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

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

The linear model equation

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

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

Ordinary least squares

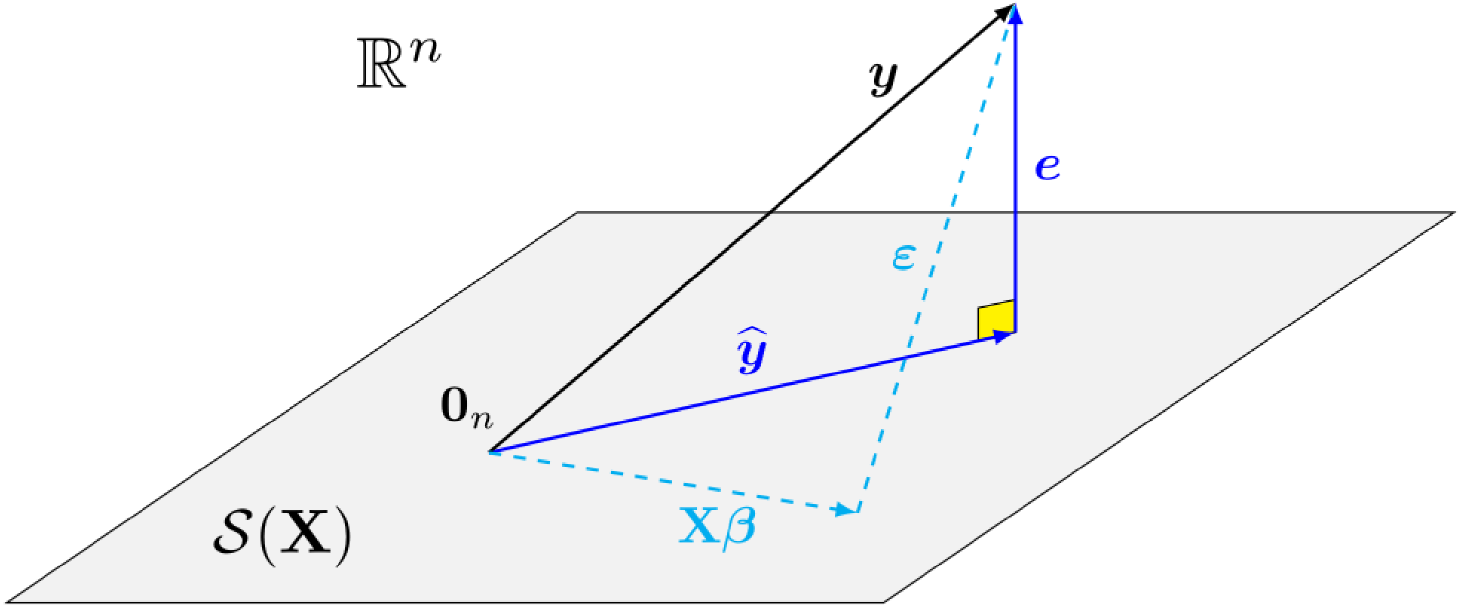
Find the element of $\mathcal{S}(\mathbf{X})$ with the minimum distance from $\boldsymbol{y}$ (ordinary least squares), i.e.

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

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 $\boldsymbol{p}$-dimensional approximation onto $\mathcal{S}(\mathbf{X})$.



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

Orthogonal decomposition

Write

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

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

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

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

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