The Minimum Description Length Principle

Peter Grünwald
CWI Amsterdam
www.grunwald.nl
(slides edited by Tim van Erven)

Machine Learning Course,
Vrije Universiteit Amsterdam
December 5th 2007

Minimum Description Length Principle

Rissanen 1978, 1987, 1996,
Barron, Rissanen and Yu 1998

- 'MDL' is a method for inductive inference…
  - machine learning
  - pattern recognition
  - statistics
- …based on ideas from data compression (information theory)
- In contrast to most other methods, MDL automatically deals with overfitting, arguably the central problem in machine learning and statistics

Minimum Description Length Principle

- MDL is based on the correspondence between ‘regularity’ and ‘compression’:
  - The more you are able to compress a sequence of data, the more regularity you have detected in the data
  - Example:

001001001001001001001001:::001

010110111001001110100010101:::010

Minimum Description Length Principle

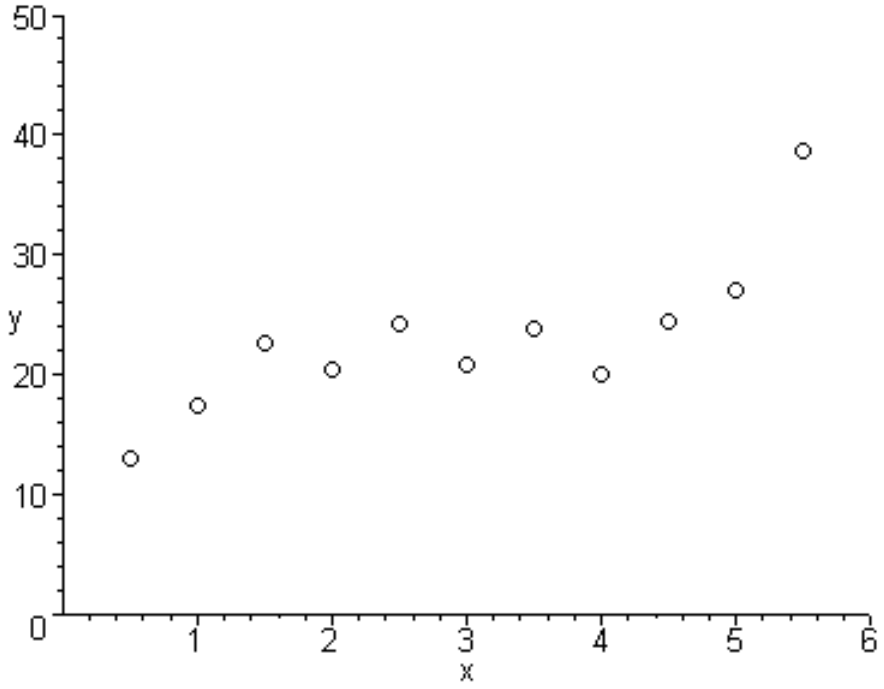
- MDL is based on the correspondence between ‘regularity’ and ‘compression’:
  - The more you are able to compress a sequence of data, the more regularity you have detected in the data…
  - …and thus the more you have learned from the data:
    - ‘inductive inference’ as trying to find regularities in data (and using those to make predictions of future data)

Model Selection/Overfitting

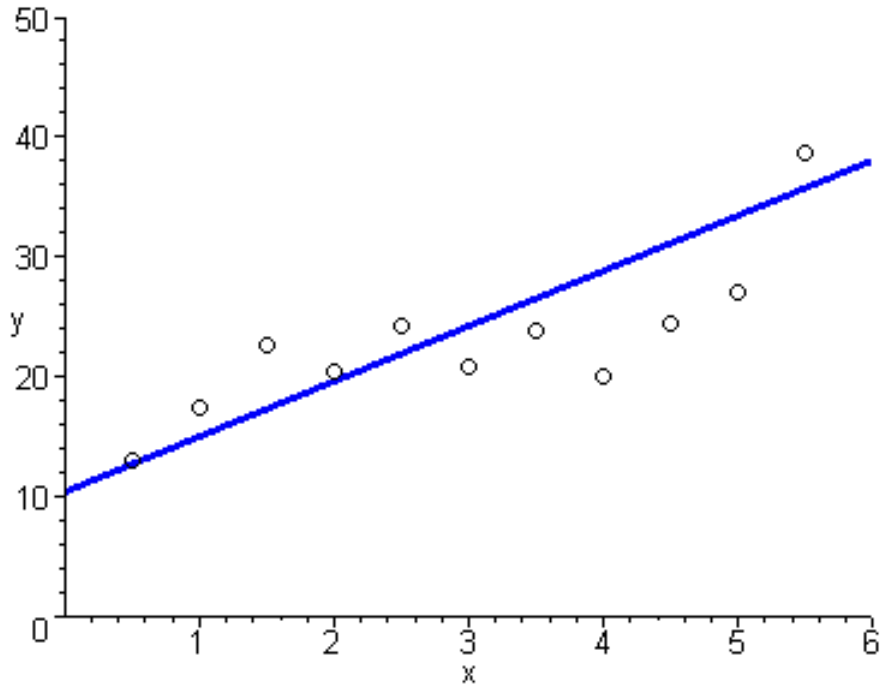
Given data D and hypothesis spaces/models $M_1, M_2 M_3, \ldots$, which model *best explains* the data?

- Need to take into account
  - Complexity of models
  - Error (minus Goodness-of-fit)
- Example:
  - Selecting the degree of a polynomial in regression
  - Sum of squared errors

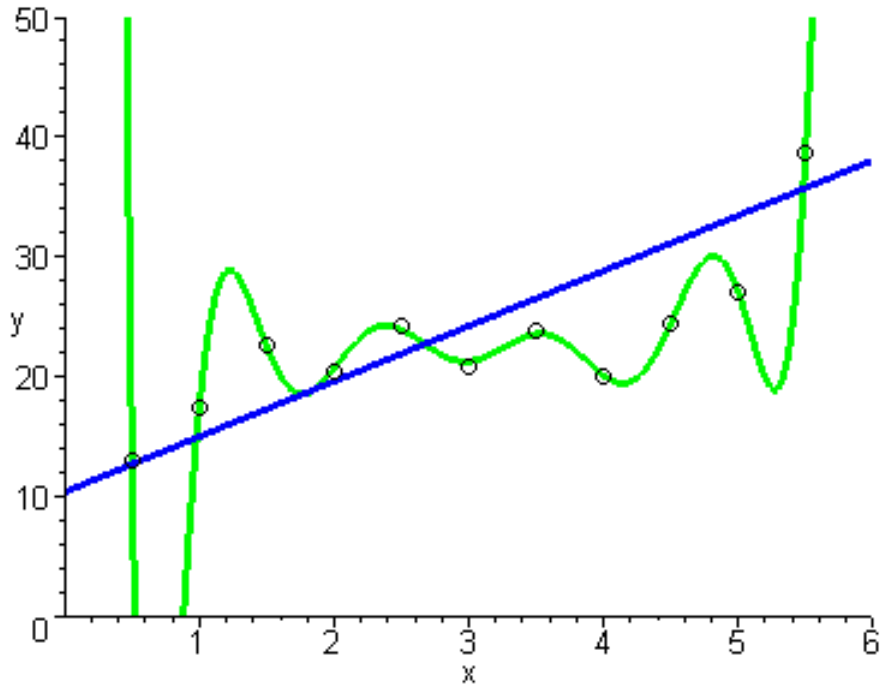
Example: Regression



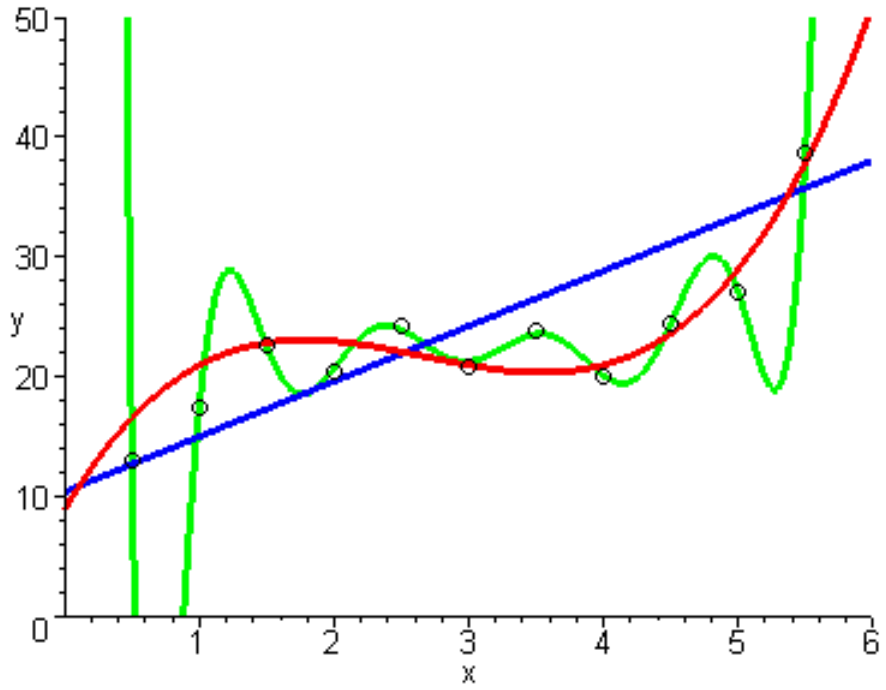
Example: Regression



Example: Regression



Example: Regression



Example: Regression