Least squares estimates of the regression parameters

2019-01-16

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

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*)

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 $(uX)^T = X^T u^T$.

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

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)

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 $\hat{\beta}$

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

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

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

Hat matrix

The hat matrix is:

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

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 $(I - H)\, X = \mathbf{0}$

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 $e^T X = \mathbf{0}$

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

Next time

What are the properties of the least squares estimates?