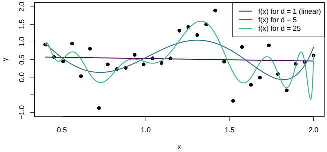
Introduction to Machine Learning

Polynomial Regression Models



Learning goals

- Understand how to add flexibility to the linear model by using polynomials
- Understand that this only affects the hypothesis space, not risk or optimization
- Understand that more flexibility is not equivalent to a better model

REGRESSION: POLYNOMIALS

We can make linear regression models much more flexible by using *polynomials* $\mathbf{x}_j^d$ – or any other *derived features* like $\sin(\mathbf{x}_j)$ or $(\mathbf{x}_j \cdot \mathbf{x}_k)$ – as additional features.

The optimization and risk of the learner remain the same.

Only the hypothesis space of the learner changes:
instead of linear functions

$$f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right) = \theta_0 + \theta_1 \mathbf{x}_1^{(i)} + \theta_2 \mathbf{x}_2^{(i)} + \dots$$

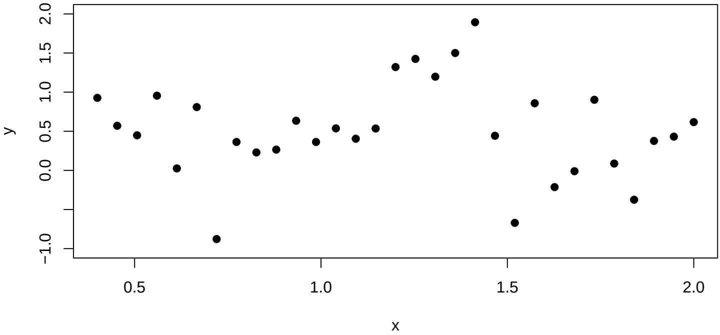
of only the original features,

it now includes linear functions of the derived features as well, e.g.

$$f\left(\mathbf{x}^{(i)} \mid \boldsymbol{\theta}\right) = \theta_0 + \sum_{k=1}^{d} \theta_{1k} \left(\mathbf{x}_1^{(i)}\right)^k + \sum_{k=1}^{d} \theta_{2k} \left(\mathbf{x}_2^{(i)}\right)^k + \dots$$

REGRESSION: POLYNOMIALS

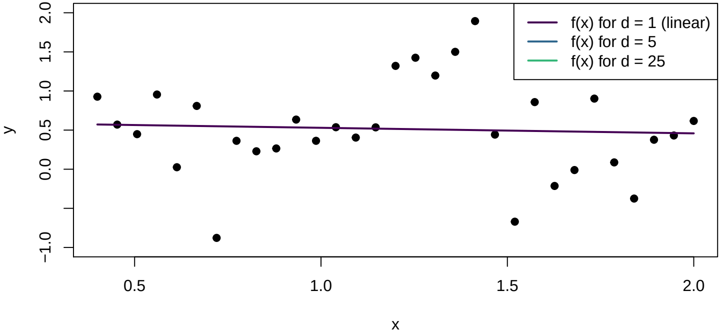
Polynomial regression example



REGRESSION: POLYNOMIALS

Polynomial regression example

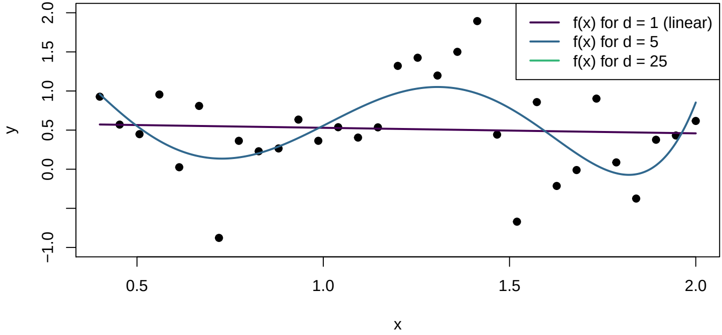
Models of different *complexity*, i.e., of different polynomial order d, are fitted to the data:



REGRESSION: POLYNOMIALS

Polynomial regression example

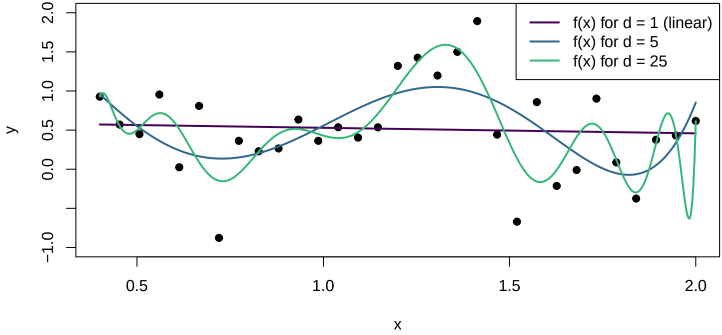
Models of different *complexity*, i.e., of different polynomial order d, are fitted to the data:



REGRESSION: POLYNOMIALS

Polynomial regression example

Models of different *complexity*, i.e., of different polynomial order d, are fitted to the data:



REGRESSION: POLYNOMIALS

The higher d is, the more **capacity** the learner has to learn complicated functions of $\mathbf{x}$, but this also increases the danger of **overfitting**:

The model space $\mathcal{H}$ contains so many complex functions that we are able to find one that approximates the training data arbitrarily well.

However, predictions on new data are not as successful because our model has learnt spurious “wiggles” from the random noise in the training data (much, much more on this later).

REGRESSION: POLYNOMIALS