Machine Learning Course - CS-433

Overfitting

Oct 6, 2016

minor changes by Martin Jaggi 2016

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Motivation

Most models can be either *too limited*, or also *too powerful*. In other words, models can either underfit or overfit.

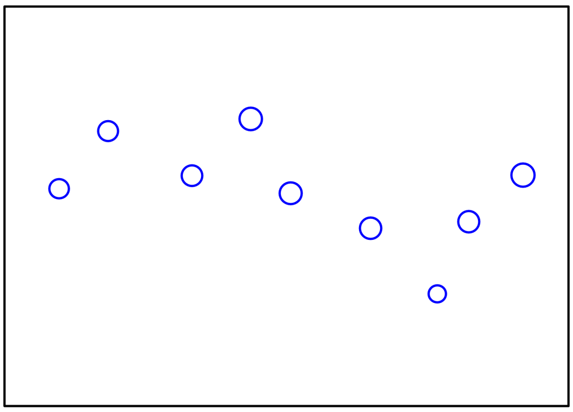
Can Linear Models Overfit?

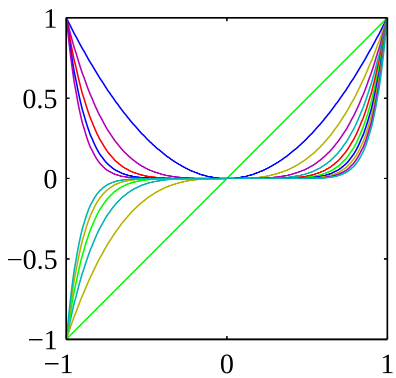
Yes! #1: Consider simple linear regression. Given one-dimensional input x_n, we can generate a polynomial basis.

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

Then we fit a linear model on the generated feature**s**, instead of the original feature:

$$y_n \approx w_0 + w_1 x_n + w_2 x_n^2 + \ldots + w_M x_n^M$$



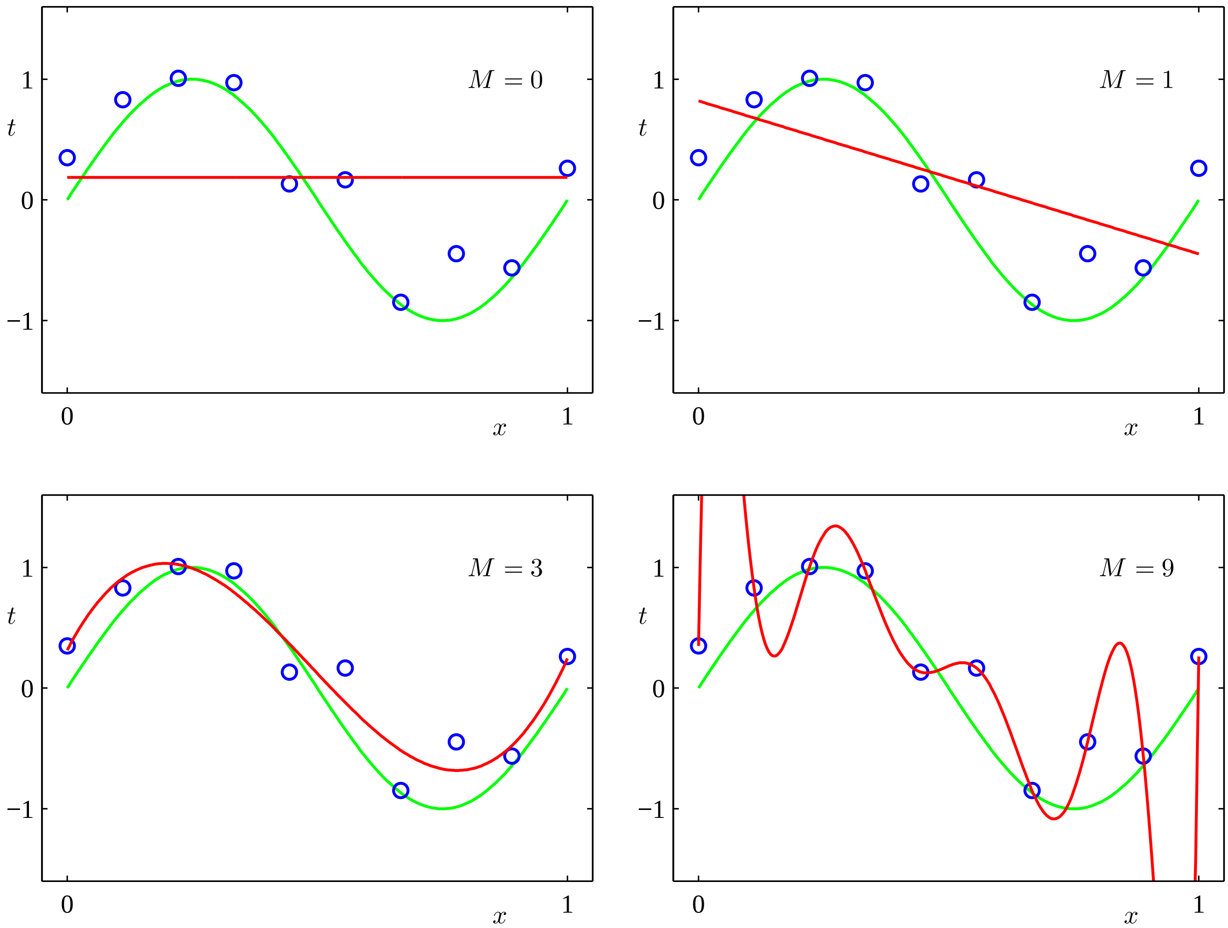


Overfitting and Underfitting

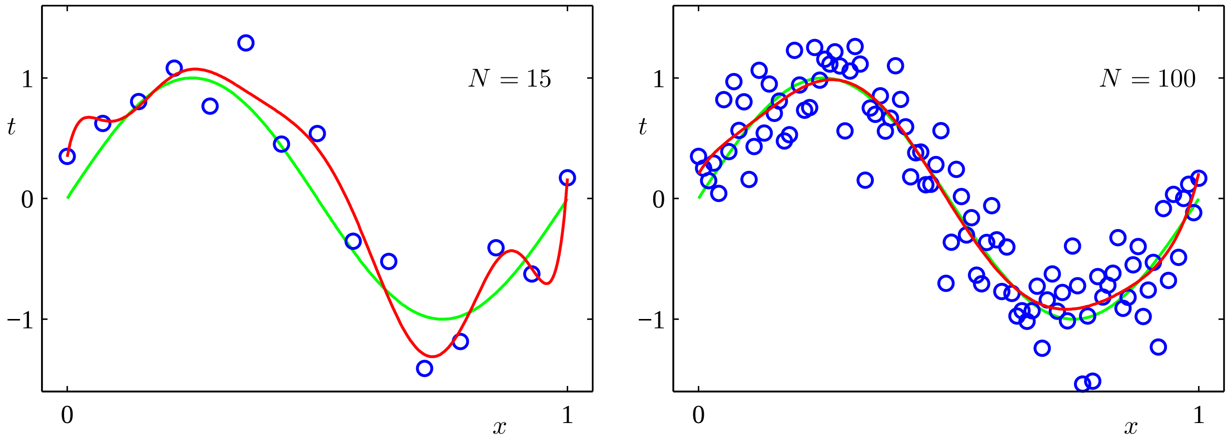
Overfitting is fitting the noise in addition to the signal. Underfitting is not fitting the signal well.
In real life, it is difficult to distinguish signal from noise.

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.



If you increase the amount of data (increase N, but keep M fixed), overfitting *might* reduce.



Preventing Overfitting

- Get more data!
- Regularization - Force the model to be not too complex

Occam's Razor

One solution is dictated by Occam's razor which states that "Plurality is not to be posited without necessity" or rephrased "Simpler models are better - only use complicated ones if strictly necessary".

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

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

$$\min_{\mathbf{w}} \quad \frac{1}{2N}\sum_{n=1}^{N}\left[y_n - \boldsymbol{\phi}(\mathbf{x}_n)^\top \mathbf{w}\right]^2 \; + \; \Omega(\mathbf{w})$$

ToDo

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