Outline

- What is overfitting?
- The role of noise
- Deterministic noise
- Dealing with overfitting

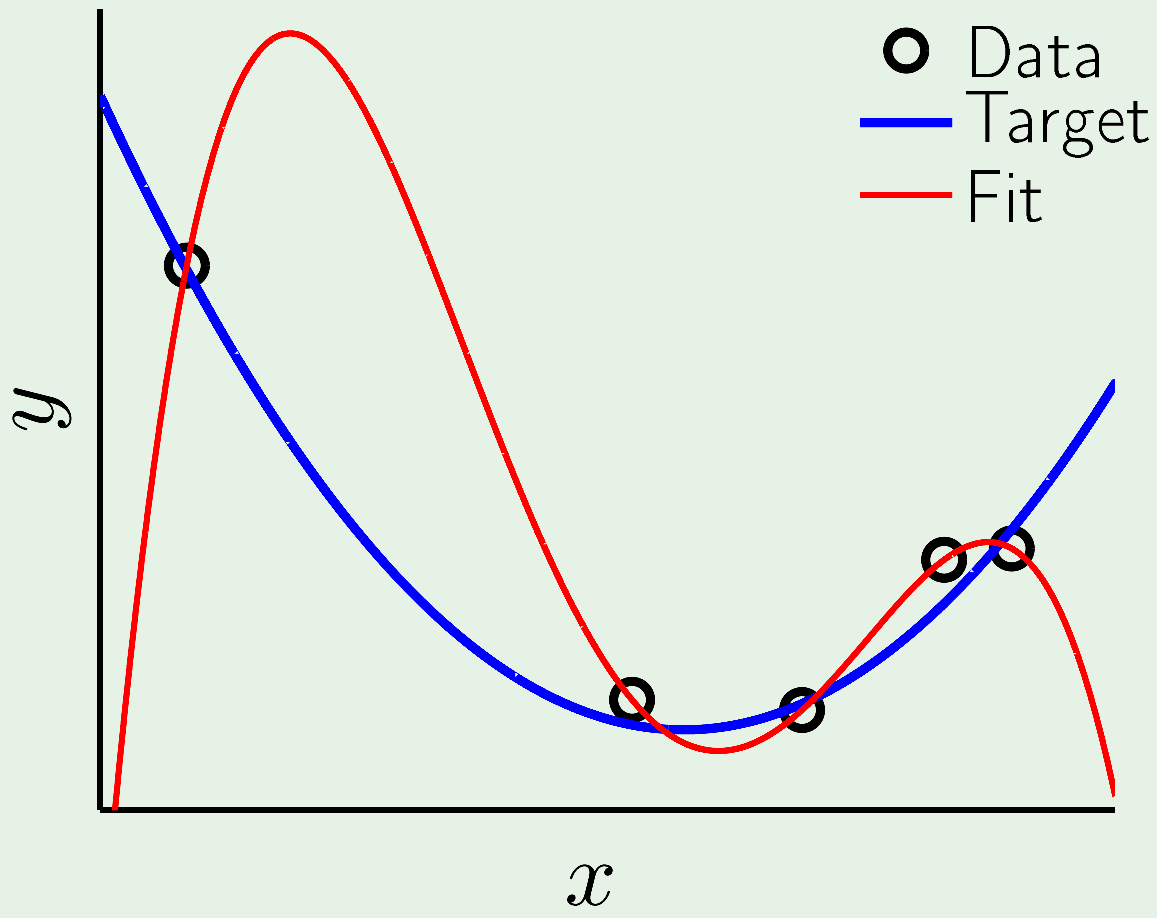
Illustration of overfitting

Simple target function

5 data points- **noisy**

4th-order polynomial fit

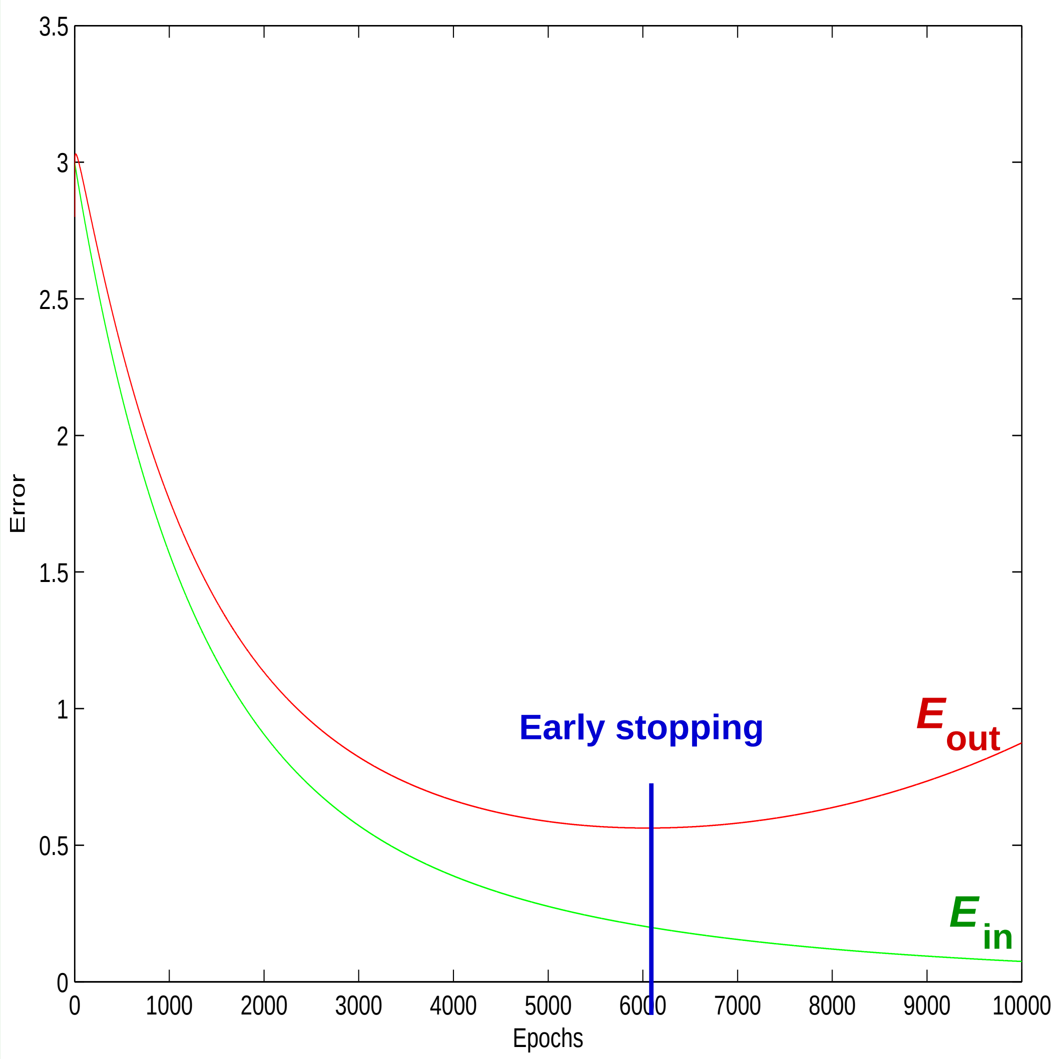
$$E_{\text{in}} = 0, \quad E_{\text{out}} \text{ is huge}$$



Overfitting versus bad generalization

Neural network fitting noisy data

Overfitting: $E_{\text{in}} \downarrow \quad E_{\text{out}} \uparrow$



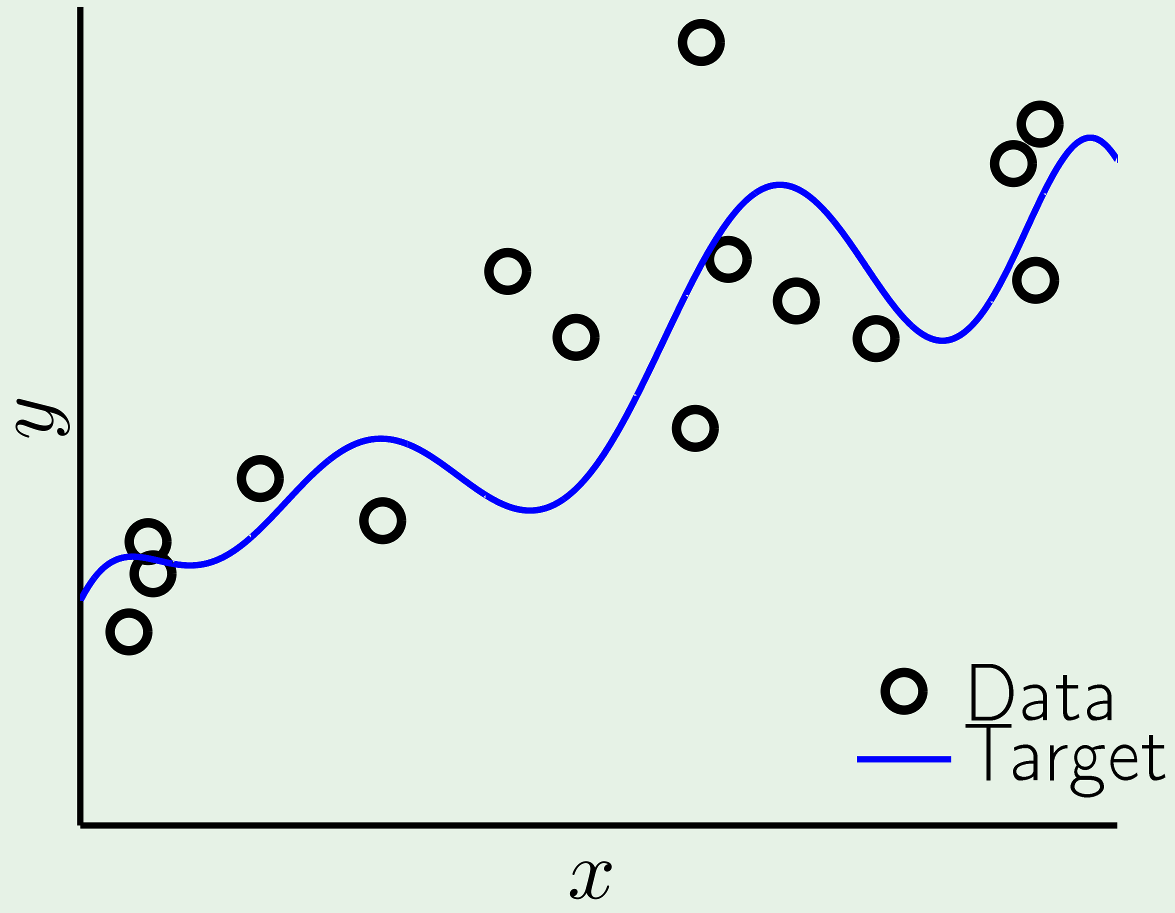
The culprit

Overfitting: "fitting the data more than is warranted"

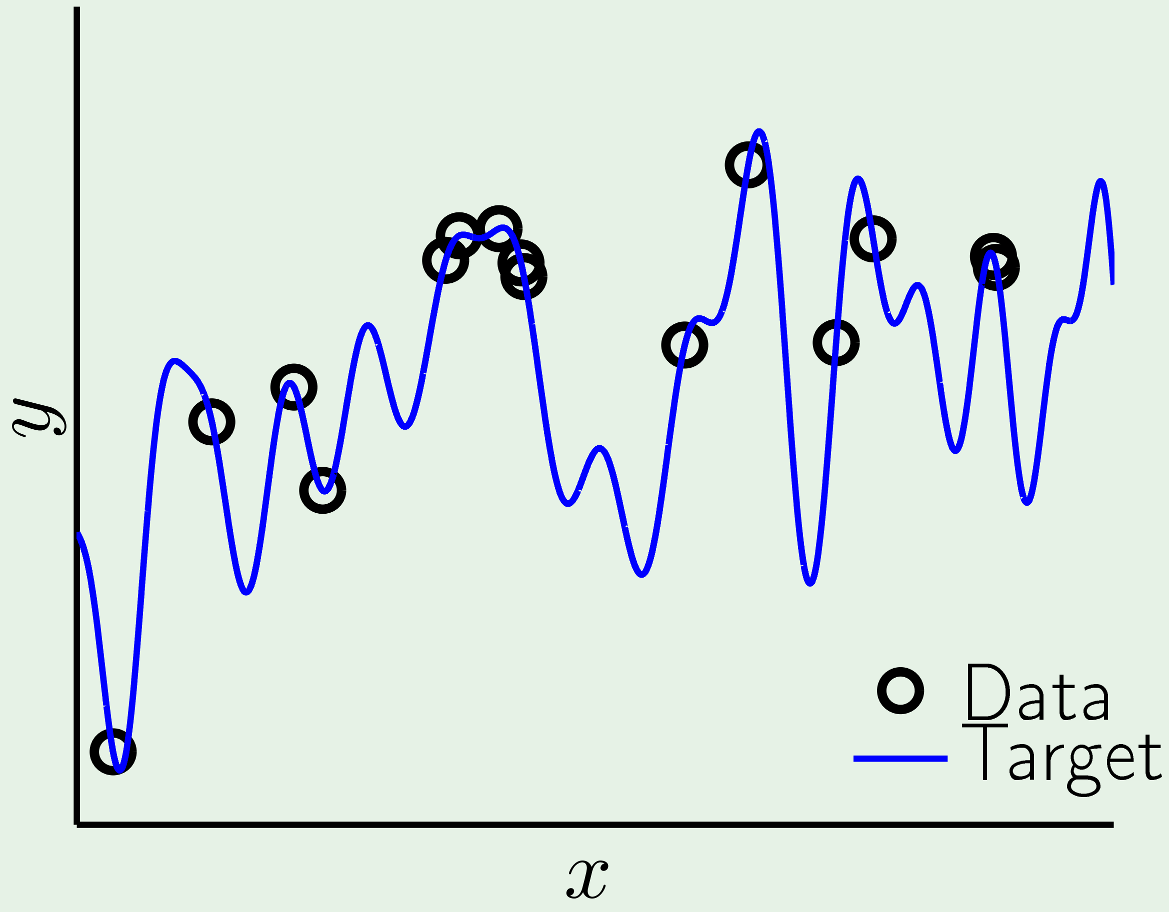
Culprit: fitting the noise - harmful

Case study

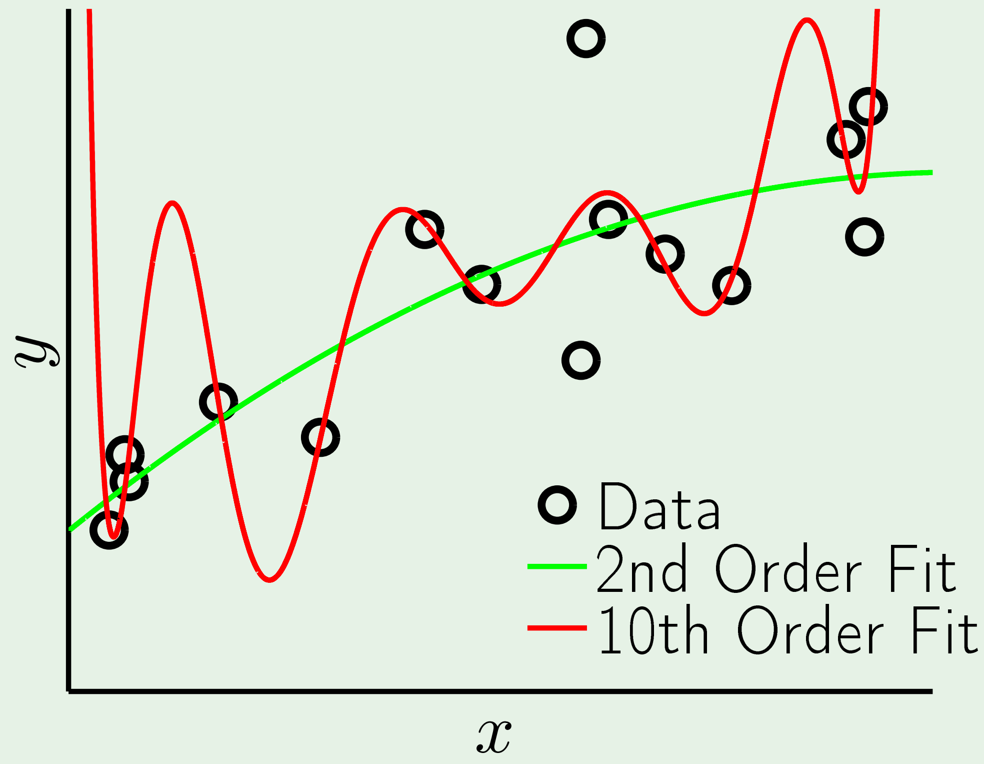
10th-order target + noise

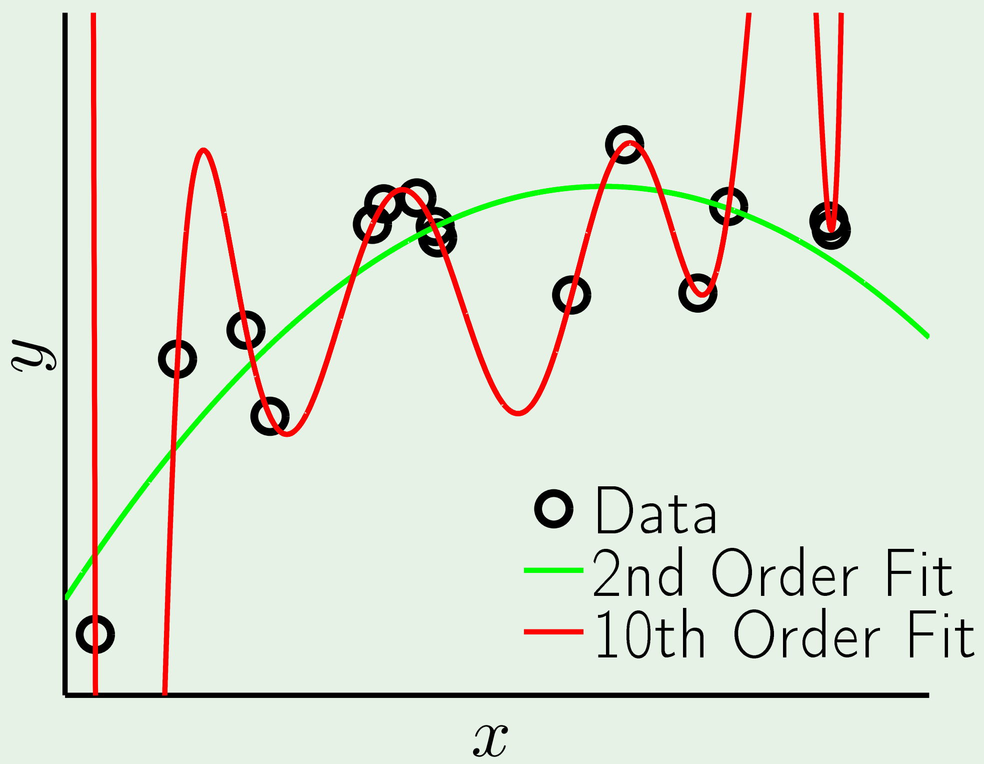


50th-order target



Two fits for each target





Noisy low-order target

	2nd Order	10th Order
E_{in}	0.050	0.034
E_{out}	0.127	**9.00**

Noiseless high-order target

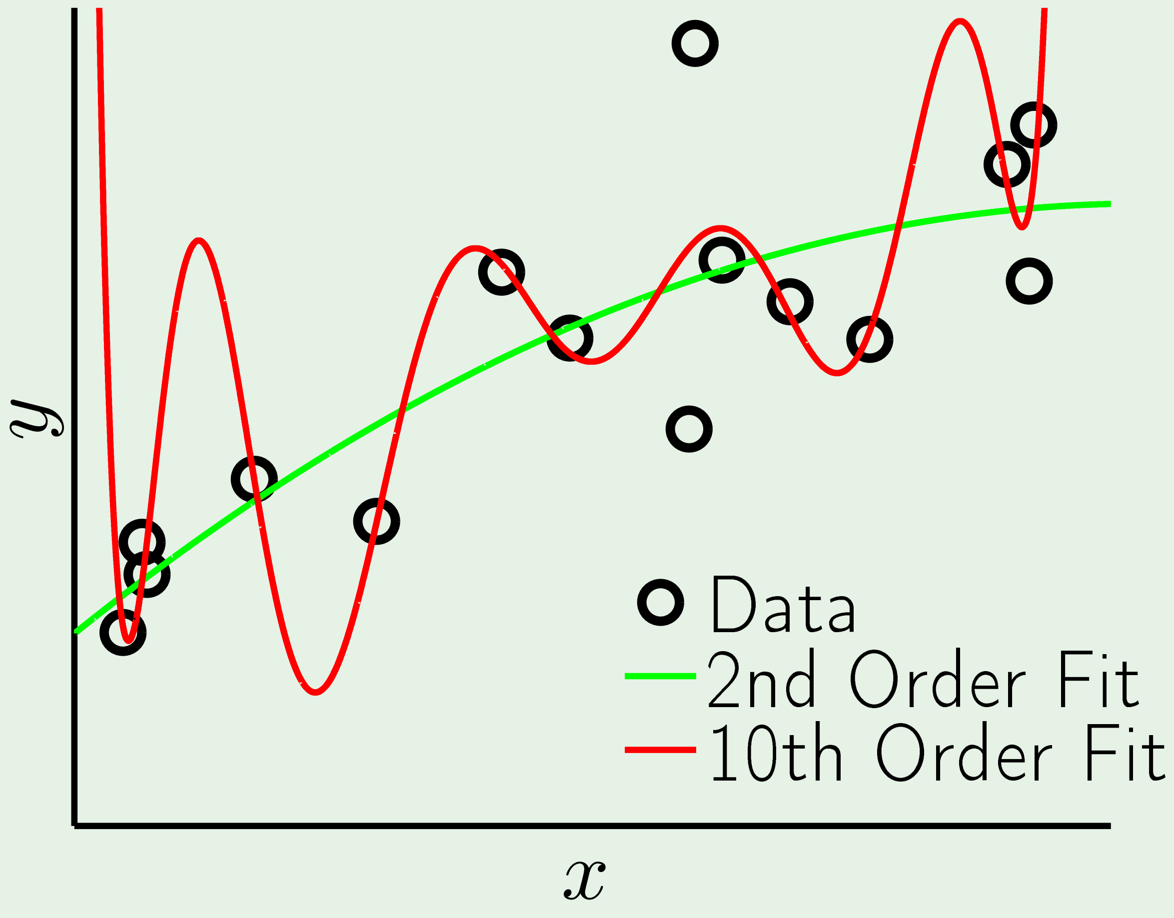
	2nd Order	10th Order
E_{in}	0.029	10^{-5}
E_{out}	0.120	**7680**

An irony of two learners

Two learners O and R

They <u>know</u> the target is 10th order

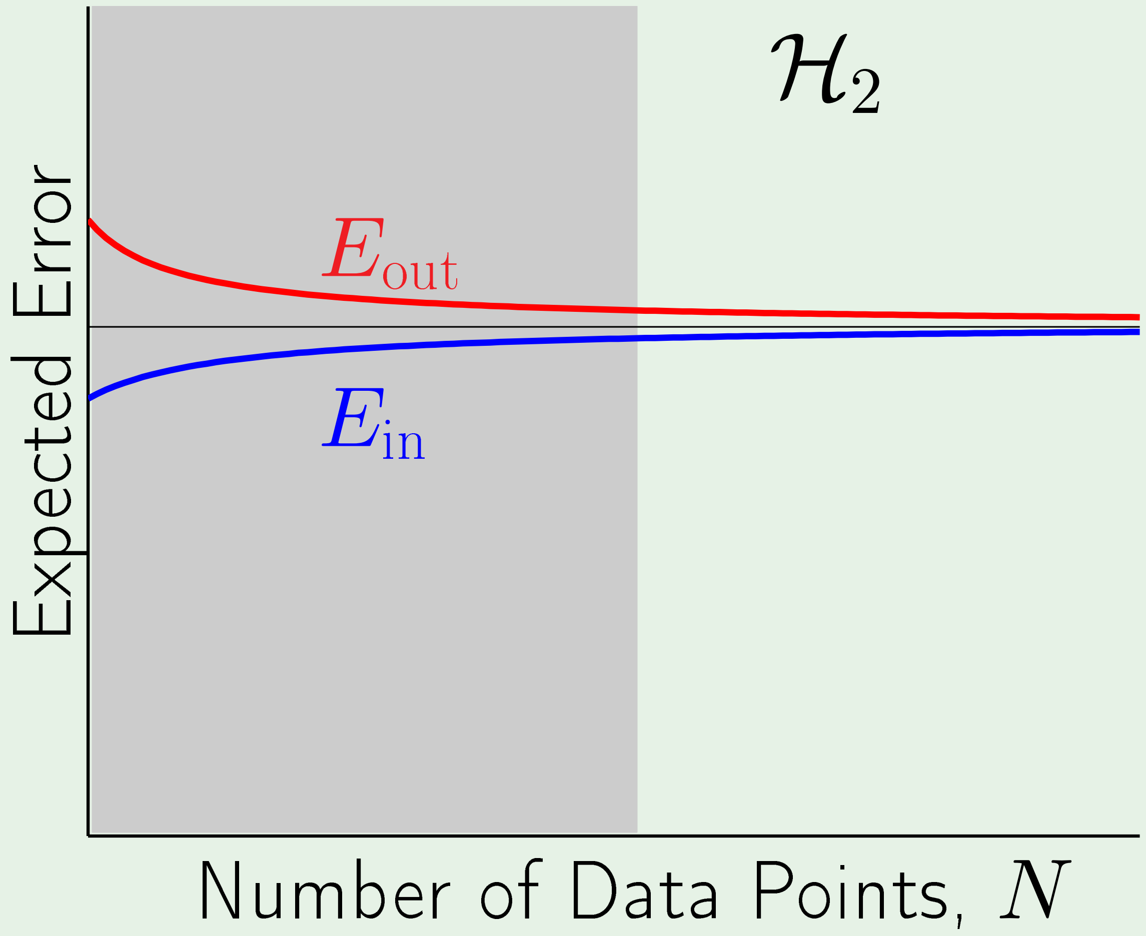
O chooses $\mathcal{H}_{10}$ R chooses $\mathcal{H}_2$

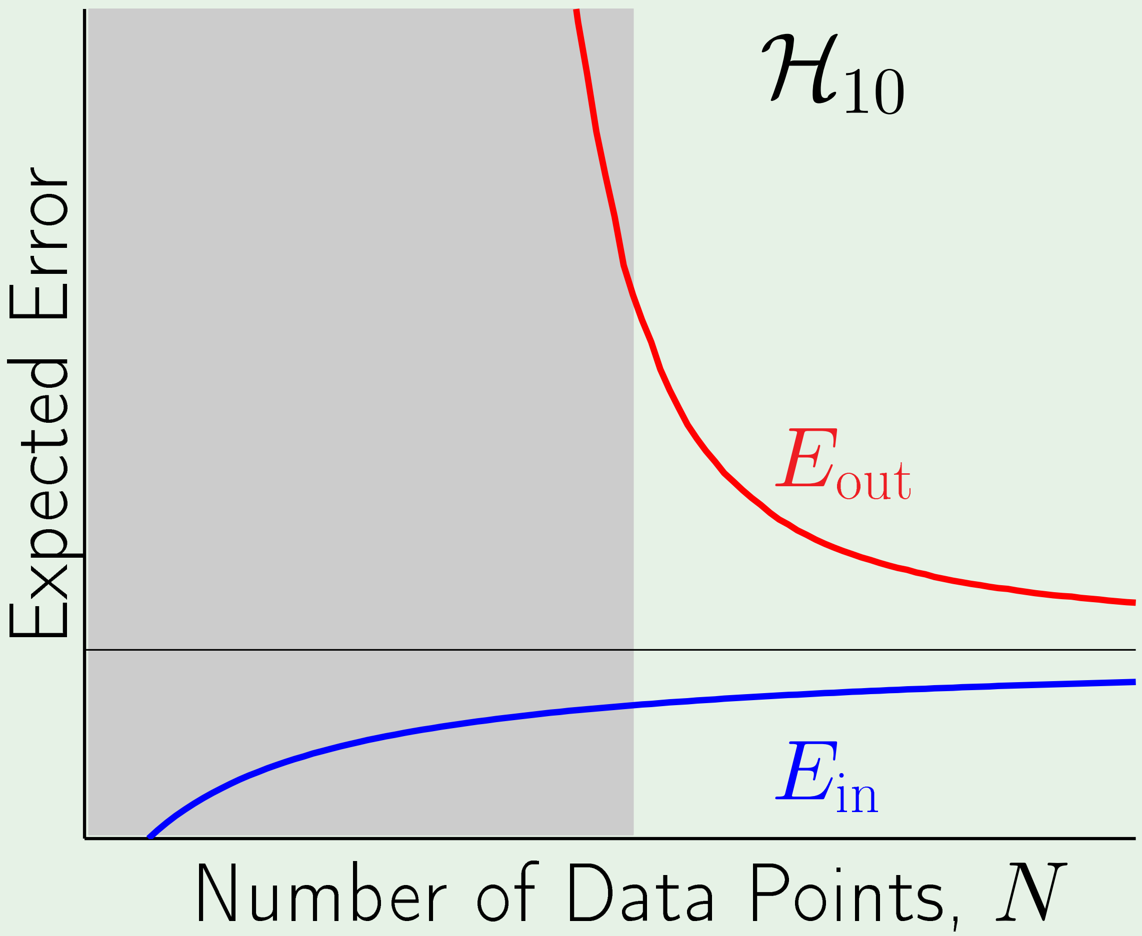


Learning a 10th-order target

We have seen this case

Remember learning curves?



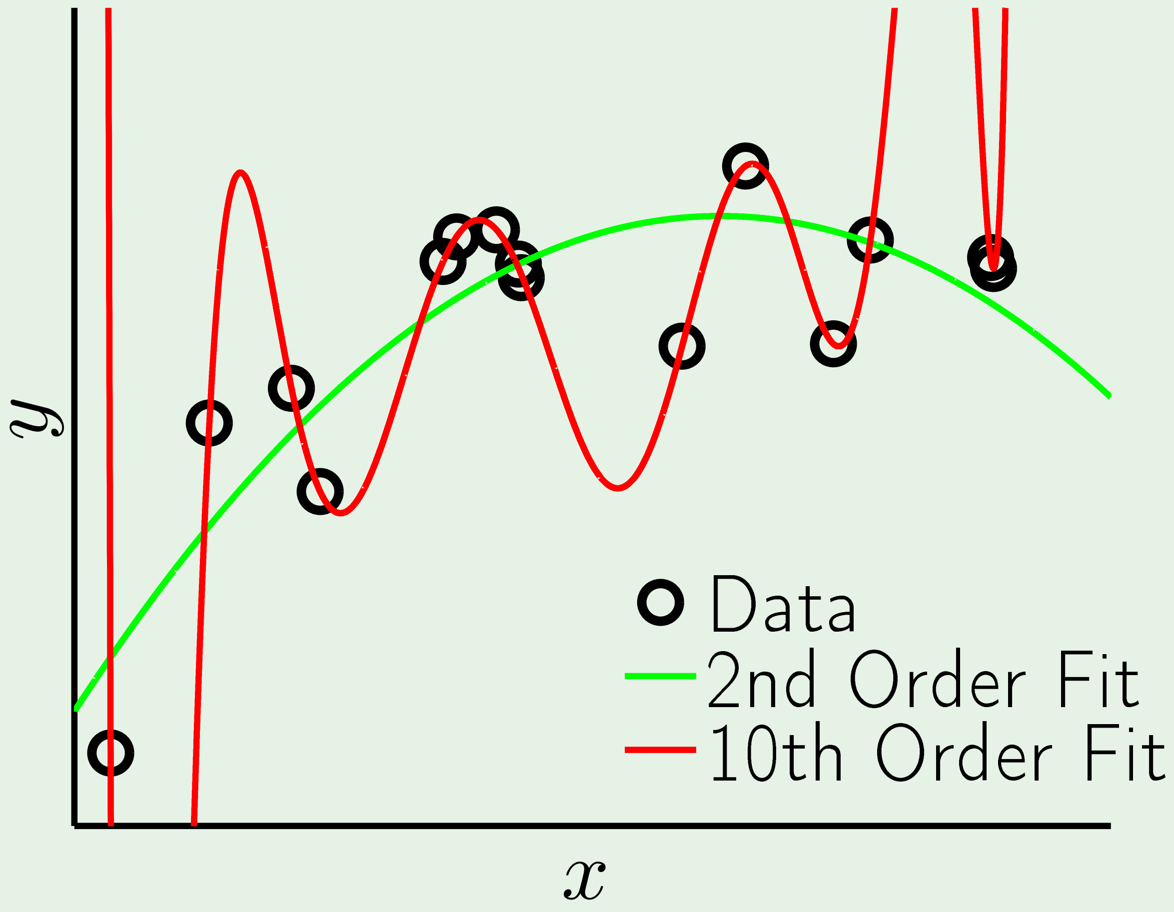


Even without noise

The two learners $\mathcal{H}_{10}$ and $\mathcal{H}_2$

They know there is no noise

Is there really no noise?



Learning a 50th-order target