OLS 係数についての補足

佐藤

2019 年 7 月 10 日

$Y \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$, $X^{n\times p} \in \mathbb{R}$ とする[1]。n は観測の数，p は定数項を含めた説明変数の数である。線形回帰モデル

$$Y = X\beta + u$$

の回帰係数を最小 2 乗法で計算したものを $\hat{\beta} \in \mathbb{R}^p$ とする。係数が β であるときの Y の予測を

$$\hat{Y}\left(\beta\right) = X\beta$$

と書けば，予測誤差

$$e\left(\beta\right) = Y - \hat{Y}\left(\beta\right) \in \mathbb{R}^n$$

のユークリッドノルムの 2 乗（成分ごとの誤差の 2 乗和）を最小化するように $\beta = \hat{\beta}$ が定まっている。

$$\min \|e\left(\beta\right)\|^2$$

の必要十分条件は

$$\frac{\partial \|e\left(\beta\right)\|^2}{\partial \beta_0} = 0, \quad \frac{\partial \|e\left(\beta\right)\|^2}{\partial \beta_1} = 0, \ldots, \quad \frac{\partial \|e\left(\beta\right)\|^2}{\partial \beta_{p-1}} = 0 \tag{1}$$

であるから，連立方程式 (1) を解けば $\hat{\beta}$ が求まる。

ノルムの計算

式 (1) のような連立方程式を

$$\frac{\partial \|e\left(\beta\right)\|^2}{\partial \beta} = 0$$

と書くことにしよう。もちろん，β はベクトルなので文字通り偏微分を計算するわけではなく，成分ごとに偏微分した結果を並べたものと理解する。すなわち，

$$\begin{bmatrix} \frac{\partial \|e(\beta)\|^2}{\partial \beta_0} \\ \frac{\partial \|e(\beta)\|^2}{\partial \beta_1} \\ \vdots \\ \frac{\partial \|e(\beta)\|^2}{\partial \beta_{p-1}} \end{bmatrix} = 0$$

実は，$\|e\left(\beta\right)\|^2$ は β の 2 次式なので左辺の偏微分ベクトルを行列で書くことができる。

$$\begin{aligned} \|e\left(\beta\right)\|^2 &= e\left(\beta\right)^\top e\left(\beta\right) \\ &= \left(Y - X\beta\right)^\top \left(Y - X\beta\right) \\ &= Y^\top Y - \beta^\top X^\top Y - Y^\top X\beta + \beta^\top X^\top X\beta \end{aligned}$$

[1]ベクトル・行列であっても太字で書かないので注意する

ここで，最右辺の各項がスカラーであることに注意すると，

$$\beta^\top X^\top Y = \left(\beta^\top X^\top Y\right)^\top = Y^\top X\beta$$

であることに気がつく。したがって，

$$\|e\left(\beta\right)\|^2 = Y^\top Y - 2\beta^\top X^\top Y + \beta^\top X^\top X\beta \tag{2}$$

である。

ベクトルによる微分

式 (2) で β に関係があるのは

$$\beta^\top X^\top Y, \qquad \beta^\top X^\top X\beta$$

の 2 つであるから，これらの項の微分がわかればよい。（$Y^\top Y$ は微分するとゼロになる）

$X^\top Y \in \mathbb{R}^p, X^\top X \in \mathbb{R}^{p\times p}$ であることに注意しておこう。

定理 1. $\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^p$ とする。このとき

$$\frac{\partial}{\partial\beta}\left[\beta^\top\alpha\right] = \alpha.$$

証明. $\beta^\top\alpha = \sum_{i=0}^{p-1}\beta_i\alpha_i$ なので，$\frac{\partial}{\partial\beta_i}\left[\beta^\top\alpha\right] = \alpha_i, i = 0, 1, \ldots, p-1$。したがって，

$$\frac{\partial}{\partial\beta}\left[\beta^\top\alpha\right] = \begin{bmatrix} \frac{\partial}{\partial\beta_0}\left[\beta^\top\alpha\right] \\ \frac{\partial}{\partial\beta_1}\left[\beta^\top\alpha\right] \\ \vdots \\ \frac{\partial}{\partial\beta_{p-1}}\left[\beta^\top\alpha\right] \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{p-1} \end{bmatrix} = \alpha.$$

□

定理 2. $A \in \mathbb{R}^{p\times p}, \beta \in \mathbb{R}^p$ とする。このとき

$$\frac{\partial}{\partial\beta}\left[\beta^\top A\beta\right] = \left(A + A^\top\right)\beta.$$

証明. $\left[A\beta\right]_j$ を $A\beta \in \mathbb{R}^p$ の第 j 成分，A_{jk} を行列 A の第 (j, k) 成分とすると，

$$\begin{aligned} \beta^\top A\beta = \sum_{j=0}^{p-1}\beta_j\left[A\beta\right]_j &= \\ &= \sum_{j=0}^{p-1}\beta_j\left[\sum_{k=0}^{p-1}A_{jk}\beta_k\right] \\ &= \sum_{j=0}^{p-1}\sum_{k=0}^{p-1}A_{jk}\beta_j\beta_k \end{aligned}$$

と書ける。したがって，

$$
\begin{aligned}
\frac{\partial}{\partial \beta_i}\left[\beta^\top A \beta\right] &= \frac{\partial}{\partial \beta_i}\left[\sum_{j=0}^{p-1}\sum_{k=0}^{p-1} A_{jk}\beta_j\beta_k\right] \\
&= \sum_{k=0}^{p-1} A_{ik}\beta_k + \sum_{j=0}^{p-1} A_{ji}\beta_j \\
&= \sum_{j=0}^{p-1} A_{ij}\beta_j + \sum_{j=0}^{p-1} A_{ji}\beta_j \\
&= A\beta + A^\top \beta \\
&= \left(A + A^\top\right)\beta
\end{aligned}
$$

□

系 3. $A \in \mathbb{R}^{p\times p}$, $\beta \in \mathbb{R}^p$, A は対称行列とする。このとき，

$$
\frac{\partial}{\partial \beta}\left[\beta^\top A \beta\right] = 2A\beta.
$$

証明. $A^\top = A$ と定理 2 より直ちに示せる。 □

正規方程式の解

上で示した定理，系を式 2 に適用しよう。偏微分がゼロになる条件は

$$
\begin{aligned}
\frac{\partial}{\partial \beta}\|e\left(\beta\right)\|^2 &= \frac{\partial}{\partial \beta}\left[Y^\top Y - 2\beta^\top X^\top Y + \beta^\top X^\top X \beta\right] \\
&= -2X^\top Y + 2X^\top X\beta \\
&= 0
\end{aligned}
$$

である。これを β について解けばよい。解 $\hat{\beta}$ は次の公式によって定まる。

$$
\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y
$$