Least squares estimates of the regression parameters

2018-01-16

Warmup

Recall from last time we can set up a multiple linear regression model in the matrix form:

$$y = X\beta + \epsilon$$

Give the name and dimensions of each term

Solution: $y_{n\times 1}$ the response vector, $X_{n\times p}$ the design matrix, $\beta_{p\times 1}$ the parameter vector, $\epsilon_{n\times 1}$ the error vector

Today's goal

Derive the form of the estimates for the parameter vector β.

Least Squares

Just like in simple linear regression, we'll estimate β by **least squares**. In simple linear regression this involved finding $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimise the sum of squared residuals:

$$\text{sum of squared residuals SLR} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

Your turn: What procedure do you use to minimise a function? E.g. if $f(x)$ is a function of a single real value x, how do you find the x that minimises $f(x)$?

(*2 min discussion*)

Solution: 1. Find derivative of $f(x)$ with respect to x, $\frac{\partial f(x)}{\partial x}$.

2. Set the derivative to zero, and solve for x.

3. Check second derivative at minimum is positive, i.e. that this is a minimum.

For multiple linear regression the least squares estimate of the β is the **vector** $\hat{\beta}$ that minimizes the sum of squared residuals:

$$\text{sum of squared residuals MLR} = \sum_{i=1}^{n} e_i^2 = ||e||^2 = \left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right)$$

Your turn Expand the matrix product on the right into four terms. Be careful with the order of matrix multiplication and recall $(uX)^T = X^T u^T$.

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

Solution:

$$\sum_{i=1}^{n} e_i^2 = ||e||^2 = y^T y - \hat{\beta}^T X^T y - y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta}$$

Consider the terms:

$$-\hat{\beta}^T X^T y \quad \text{and} - y^T X\hat{\beta}$$

Argue that these can be combined into the single term

$$-2\hat{\beta}^T X^T y$$

(Hint: consider the dimensions of these terms)

Solution: Note the the dimensions of the terms

$$y^T_{1\times n} X_{n\times n} \hat{\beta}_{p\times 1}$$

so the result is 1×1, a scalar, and therefore

$$y^T X\hat{\beta} = \left(y^T X\hat{\beta}\right)^T = \hat{\beta}^T X^T y$$

Finding the minimum

Now our objective is to find $\hat{\beta}$ that minimises:

$$y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}$$

The usual procedure would be to take derivative with respect to $\hat{\beta}$, set to zero and solve for $\hat{\beta}$. **Except** $\hat{\beta}$ is a **vector**! We need to use vector calculus.

Vector calculus

You should be familiar with the usual differentiation rules for scalar a and x:

- $\frac{\partial}{\partial x} a = 0$
- $\frac{\partial}{\partial x} ax = a$
- $\frac{\partial}{\partial x} ax^2 = 2ax$

There are analogs when we want to take derivative with respect to a vector $\mathbf{x}$:

- $\frac{\partial}{\partial \mathbf{x}} a = 0$, where a is a scalar
- $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T u = u$, where u is a vector
- $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A \mathbf{x} = (A + A^T)\mathbf{x}$, where A is a matrix

Use the rules above to take the derivative of the sum of squared residuals with respect to the vector $\hat{\beta}$

$$\frac{\partial}{\partial \hat{\beta}} \left(y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} \right) =$$

Solution:

$$\begin{aligned} \frac{\partial}{\partial \hat{\beta}} \left(y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} \right) &= 0 - 2X^T y + (X^T X + (X^T X)^T)\hat{\beta} \\ &= -2X^T y + (X^T X + X^T X)\hat{\beta} \\ &= -2X^T y + 2X^T X \hat{\beta} \end{aligned}$$

Normal Equations

Setting the above derivative to zero leads to the **Normal Equations**. The least squares estimates satisfy:

$$X^T y = X^T X \hat{\beta}$$

If $X^T X$ is invertible, the least squares estimates are (**fill me in**):

$$\hat{\beta} = (\qquad)^{-1} \quad {}^T y$$

If X has rank p then $X^T X$ will be invertible.

Fitted Values and Residuals

Plug in the least squares estimate for $\hat{\beta}$ to find the fitted values and residuals

$$\hat{y} = X\hat{\beta} =$$

$$\hat{\epsilon} = e = y - X\hat{\beta} =$$

Solution:

$$
\begin{aligned}
\hat{y} &= X\hat{\beta} \\
&= X(X^TX)^{-1}X^Ty \\
\hat{\epsilon} = e &= y - X\hat{\beta} \\
&= y - X(X^TX)^{-1}X^Ty \\
&= \left(I - X(X^TX)^{-1}X^T\right)y
\end{aligned}
$$

Hat matrix

The hat matrix is:

$$H = X\left(X^TX\right)^{-1}X^T$$

named because it puts "hats" on the response, i.e. multiplying the response by the hat matrix gives the fitted values:

$$Hy = \hat{y}$$

Your Turn: Show $(I - H)\,X = \mathbf{0}$

Solution:

$$
\begin{aligned}
(I - H)\,X &= \left(I - X\left(X^TX\right)X^T\right)X \quad \text{expand } H \\
&= X - X\left(X^TX\right)^{-1}X^TX \quad \text{distribute } X \\
&= X - XI \quad \text{since } A^{-1}A = I \\
&= 0
\end{aligned}
$$

Other properties of H

- H is symmetric, so is $(I - H)$
- H is idempotent ($H^2 = H$), and so is $I - H$
- X is invariant under H (i.e. $HX = X$)
- $(I - H)H = H(I - H) = 0$

You can use these results to argue that the residuals are orthogonal to the columns of X, i.e. show $e^TX = \mathbf{0}$

$$
\begin{aligned}
e^TX &= ((I - H)Y)^TX \quad \text{plug in form for residuals} \\
&= Y^T(I - H)^TX \quad \text{distribute transpose} \\
&= Y^T(I - H)X \quad \text{symmetry} \\
&= Y^T0 \quad \text{from above} \\
&= 0
\end{aligned}
$$

Next time

What are the properties of the least squares estimates?