

Machine Learning

Model Selection and Validation

Fabio Vandin

November 18th, 2022

Model Selection

When we have to solve a machine learning task:

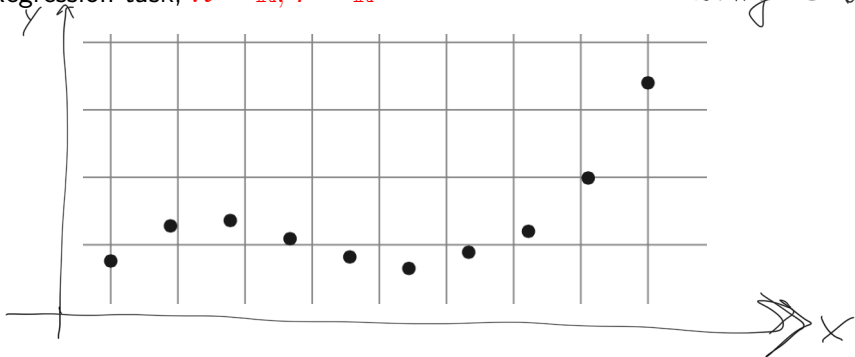
- there are different algorithms/classes
- algorithms have parameters

Question: how do we choose a algorithm or value of the parameters?

Example

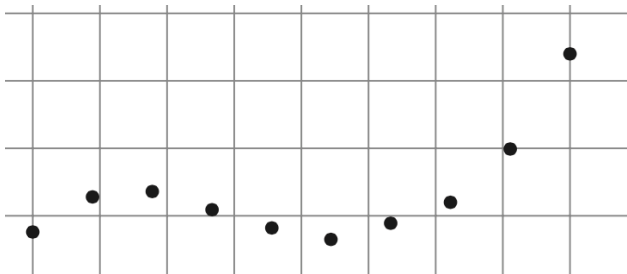
Regression task, $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$

training data



Example

Regression task, $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}$



Decision: $\mathcal{H} = \text{polynomials}$.

Note: can be done using the linear regression machinery we have seen!

How do we pick the degree d of the polynomial?

What about considering the empirical risk of best hypothesis of various degrees (e.g., $d=2, 3, 10$)?