LEAST-SQUARES ESTIMATORS IN LINEAR MODELS

DAVAR KHOSHNEVISAN

1. The General Linear Model

Let Y be the response variable and $X_1, \dots, X_m$ be the explanatory variables. The following is the linear model of interest to us:

$$Y = \beta_1 X_1 + \cdots + \beta_m X_m + \varepsilon, \tag{1}$$

where $\beta_1, \dots, \beta_m$ are unknown parameters, and ε is "noise."

Now we take a sample $Y_1, \dots, Y_n$. The linear model becomes

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_m X_{im} + \varepsilon_i \qquad i = 1, \dots n. \tag{2}$$

Define

$$\boldsymbol{X} = \begin{pmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}. \tag{3}$$

Note that

$$\boldsymbol{X}\boldsymbol{\beta} = \begin{pmatrix} \beta_1 X_{11} + \cdots + \beta_m X_{1m} \\ \vdots \\ \beta_m X_{n1} + \cdots + \beta_m X_{nm} \end{pmatrix}. \tag{4}$$

Therefore, the linear model (2) can be written more neatly as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$. To be sure, $\boldsymbol{Y}$ is $n \times 1$, $\boldsymbol{X}$ is $n \times m$, $\boldsymbol{\beta}$ is $m \times 1$, and $\boldsymbol{\varepsilon}$ is $n \times 1$.

The matrix $\boldsymbol{X}$ is treated as if it were non-random; it is called the "design matrix" or the "regression matrix."

2. Least Squares

Let $\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta}$, and minimize, over all $\boldsymbol{\beta}$, the following quantity:

$$\|\boldsymbol{Y} - \boldsymbol{\theta}\|^2 = (\boldsymbol{Y} - \boldsymbol{\theta})'(\boldsymbol{Y} - \boldsymbol{\theta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{i=1}^n \varepsilon_i^2. \tag{6}$$

Note that

$$\boldsymbol{\theta} = \beta_1 \boldsymbol{X}_1 + \cdots + \beta_m \boldsymbol{X}_m, \tag{7}$$

Date: August 30, 2004.

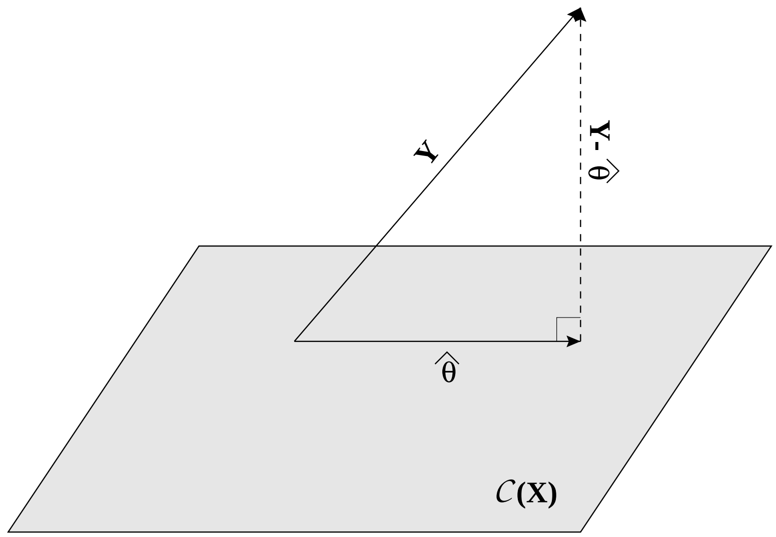


FIGURE 1. The projection $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{Y}$ onto the subspace $\mathscr{C}(\boldsymbol{X})$

where $\boldsymbol{X}_i$ denotes the ith column of the matrix $\boldsymbol{X}$. That is, $\boldsymbol{\theta} \in \mathscr{C}(\boldsymbol{X})$—the column space of $\boldsymbol{X}$. So our problem has become: Minimize $\|\boldsymbol{Y} - \boldsymbol{\theta}\|$ over all $\boldsymbol{\theta} \in \mathscr{C}(\boldsymbol{X})$.

A look at Figure 1 will convince you that the closest point $\widehat{\boldsymbol{\theta}} \in \mathscr{C}(\boldsymbol{X})$ is the projection of $\boldsymbol{Y}$ onto the subspace $\mathscr{C}(\boldsymbol{X})$. To find a formula for this projection we first work more generally.

3. Some Geometry

Let S be a subspace of $\boldsymbol{R}^n$. Recall that this means that:

(1) If $x, y \in S$ and $\alpha, \beta \in \boldsymbol{R}$, then $\alpha x + \beta y \in S$; and
(2) $\mathbf{0} \in S$.

Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ forms a basis for S; that is, any $x \in S$ can be represented as a linear combination of the $\boldsymbol{v}_i$'s. Define $\boldsymbol{V}$ to be the matrix whose ith column is $\boldsymbol{v}_i$; that is,

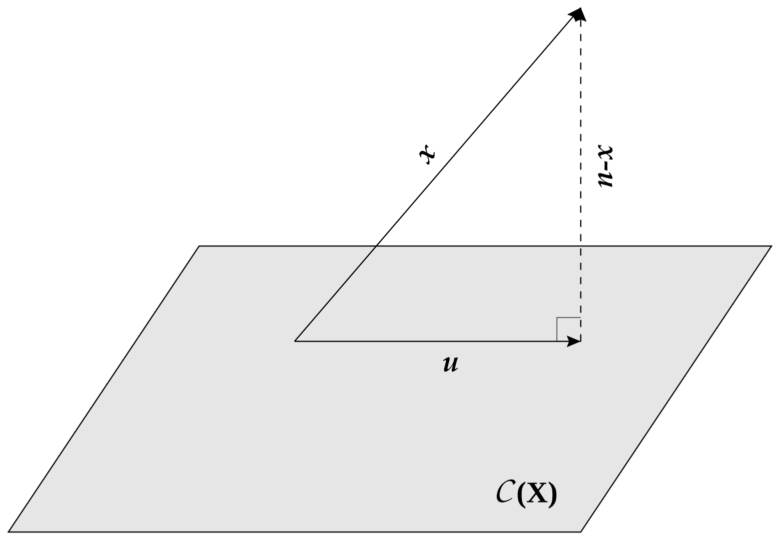
$$\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k]. \tag{8}$$

Then any $\boldsymbol{x} \in \boldsymbol{R}^n$ is *orthogonal* to S if and only if $\boldsymbol{x}$ is orthogonal to every $\boldsymbol{v}_i$; that is, $\boldsymbol{x}'\boldsymbol{v}_i = 0$. Equivalently, $\boldsymbol{x}$ is orthogonal to S if and only if $\boldsymbol{x}'\boldsymbol{V} = \mathbf{0}$. In summary,

$$\boldsymbol{x} \perp S \iff \boldsymbol{x}'\boldsymbol{V} = \mathbf{0}. \tag{9}$$

Now the question is: If $\boldsymbol{x} \in \boldsymbol{R}^n$ then how can we find its projection $\boldsymbol{u}$ onto S? Consider Figure 3. From this it follwos that $\boldsymbol{u}$ has two properties.

(1) First of all, $\boldsymbol{u}$ is perpendicular S, so that $(\boldsymbol{u} - \boldsymbol{x})'\boldsymbol{V} = 0$. Equivalently, $\boldsymbol{u}'\boldsymbol{V} = \boldsymbol{x}'\boldsymbol{V}$.



(2) Secondly, $\boldsymbol{u} \in S$, so there exist $\alpha_1, \ldots, \alpha_k \in \boldsymbol{R}^k$ such that $\boldsymbol{u} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_k \boldsymbol{v}_k$. Equivalently,

$$\boldsymbol{u} = \boldsymbol{V}\boldsymbol{\alpha}. \tag{10}$$

Plug (2) into (1) to find that $\boldsymbol{\alpha}'\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{x}'\boldsymbol{V}$. Therefore, if $\boldsymbol{V}'\boldsymbol{V}$ is invertible, then $\boldsymbol{\alpha}' = (\boldsymbol{x}'\boldsymbol{V})(\boldsymbol{V}'\boldsymbol{V})^{-1}$. Equivalently,

$$\boldsymbol{u} = \boldsymbol{V}(\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}\boldsymbol{x}. \tag{11}$$

Define

$$\mathbf{P}_S = \boldsymbol{V}(\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'. \tag{12}$$

Then, $\boldsymbol{u} = \mathbf{P}_S\boldsymbol{x}$ is the projection of $\boldsymbol{x}$ onto the subspace S.

Note that $\mathbf{P}_S$ is *idempotent* (i.e., $\mathbf{P}_S^2 = \mathbf{P}_S$) and *symmetric* ($\mathbf{P}_S = \mathbf{P}_S'$).

4. Application to Linear Models

Let $S = \mathscr{C}(\boldsymbol{X})$ be the subspace spanned by the columns of $\boldsymbol{X}$—this is the *column space* of $\boldsymbol{X}$. Then, provided that $\boldsymbol{X}'\boldsymbol{X}$ is invertible,

$$\mathbf{P}_{\mathscr{C}(\boldsymbol{X})} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'. \tag{13}$$

Therefore, the LSE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \mathbf{P}_{\mathscr{C}(\boldsymbol{X})}\boldsymbol{Y} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{14}$$

This is equal to $\boldsymbol{X}'\widehat{\boldsymbol{\beta}}$. So $\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y}$. Equivalently,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{15}$$