Machine Learning

Model Selection and Validation

Fabio Vandin   November 10th, 2023

Model Selection

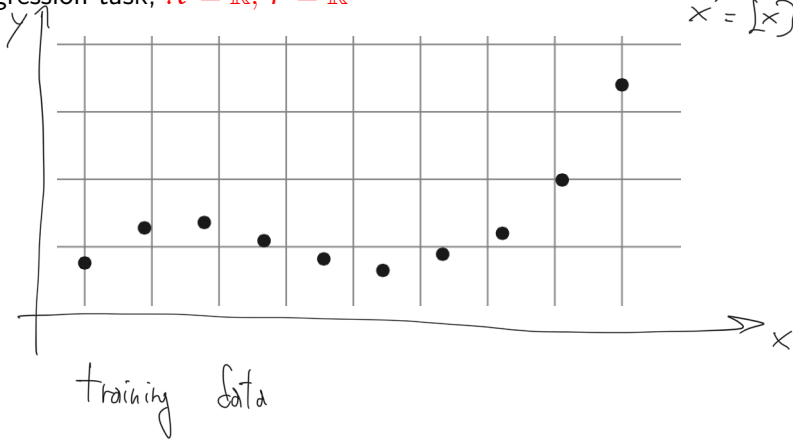When we have to solve a machine learning task:

- there are different algorithms/classes
- algorithms have parameters

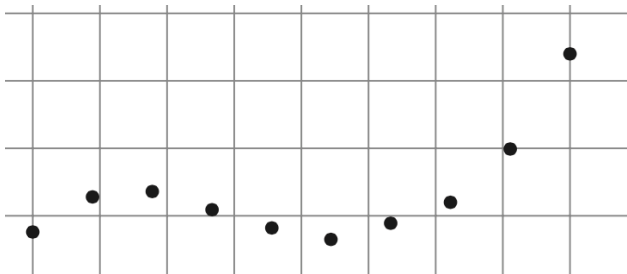**Question:** how do we choose a algorithm or value of the parameters?

Example

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

$\vec{x} \in \mathcal{X}$

$\vec{x} = [x]$

training data

Example

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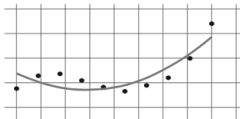

Decision: $\mathcal{H} =$ polynomials.

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

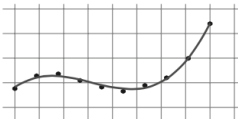How do we pick the degree d of the polynomial?

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

Best hypotheses for degree $d \in \{2, 3, 10\}$
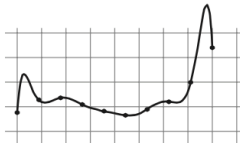

Degree 2   Degree 3   Degree 10

Empirical risk is not enough!

Approach we will consider: validation!