

Validating Learning Algorithms

Noel Welsh

18 October 2010

Announcements

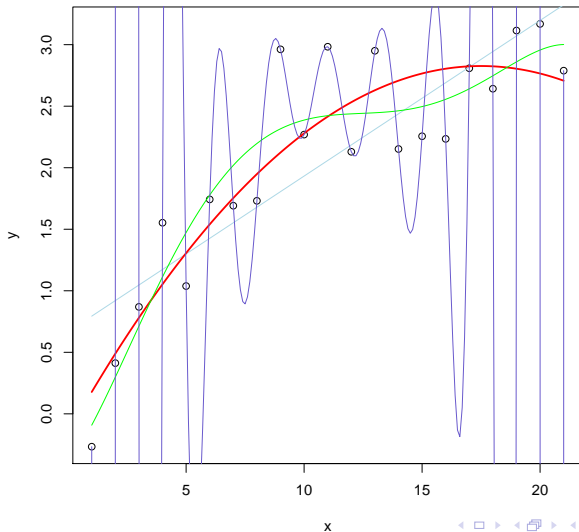
- Swipe card access
- In class assessment
- Broken kits

- In the last lecture we saw several simple models appropriate for classification or regression given the abilities of the robotics kit we're using.
- In this lecture we're going to see how we can assess one or more models

Motivating Problem

- Suppose we are considering several models for a learning problem. How should we select between them?
- E.g. train several polynomial regression models to fit data.

Example



- We see the model with highest likelihood doesn't fit the true distribution at all well.
- This is a general problem with increasing model complexity. At some point the model will start to fit the noise in the data, rather than the true trend. This known as *overfitting*

- We are interested in the performance on the model on data we have not yet observed. This is known as *generalisation*.
- The *generalisation error* is the performance measure, in our case likelihood, on all possible data.

- We would like to estimate the generalisation error using only the data we have. This is known as *validation*
- We're going to look at a general purpose algorithm for validation. Alternatively one can derive bounds on generalisation for many algorithms. We're trading computation for analysis.

Hold-out Cross Validation

- Divide data into training and testing sets. E.g. 70/30 split
- Estimate generalisation error from performance on testing set.
- Testing set also known as *hold-out* set.

k -fold Cross Validation

- Hold-out CV is wasteful of data.
- k -fold cross validation:
 - divide data into k equally sized sets.
 - Perform hold-out CV using each set as the hold-out set
 - Estimate generalisation error as average of the hold-out CV runs
- Typical values of k are 5-10.
- If each hold-out set is a single example, known as *leave-one-out cross validation*.

Assessing a Single Model

- k -fold cross validation is a valid method for assessing a single model, as well as choosing between models.

- Statistical learning theory is concerned with establishing generalisation bounds (amongst other topics).