

Statistical modelling

#2.c Geometry of least squares

Dr. Léo Belzile
HEC Montréal

Linear algebra reminders

For an $n \times p$ matrix, the column space of \mathbf{X} is

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

The linear model equation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

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

Ordinary least squares

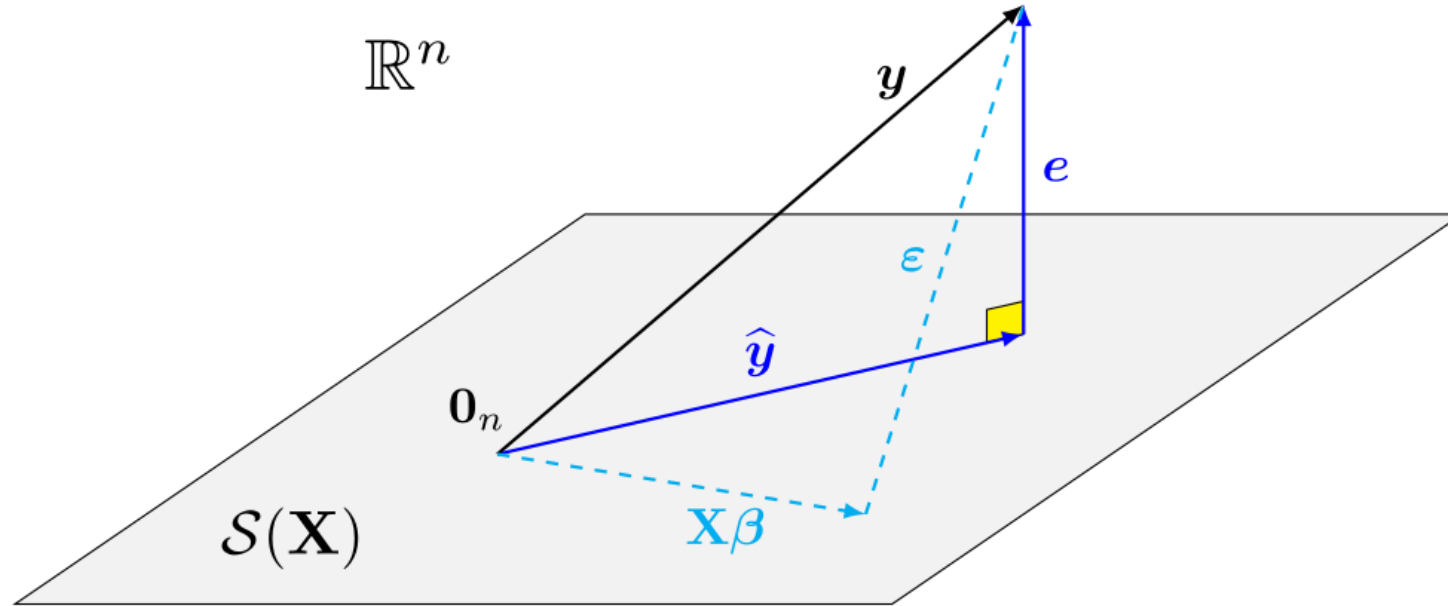
Find the element of $\mathcal{S}(\mathbf{X})$ with the minimum distance from \mathbf{y} (ordinary least squares), i.e.

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

Intuition: $\varepsilon_1, \dots, \varepsilon_n$ and $\beta_0, \dots, \beta_{p-1}$ 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 \mathbf{y} onto $\mathcal{S}(\mathbf{X})$, i.e., $\mathbf{H}\mathbf{y}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$.

Orthogonal decomposition

Write

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\mathbf{y} + (\mathbf{I}_n - \mathbf{H})\mathbf{y} \\ &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}\end{aligned}$$

The residuals \mathbf{e} are orthogonal to the columns of \mathbf{X} and the fitted values $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

✚ By Pythagoras' theorem, $\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\mathbf{e}\|^2$.

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

✚ The sample mean of \mathbf{e} is zero.

✚ A linear regression of $\hat{\mathbf{y}}$ onto \mathbf{e} has zero intercept and slope (they are uncorrelated).