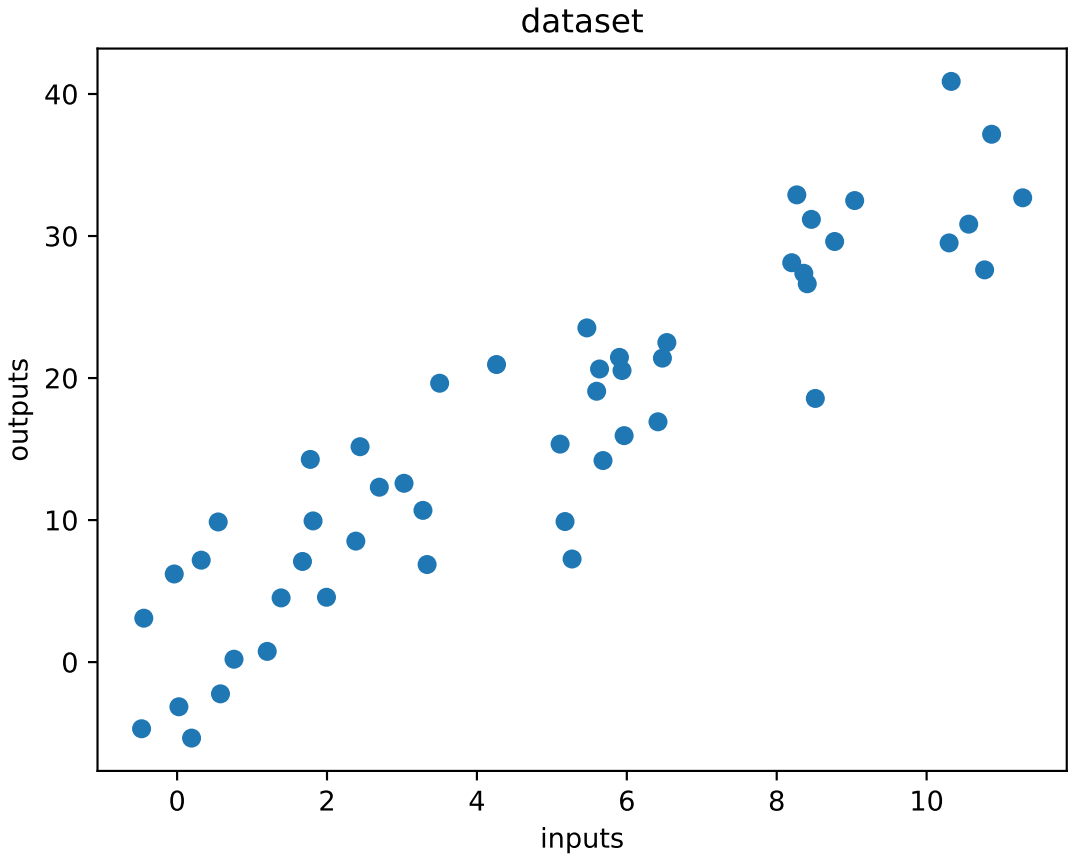
Machine learning I, supervised learning: linear regression



Linear regression

We consider a regression problem in one dimension.

- $\mathcal{X} = \mathbb{R}$
- $\mathcal{Y} = \mathbb{R}$
- $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$.

Our estimator writes :

$$h(x) = \theta x + b \tag{1}$$

Empirical risk

With the square loss, the empirical risk writes

$$R_n(\theta) = \sum_{i=1}^{n} (\theta x_i + b - y_i)^2 \tag{2}$$

Exercice 1 : Compute the empirical risk minimizer

Generalization

- Ordinary least square
- Ridge regression
- Cross-validation