，「仮に x_{ij} 以外の変数の値を一定水準に固定したときに，$x_{i1}, x_{i2}, \cdots, x_{ik}$ を所与とした y_i の条件付き期待値に x_{ij} が与える影響」を測る．
  - e.g., 仮に熟練度がすべての個人で一定だったときの，修学年数が年収の条件付き期待値に与える影響．

- 他の変数の値を一定水準に固定することを「他の条件を一定とする（*ceteris paribus*）」という．
    - e.g.,「他の条件を一定としたうえで，修学年数が1年長くなると年収が平均して○円高くなる傾向がある」などと表現．

- y_i の条件付き期待値をとった

  $$E(y_i \mid x_{i1}, x_{i2}, \cdots, x_{ik}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik},$$

  を x_{ij} で偏微分（他の説明変数の値は一定）すると，β_j になる．

  $$\frac{\partial E(y_i \mid x_{i1}, x_{i2}, \cdots, x_{ik})}{\partial x_{ij}} = \beta_j.$$

- その他の変数の影響が一定であるという状況を作り出すための説明変数をコントロール変数（control variable）という．
- x_{ij} が y_i に与える影響に興味がある場合，x_{ij} 以外の説明変数はコントロール変数．

古典的線形回帰モデル

- 説明変数を所与として，誤差項の分散が一定で，異なる個体の誤差項同士が無相関な線形回帰モデルを古典的線形回帰モデル（classical linear regression model）という．
- 重回帰モデルの場合，

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \\
E(\boldsymbol{u} \mid \boldsymbol{X}) &= \mathbf{0}, \\
V(\boldsymbol{u} \mid \boldsymbol{X}) &= \sigma^2 \boldsymbol{I}_n
\end{aligned}
$$

であれば古典的線形回帰モデルとなる．

※ $E(\boldsymbol{u} \mid \boldsymbol{X}) = \mathbf{0}$ の条件は，回帰モデルであるために必要．

OLS 推定量の性質

- 古典的線形回帰モデルの偏回帰係数ベクトルの OLS 推定量 $\hat{\boldsymbol{\beta}}$ の期待値は,

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}.$$

➥ 古典的線形回帰モデルの（偏）回帰係数の OLS 推定量は不偏性をもつ不偏推定量である.
（証明は省略）

- 被説明変数の１次関数で表される推定量を線形推定量（linear estimator）という．
  - 単回帰モデル $y_i = \beta_0 + \beta_1 x_i + u_i$ の場合，

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

  （証明は省略）
  - 重回帰モデル $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ の場合，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

  $\Rightarrow$ どちらも被説明変数 y_i または $\boldsymbol{y}$ の１次関数なので，線形推定量である．

- 不偏性をもつ線形推定量を線形不偏推定量（linear unbiased estimator）という．

$$\Downarrow$$

古典的線形回帰モデルの（偏）回帰係数の OLS 推定量は線形不偏推定量である.

- 線形不偏推定量の中で分散が最小の推定量を最良線形不偏推定量（Best Linear Unbiased Estimator, BLUE）という.
  - 有効性と不偏性をもつ線形推定量.

- ガウス＝マルコフ定理：
  古典的線形回帰モデルの（偏）回帰係数の OLS 推定量は BLUE である.
  （証明は省略）

実証分析例：ミンサー方程式の推定

就業可能年数をコントロールしたうえで，「修学年数が増えると，年収がどれだけ増えるのか」を分析するためのモデル（ミンサー方程式）

$$\ln income_i = \beta_0 + \beta_1 yeduc_i + \beta_2 exper_i + \beta_3 exper_i^2 + u_i$$

- $income_i$：年収（万円）
- $yeduc_i$：修学年数（年）
- $exper_i$：就業可能年数（年）
- i：個人番号

を推定する．

➥ 「年収の対数値」を「修学年数」と「就業可能年数」と「就業可能年数の２乗」に回帰する．

- 就業可能年数
  - 最後の学校を卒業してからの年数

$$\text{就業可能年数} = \text{年齢} - \text{修学年数} - 6.$$

※ 小学校に入学する年齢が 6 歳のため，6 を引いている.

  - 熟練度を表す.
    $\Rightarrow$ 賃金に影響を与える.

$$\Downarrow$$

修学年数が年収に与える純粋な効果を計測するには，熟練度（を表す就業可能年数) をコントロールする必要がある.

モデル推定結果

gretl: モデル1

ファイル 編集(E) 検定(T) 保存(S) グラフ(G) 分析(A) LaTeX

```
モデル 1: 最小二乗法(OLS), 観測: 1-4299
従属変数: lincome

               係数          標準誤差        t値          p値
  ---------------------------------------------------------
  const        2.48550       0.110782        22.44    1.64e-105 ***
  yeduc        0.117547      0.00706026      16.65    2.31e-060 ***
  exper        0.196174      0.00749354      26.18    2.75e-140 ***
  exper2      -0.00638115    0.000316188    -20.18    1.32e-086 ***

Mean dependent var   5.290452   S.D. dependent var   0.895883
Sum squared resid    2736.906   回帰の標準誤差             0.798267
R-squared            0.206603   Adjusted R-squared   0.206049
F(3, 4295)           372.8097   P-value(F)           3.4e-215
Log-likelihood      -5129.400   Akaike criterion     10266.80
Schwarz criterion    10292.27   Hannan-Quinn         10275.79
```

出力結果の見方

- 係数: （偏）回帰係数推定値
- 標準誤差: （偏）回帰係数の標準誤差
  - 次回の授業で説明
- t 値: 「（偏）回帰係数が 0」という帰無仮説の両側 t 検定における検定統計量の実現値（t 値）
  - 次回の授業で説明
- p 値: 両側 p 値
  - 次回の授業で説明
- 回帰の標準誤差：誤差項の標準誤差
- R-squared: 決定係数
- Adjusted R-squared: 自由度修正済み決定係数

誤差項の分散の推定

定数項以外に説明変数が k 個ある重回帰モデルの場合，誤差項 u_i の（条件付き）分散

$$V(u_i \mid \boldsymbol{X}) = \sigma^2,$$

は，以下のように推定できる．

- 誤差項の分散の推定量：

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1}.$$

この s^2 は不偏性をもつ，すなわち $E(s^2) = \sigma^2$ となることが知られている．（証明は省略）

誤差項の標準偏差の推定

定数項以外に説明変数が k 個ある重回帰モデルの場合，誤差項 u_i の（条件付き）標準偏差 σ の推定値である標準誤差は，以下のように推定できる．

- 誤差項の標準偏差の推定量：

$$s = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k-1}} = \sqrt{\frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1}}.$$

この式の n と k と $e_1, e_2, \cdots, e_n$ に具体的な値を代入すれば，誤差項の標準偏差の推定値（誤差項の標準誤差）を計算できる．

自由度修正済み決定係数

- 決定係数 R^2 は説明変数の数（推定するパラメータの数）を増やすと必ず上昇する．
  ➡ 関係のない説明変数を追加しても R^2 は上昇する．
  ➡ それを回避するには，R^2 を修正する．

自由度修正済み決定係数（adjusted R-squared）は，

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-k-1}.$$

- $\bar{R}^2$ はマイナスになることがある．
- 「重回帰の場合」や「単回帰と重回帰の結果を比較する場合」は，自由度修正済み決定係数 $\bar{R}^2$ を見るのが一般的．

モデル推定結果

- 修学年数の係数
  - 0.117547（符号は正）
  - 年収と修学年数についてはログ＝レベル・モデルの関係.
    ➥ 就業可能年数（とその 2 乗）を一定としたうえで，修学年数が 1 年長くなると，年収が平均して 11.7547%高くなる傾向がある.
- 就業可能年数の係数
  - 1 乗の係数は 0.196174（符号は正）
  - 2 乗の係数は −0.00638115（符号は負）

  これらの係数の解釈は，第 9 回授業で説明する.
- 定数項
  - 2.4855（符号は正）

- 誤差項の標準誤差
  - 0.798267
- 自由度修正済み決定係数
  - $\bar{R}^2 = 0.206049.$
    ➥ 「年収」の動きの約 20.6%を「修学年数」と「就業可能年数」と「就業可能年数の 2 乗」の動きで説明できている.

今日のキーワード

重回帰モデル，偏回帰係数，他の条件を一定とする，コントロール変数，古典的線形回帰モデル，線形推定量，線形不偏推定量，最良線形不偏推定量，自由度修正済み決定係数

次回までの準備

- 今回の講義スライドを読み直す.
- 「提出課題 4」に取り組む.
- 教科書第 6 章第 4 節〜第 5 節を読む.

付録：重回帰モデルの OLS 推定量の導出

残差二乗和は,

$$
\begin{aligned}
\sum_{i=1}^{n} e_i^2 &= \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= \left[\boldsymbol{y}' - (\boldsymbol{X}\hat{\boldsymbol{\beta}})'\right](\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{y}' - \hat{\boldsymbol{\beta}}'\boldsymbol{X}')(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= \boldsymbol{y}'\boldsymbol{y} - \underbrace{\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}}_{1\times 1\text{（スカラー）}} - \underbrace{\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}}_{1\times 1\text{（スカラー）}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})' + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{y}'(\boldsymbol{X}')'(\hat{\boldsymbol{\beta}}')' + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}
\end{aligned}
$$

この残差二乗和

$$\sum_{i=1}^{n} e_i^2 = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

が最小になるように OLS 推定量を求める．
残差二乗和最小化問題は，

$$\min_{\hat{\boldsymbol{\beta}} \in \mathbb{R}^{1+k}} \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

1 階条件は，

$$\frac{d}{d\hat{\boldsymbol{\beta}}}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}.$$

参考：内積の，ベクトルでの微分

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

つまり，$\boldsymbol{a}$ と $\boldsymbol{x}$ をそれぞれ $n \times 1$ ベクトルとすると，

$$\frac{d}{d\boldsymbol{x}}\boldsymbol{a}'\boldsymbol{x} = \begin{bmatrix} \frac{\partial}{\partial x_1}\boldsymbol{a}'\boldsymbol{x} \\ \frac{\partial}{\partial x_2}\boldsymbol{a}'\boldsymbol{x} \\ \vdots \\ \frac{\partial}{\partial x_n}\boldsymbol{a}'\boldsymbol{x} \end{bmatrix} = \boldsymbol{a}.$$

（証明）

$$\boldsymbol{a}'\boldsymbol{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

なので,

$$\frac{d}{d\boldsymbol{x}}\boldsymbol{a}'\boldsymbol{x} = \begin{bmatrix} \frac{\partial}{\partial x_1}\boldsymbol{a}'\boldsymbol{x} \\ \frac{\partial}{\partial x_2}\boldsymbol{a}'\boldsymbol{x} \\ \vdots \\ \frac{\partial}{\partial x_n}\boldsymbol{a}'\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\ \frac{\partial}{\partial x_2}(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\ \vdots \\ \frac{\partial}{\partial x_n}(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \boldsymbol{a}. \quad \text{（証明終）}$$

参考：2 次形式の，ベクトルでの微分

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix},$$

つまり，$\boldsymbol{x}$ を $n \times 1$ ベクトル，$\boldsymbol{A}$ を $n \times n$ 対称行列とすると，

$$\frac{d}{d\boldsymbol{x}}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} \frac{\partial}{\partial x_1}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \\ \frac{\partial}{\partial x_2}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \\ \vdots \\ \frac{\partial}{\partial x_n}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \end{bmatrix} = 2\boldsymbol{A}\boldsymbol{x}.$$

※ $\boldsymbol{A}$ が対称でなければ，$\frac{d}{d\boldsymbol{x}}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}$.

（証明【A が対称の場合】）簡単化のため，

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix},$$

とすると，

$$\begin{aligned} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 & a_{12}x_1 + a_{22}x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= (a_{11}x_1 + a_{12}x_2)x_1 + (a_{12}x_1 + a_{22}x_2)x_2 \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{12}x_1x_2 + a_{22}x_2^2 \\ &= a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2, \end{aligned}$$

なので，

$$
\begin{aligned}
\frac{d}{d\boldsymbol{x}}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} \frac{\partial}{\partial x_1}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \\ \frac{\partial}{\partial x_2}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2) \\ \frac{\partial}{\partial x_2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2) \end{bmatrix} \\
&= \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 \\ 2a_{12}x_1 + 2a_{22}x_2 \end{bmatrix} = \begin{bmatrix} 2(a_{11}x_1 + a_{12}x_2) \\ 2(a_{12}x_1 + a_{22}x_2) \end{bmatrix} \\
&= 2\begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{12}x_1 + a_{22}x_2 \end{bmatrix} \\
&= 2\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= 2\boldsymbol{A}\boldsymbol{x}. \quad \text{（証明終）}
\end{aligned}
$$

残差二乗和最小化問題の 1 階条件は，

$$
\begin{aligned}
&\frac{d}{d\hat{\boldsymbol{\beta}}}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{0} \\
\Leftrightarrow\ &\underbrace{\frac{d}{d\hat{\boldsymbol{\beta}}}\boldsymbol{y}'\boldsymbol{y}}_{=\boldsymbol{0}} + \frac{d}{d\hat{\boldsymbol{\beta}}}(-2\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}) + \frac{d}{d\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0} \\
\Leftrightarrow\ &-2\cdot\frac{d}{d\hat{\boldsymbol{\beta}}}\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \frac{d}{d\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}.
\end{aligned}
$$

- $\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$ の中で，
  - $\boldsymbol{y}'\boldsymbol{X}$（$1 \times (1+k)$ ベクトル）が $\boldsymbol{a}'\boldsymbol{x}$ の $\boldsymbol{a}'$ に相当
  - $\hat{\boldsymbol{\beta}}$（$(1+k) \times 1$ ベクトル）が $\boldsymbol{a}'\boldsymbol{x}$ の $\boldsymbol{x}$ に相当
- $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}$ の中で，
  - $\hat{\boldsymbol{\beta}}'$（$1 \times (1+k)$ ベクトル）が $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ の $\boldsymbol{x}'$ に相当
  - $\boldsymbol{X}'\boldsymbol{X}$（$(1+k) \times (1+k)$ 対称行列）が $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ の $\boldsymbol{A}$ に相当
  - $\hat{\boldsymbol{\beta}}$（$(1+k) \times 1$ ベクトル）が $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ の $\boldsymbol{x}$ に相当

のように考えれば，

$$
\begin{aligned}
& -2 \cdot \frac{d}{d\hat{\boldsymbol{\beta}}} \boldsymbol{y}' \boldsymbol{X} \hat{\boldsymbol{\beta}} + \frac{d}{d\hat{\boldsymbol{\beta}}} \hat{\boldsymbol{\beta}}' \boldsymbol{X}' \boldsymbol{X} \hat{\boldsymbol{\beta}} = \mathbf{0} \\
\Leftrightarrow & -2(\boldsymbol{y}'\boldsymbol{X})' + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \\
\Leftrightarrow & -2\boldsymbol{X}'(\boldsymbol{y}')' + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \\
\Leftrightarrow & 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = 2\boldsymbol{X}'\boldsymbol{y} \\
\Leftrightarrow & \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y} \\
\Leftrightarrow & \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}}_{=\boldsymbol{I}_{1+k}} \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
\Leftrightarrow & \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.
\end{aligned}
$$