Statistical Learning II

Lecture 7 - Bias-Variance decomposition (Continued)

Bruno Loureiro
@ CSD, DI-ENS & CNRS

brloureiro@gmail.com

DL3 IASO, Université Paris Dauphine-PSL
05.11.2025

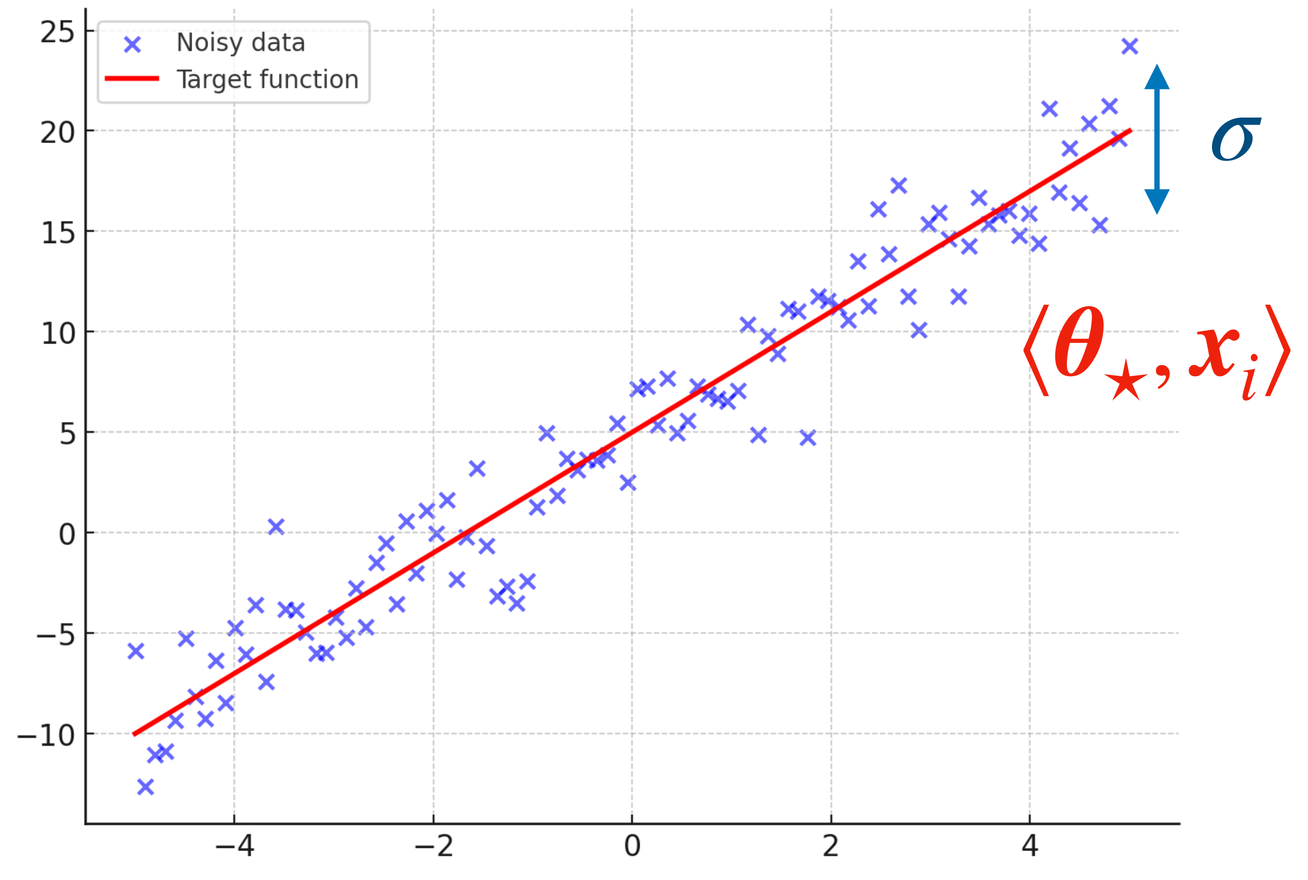
Assumptions

We now assume the following data generative model:

$$y_i = \langle \boldsymbol{\theta}_\star, \boldsymbol{x}_i \rangle + \varepsilon_i$$

With:

- Fixed $\boldsymbol{\theta}_\star \in \mathbb{R}^d$ and $\boldsymbol{x}_i \in \mathbb{R}^d$ "fixed design"
- $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i^2] = \sigma^2 < \infty$



Decomposition of OLS

$$\hat{\boldsymbol{\theta}}_{OLS}(\boldsymbol{X}, \boldsymbol{y}) = \underbrace{\boldsymbol{\theta}_\star}_{\text{“signal”}} + \underbrace{\frac{1}{n}\hat{\boldsymbol{\Sigma}}_n^{-1}\boldsymbol{X}^\top\boldsymbol{\varepsilon}}_{\text{“noise”}}$$

In particular:

- Bias: $\mathbb{E}_{\boldsymbol{\varepsilon}}\left[\hat{\boldsymbol{\theta}}_{OLS}(\boldsymbol{X}, \boldsymbol{y})\right] = \boldsymbol{\theta}_\star$ “Unbiased”
- Variance: $\mathrm{Var}_{\boldsymbol{\varepsilon}}\left[\hat{\boldsymbol{\theta}}_{OLS}(\boldsymbol{X}, \boldsymbol{y})\right] = \frac{\sigma^2}{n}\hat{\boldsymbol{\Sigma}}_n^{-1}$

Informally, if $\hat{\boldsymbol{\Sigma}}_n \to \boldsymbol{\Sigma}$ a rank d matrix as $n \to \infty$, then:

$\hat{\boldsymbol{\theta}}_{OLS} \to \boldsymbol{\theta}_\star$ as $n \to \infty$ “Consistency”

Risk of OLS

Therefore, we have the following final result for the excess risk of OLS

$$\mathbb{E}_{\boldsymbol{\varepsilon}}\left[\mathscr{R}(\hat{\boldsymbol{\theta}}_{OLS})\right] - \sigma^2 = \sigma^2 \frac{d}{n}$$

Remarks:

- Excess risk is proportional to the noise level $\mathbb{E}[\varepsilon^2] = \sigma^2$.
- Excess risk is proportional to the data dimension.
- To achieve excess risk $\Delta\mathscr{R} < \delta$, need:

$$n > \frac{\sigma^2 d}{\delta}$$

samples.