Model Selection and Assessment

Sam Reid
Advisor Greg Grudic

University of Colorado at Boulder
Department of Computer Science

January 23, 2008

Model Selection and Assessment

- Model Assessment: "How good is my model?"
- Model Selection: "How can I fine-tune my model?"
- Model Assessment $\Rightarrow$ Model Selection

Definitions

- Data Set

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$$

- Model = Classifier/Regression Model

$$\hat{y}(\mathbf{x}) = f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$$

- Model is characterized by internal parameters
- Machine Learning Algorithm = Inducer

$$\mathcal{I}(\mathcal{D}|\mathbf{\Phi}) = \hat{y}(\mathbf{x})$$

- $\mathbf{\Phi}$ are tuning parameters/hyperparameters

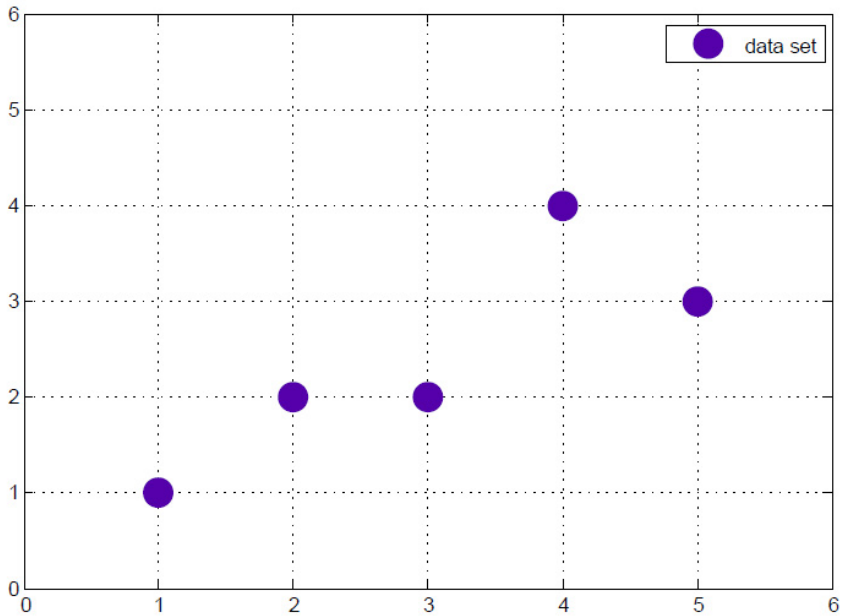
Classifier Examples

- Linear Model, e.g. $\hat{y}(\mathbf{x}) = 3x + 2$
- Polynomial Model e.g. $\hat{y}(\mathbf{x}) = 3x_2^2 + 2x_1^9 + 3 + 7x_0$
- K-Nearest Neighbor Model (regression), e.g. $\hat{y}(\mathbf{x}) = avg_y(nearest3(a, b, c, d, e))$
- Fourier Series
- Neural Network
- Decision Tree
- Support Vector Machine

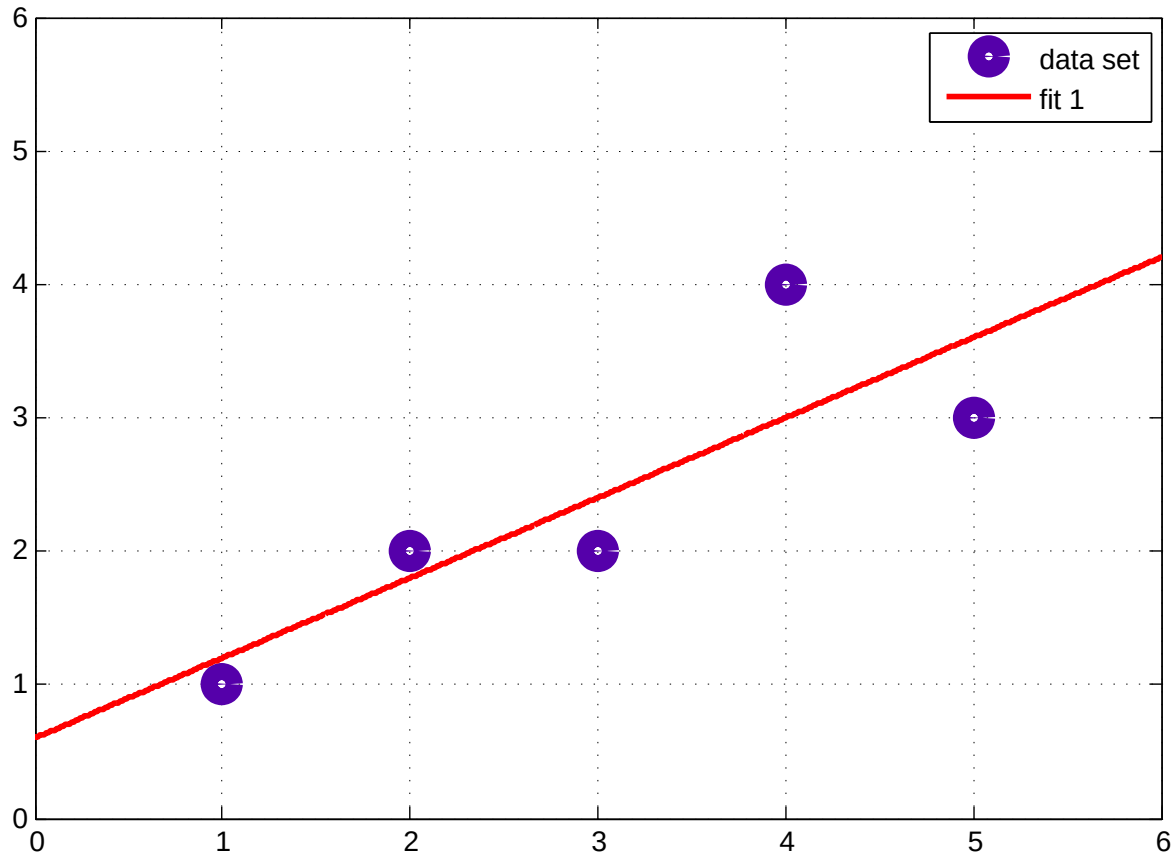
Data Example

- Example Dataset (1-dimensional regression problem):

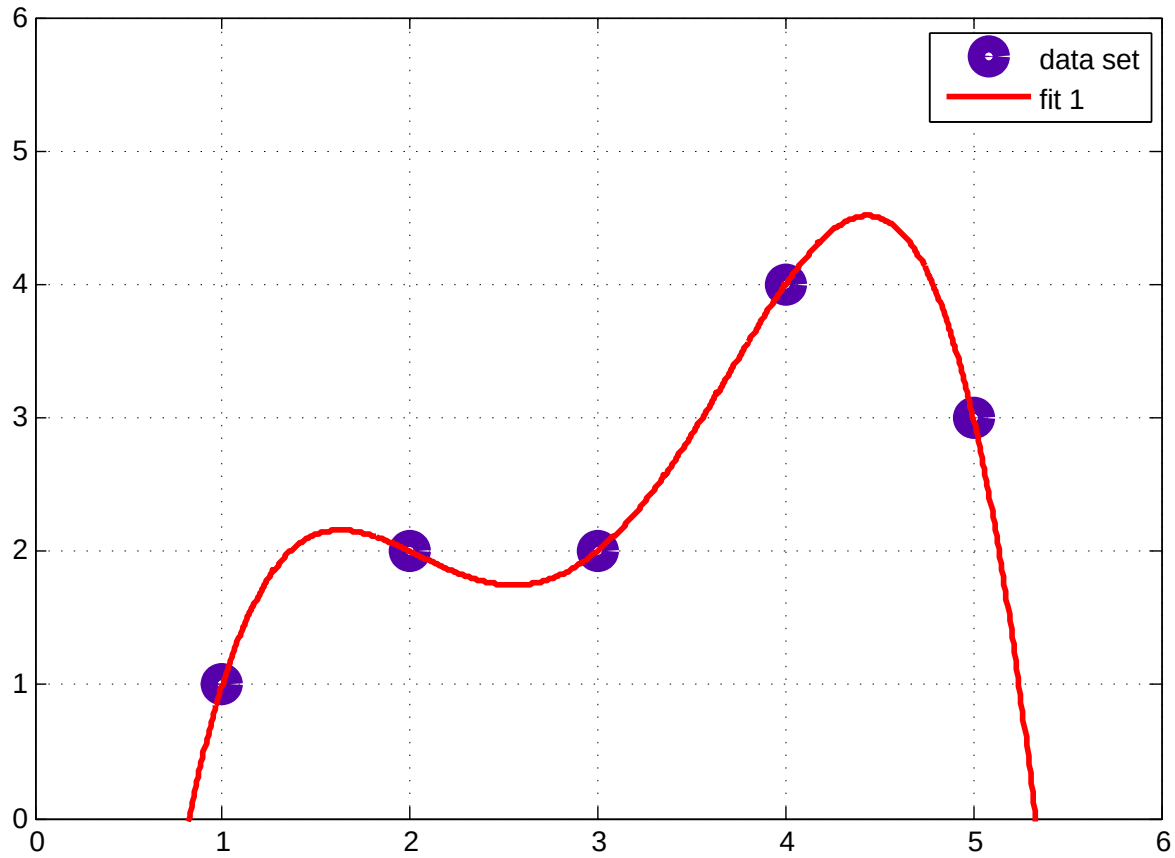
x	y
1	1
2	2
3	2
4	4
5	3



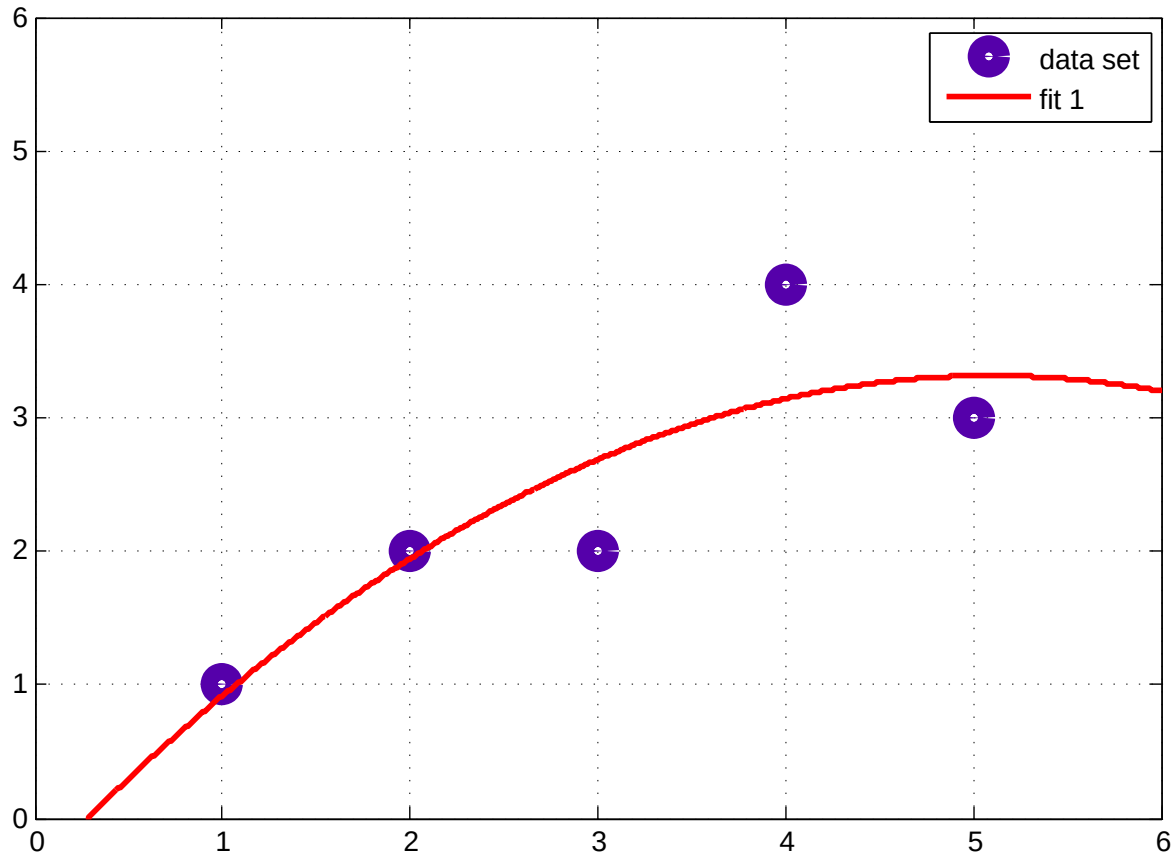
Linear Fit



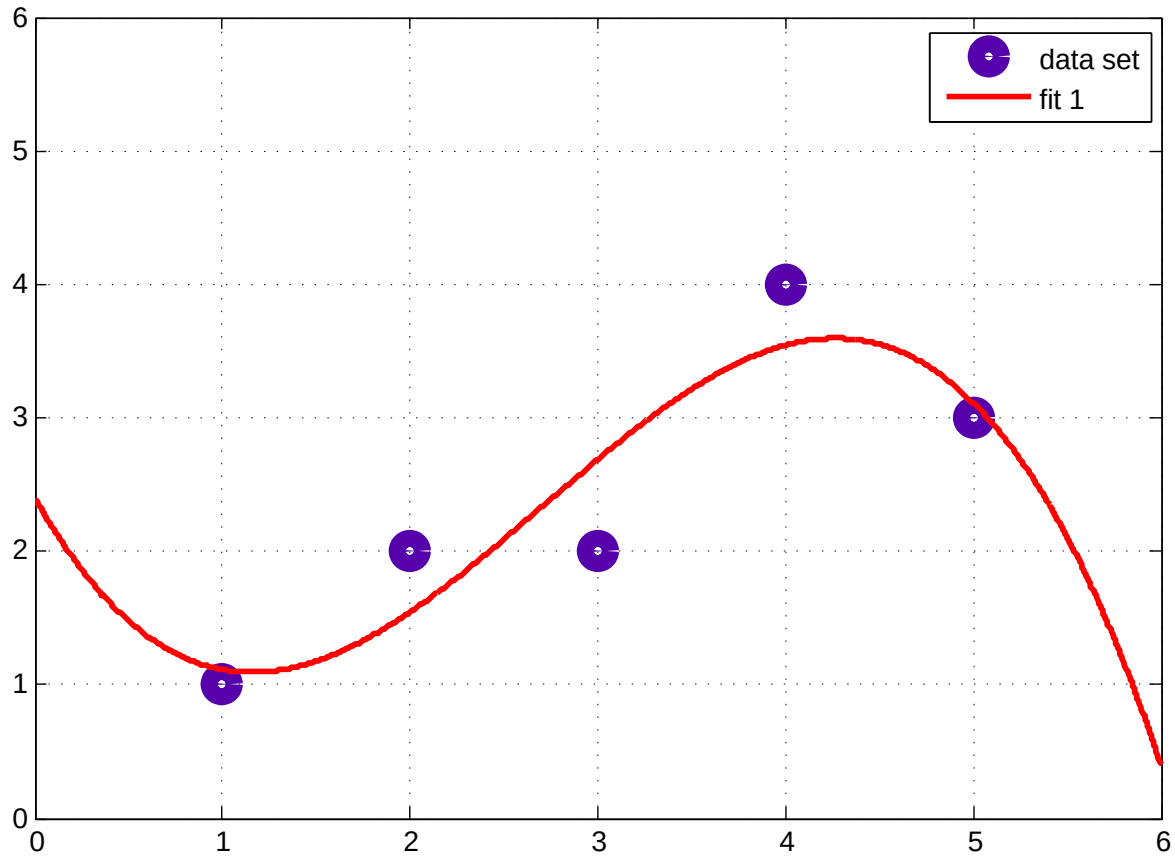
4th Degree Polynomial



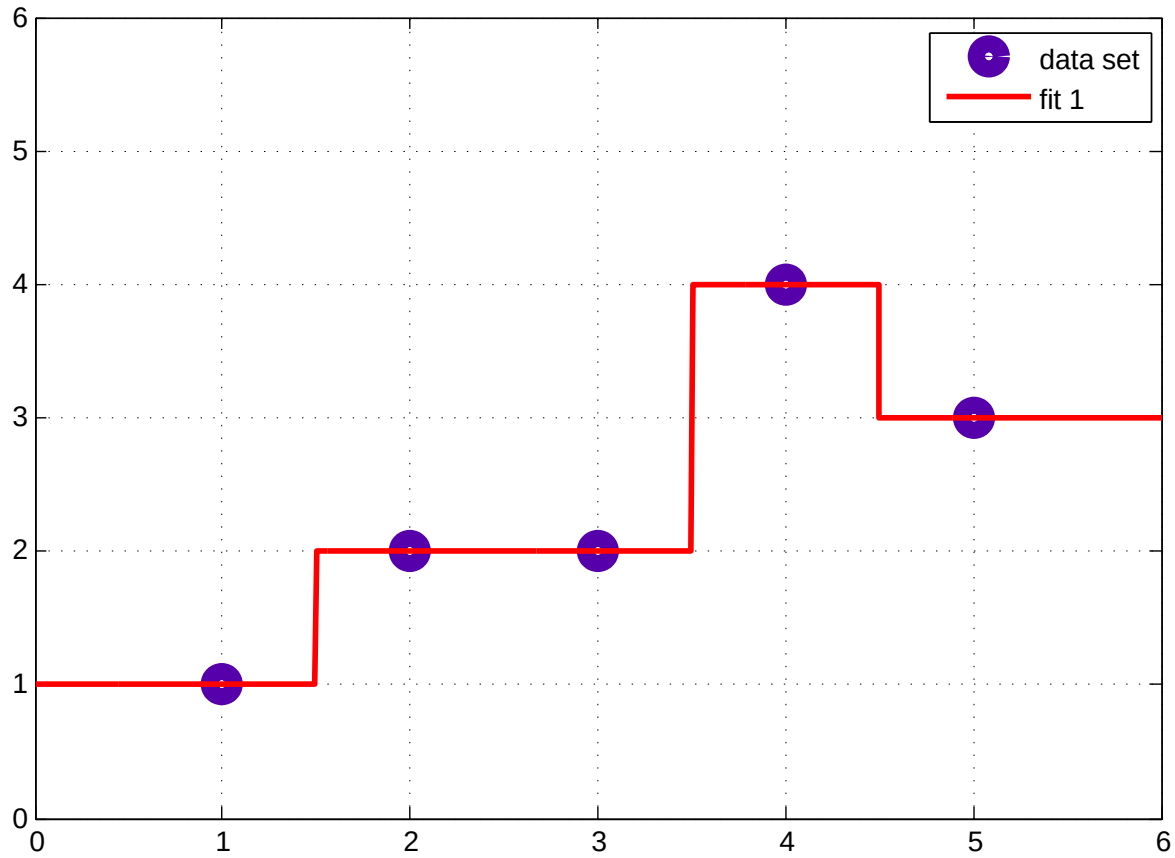
Quadratic



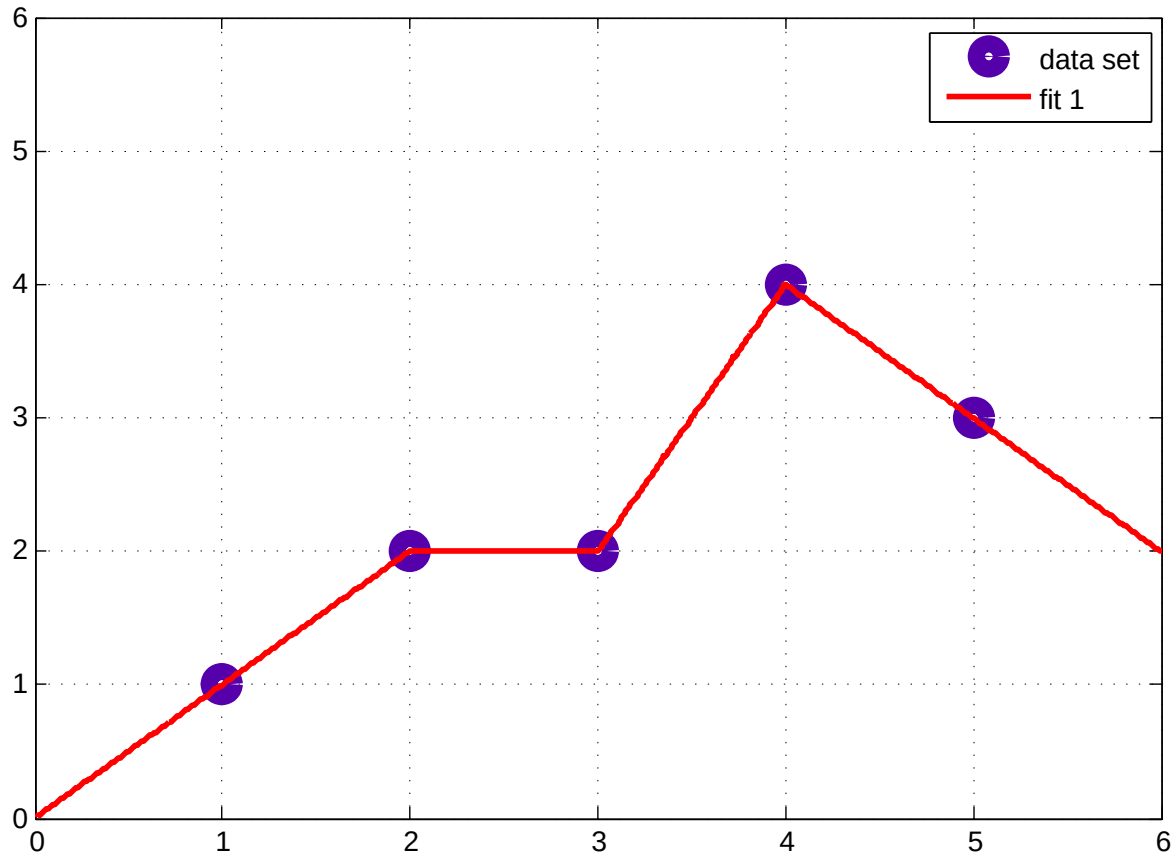
Cubic



1-Nearest Neighbor



Piecewise Linear



How can we choose a model?

- Two Techniques:
  1. The data generating model is known. (e.g. $x = x_0 + v * t + \epsilon$)
  2. Make assumptions about the data.
- For our purposes, the data generating model is unknown.

... a learner that makes no a priori assumptions regarding the identity of the target concept has no rational basis for classifying any unseen instances.

-Mitchell, 1997

How can we choose a model? - II

- Bayesian: mathematically optimal, based on probability theory, difficult to implement for real-world problems
- Frequentist: Sampling techniques, direct estimate of desired quantity
  - Re-use training set
  - Hold-out test set
  - Bootstrap sampling
  - Cross-Validation

Problems with Occam's Razor

- Occam's Razor: Choose the simplest hypothesis that explains the data.
- Problem: How to quantify "simplest"?
- Problem: Should it explain the data exactly?

No Free Lunch

- Wolpert, 1994.
- All classification algorithms have identical performance, when averaged over all data sets.

Practical Issues

- Stratification
- Standardization

Hypothesis Testing

- Null Hypothesis: Hypothesis that two treatments/classifiers are the same
- Type I Error: Identifying a difference between two treatments/classifiers when there is none
- Type II Error: Identifying that two treatments/classifiers are different when they actually are the same

- Power: Ability to detect a difference between two treatments/classifiers

Hypothesis Testing-II

- Fit the given data to a known distribution (e.g. Gaussian or t-distribution)
- Compute the confidence that the distribution has certain properties.

Summary: Things you need to know

- How to compute mean-squared error
- The difference between a classifier and a classification algorithm.
- Why can't the training set be used to perform model assessment?
- The relationship between model selection and model assessment
- The difference between parameters and hyperparameters
- How to implement k-fold cross-validation.
- What the no-free lunch theorem says, and its relevance.
- The difference between bias and variance
- What is a loss function?
- What is stratification?