Statistical modelling

#2.c Geometry of least squares

Dr. Léo Belzile HEC Montréal

Linear algebra reminders

For an n imes p matrix, the column space of ${f X}$ is

$$\mathcal{S}(\mathbf{X}) = \{\mathbf{X}oldsymbol{a}, oldsymbol{a} \in \mathbb{R}^p\}$$

The linear model equation

$$oldsymbol{Y} = \mathbf{X}oldsymbol{eta} + oldsymbol{arepsilon}$$

corresponds to an (unknown) element of the span of ${f X}$ plus a disturbance.

Ordinary least squares

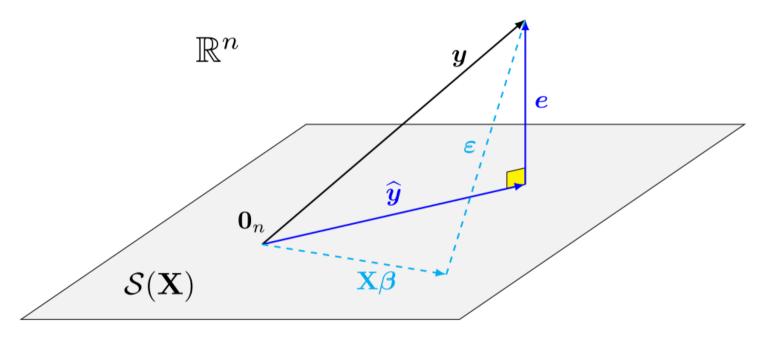
Find the element of $\mathcal{S}(\mathbf{X})$ with the minimum distance from \boldsymbol{y} (ordinary least squares), i.e.

$$oldsymbol{\widehat{oldsymbol{eta}}} = \min_{oldsymbol{eta} \in \mathbb{R}^p} \|oldsymbol{y} - \mathbf{X}oldsymbol{eta}\|^2$$

Intuition: $\varepsilon_1, \ldots, \varepsilon_n$ and $\beta_0, \ldots, \beta_{p-1}$ are unknown, but we cannot retrieve them (n observations, n+p unknowns).

Column geometry

We try to find the best p-dimensional approximation onto $\mathcal{S}(\mathbf{X})$.



The solution to the least square problem is the projection of y onto S(X), i.e., Hy, where $H = X(X^{\top}X)^{-1}X^{\top}$.

Orthogonal decomposition

Write

$$egin{aligned} oldsymbol{y} &= \mathbf{H} oldsymbol{y} + (\mathbf{I}_n - \mathbf{H}) oldsymbol{y} \ &= \mathbf{X} \widehat{oldsymbol{eta}} + oldsymbol{e} \end{aligned}$$

The residuals $m{e}$ are orthogonal to the columns of $m{X}$ and the fitted values $\widehat{m{y}} = m{X} \widehat{m{eta}}$.

lacktriangle By Pythagoras' theorem, $\|oldsymbol{y}\|^2 = \|\widehat{oldsymbol{y}}\|^2 + \|oldsymbol{e}\|^2$.

Assuming $\mathbf{1}_n \in \mathcal{S}(\mathbf{X})$ (intercept included)

- lacktriangle The sample mean of $oldsymbol{e}$ is zero.
- lacktriangle A linear regression of $\widehat{m{y}}$ onto $m{e}$ has zero intercept and slope (they are uncorrelated).