Overfitting

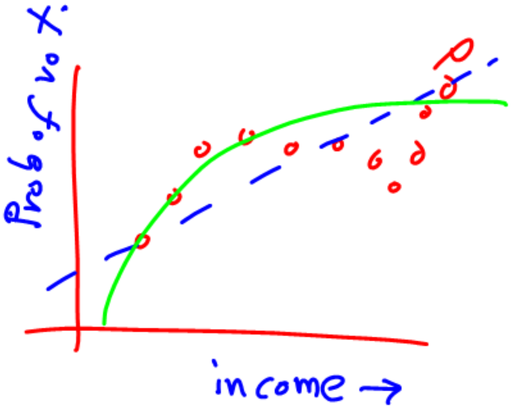
Mohammad Emtiyaz Khan
EPFL

Sep 29, 2015

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Motivation

Linear model can be too limited and usually underfit. One way is to use nonlinear basis functions instead.



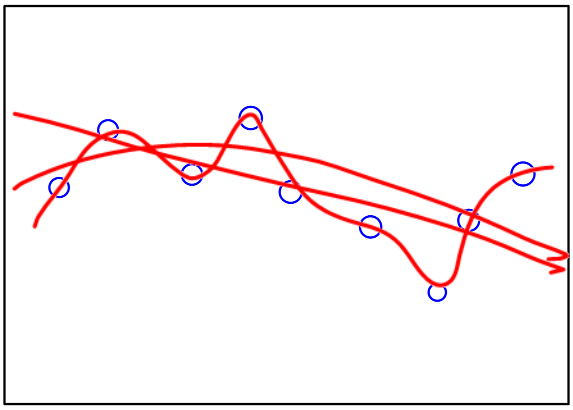
Consider simple linear regression. Given one-dimensional input x_n, we can generate a polynomial basis.

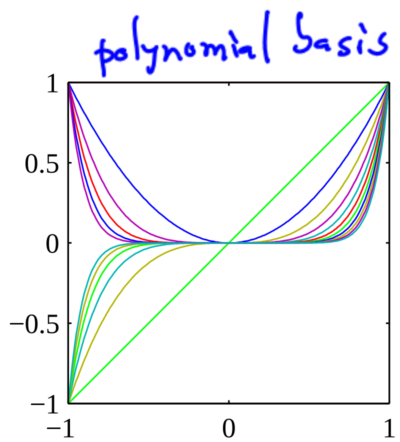
$y_n = \beta_0 + \beta_1 x_{n1}$

$$\boldsymbol{\phi}(x_n) = [1,\, x_n,\, x_n^2,\, x_n^3,\, \ldots,\, x_n^M]$$

Then we fit a linear model using the original *and* the generated features:

$$y_n \approx \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \ldots + \beta_M x_n^M$$





Overfitting and Underfitting

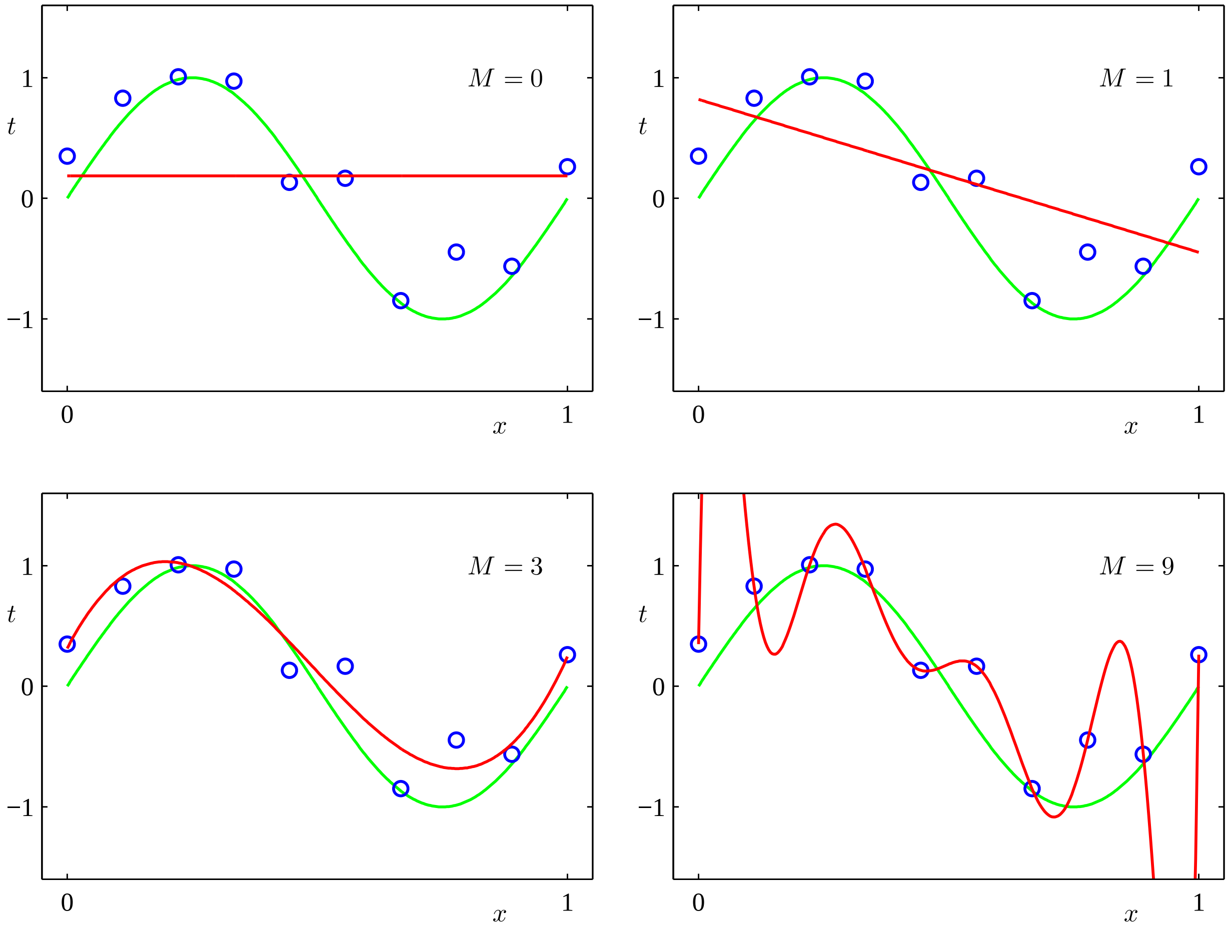
Overfitting is fitting the noise in addition to the signal. Underfitting is not fitting the signal well. In reality, it is very difficult to be able to tell the signal from the noise.

Question

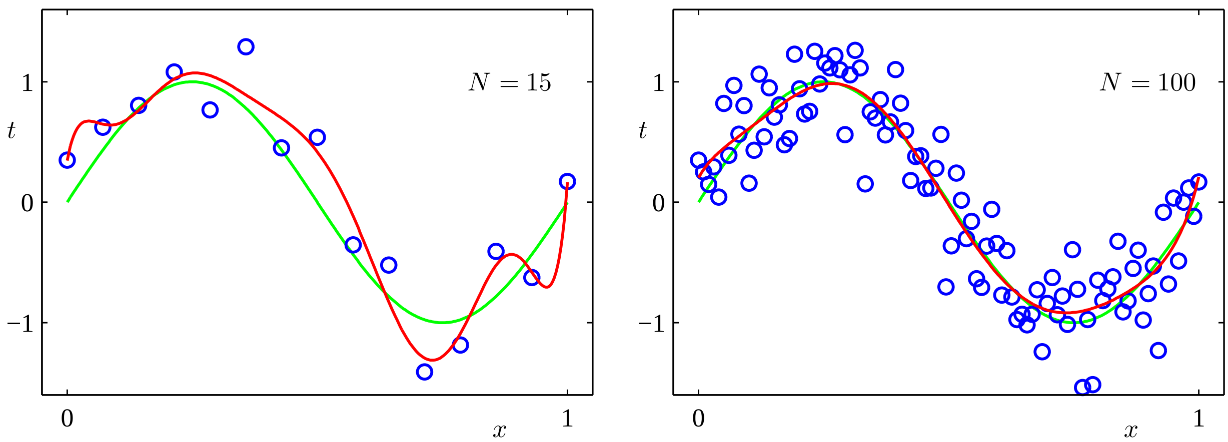
Complex models overfit easily

Circles are data points, green line is the truth & red line is the model fit. M is the maximum degree in the generated polynomial basis.

$\beta_0 + \beta_1 x_n^2 + \beta_2 x_n^2 + - - + \beta_M x_n^M \leftarrow$



If you increase the amount of data, overfitting *might* reduce.



Occam's razor

One solution is dictated by Occam's razor which states that "Simpler models are better – in absence of certainty."

Sometimes, if you increase the amount of data, you might reduce overfitting. But, when unsure, choose a simple model over a complicated one.

Question!

We can choose simpler models by adding a regularization term which 'penalizes' complex models.

$$\min_{\beta} \quad \frac{1}{2N} \sum_{n=1}^{N} [y_n - \widetilde{\boldsymbol{\phi}}(\mathbf{x}_n)^T \boldsymbol{\beta}]^2 \quad + \quad \frac{\lambda}{2N} \sum_{j=1}^{M} \beta_j^2$$

where $\lambda > 0$.

To do

Read about overfitting in the paper by Pedro Domingos (section 3 and 5 of "A few useful things to know about machine learning").