CogSci 109: Lecture 17

Friday Nov. 16, 2007
Nonlinear interpolation (Splines examples), Error analysis

Outline for today

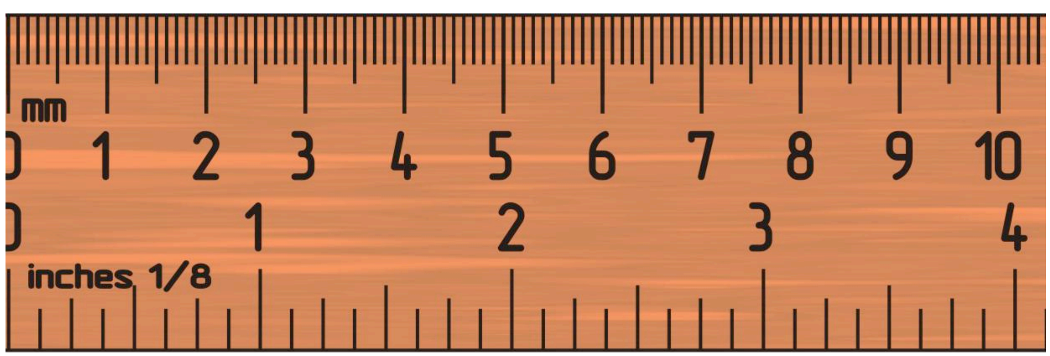
- Announcements
- Uncertainty
  - **Definition**
  - **A model for thought - the impossibility of certainty**
  - **What is the question uncertainty brings into focus?**
- Computing estimation error
- Different error estimates
- Error analysis in terms of methods discussed so far
  - **All these methods are models**
  - **Modeling based on data**
  - **Examples from last few lectures**
    - Comparing fits with error analysis

Announcements

- Midterms will be handed back on Monday
- Homework assigned tonight
  - **Simple short**
  - **Find and read a paper which applies one of the techniques outlined in the assignment**
  - **Answer conceptual questions**
- Reading posted later today

Uncertainty

- Error does not mean, in science, mistake
  - **It means the level of uncertainty in measurements and calculations**
  - **Can't eliminate by being careful, must instead minimize them**
- Basically want to have an estimate which is as reliable as possible
  - **'keep an eye on' your uncertainty**



Impossibility of certainty

- No physical quantity can be measured with absolute certainty
  - **Wood door example**
  - **Changing brain state, structure, dynamics**
- Mathematical approximations to real systems are ALWAYS approximations, no matter how good
  - **Any model you make is ONLY an approximation and should NEVER be confused with the real system**
    - **Wrong**: "The Brain is computing the inverse of this matrix"
    - **Right**: "Our model approximates what the brain is doing by computing the inverse of this matrix"

The question…

- The question is not whether you are right or not
- The question is whether your approximation is good enough to be useful, dependent on what you (and probably the rest of science) consider to be ‘good enough’

Computing the estimated error

- One way to assess how good your model is consists of computing an estimated error
  - **Typically you then decide whether your error is 'within bounds'**
    - (you create a boundary, such as the error in measuring/predicting position of a limb in space must be less than 10 inches)
- Uses one of many possible methods

Different error estimates

- There are many ways to estimate errors, here are a couple of common ones
  - **To get a single # - can use various norms**
    - 2-norm

$$\|e\|_2 = \sqrt{\sum_i (y_i - \hat{y}_i)^2}$$

    - Mean-squared-error

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

  - **Curve - simple error (for a time dependent signal y(t))**

$$e(t) = y(t) - \hat{y}(t)$$

  - **Curve - prediction error**

$$e_p(t) = y(t) - \hat{y}(t \mid t-1)$$

We’ve already developed several models and methods!

- Least squares is a very common method of fitting a model
  - **Works by minimizing an estimate of modeling error**
  - **Linear model, nonlinear model**
- Linear and nonlinear methods of interpolation are models of data approximation
  - **Computes curves which exactly pass through data points or use the data to control aspects of the curve**
  - **LERP, BERP, TERP, SLERP, Splines and lagrange**

Different ways of modeling based on data

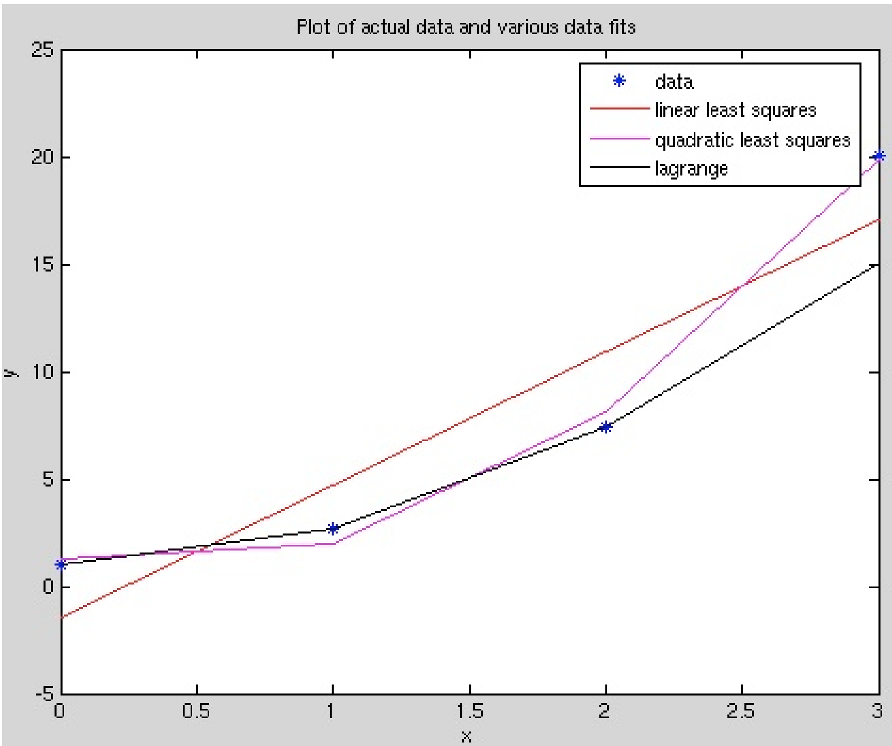
- Record all your data, then create a fit and study the resulting model
- Record all your data, split the recorded data into different groups
  - **use one group to fit a model**
  - **then the other to check and see how well does your model predict what the system does (this is called model validation - or 'invalidation')**

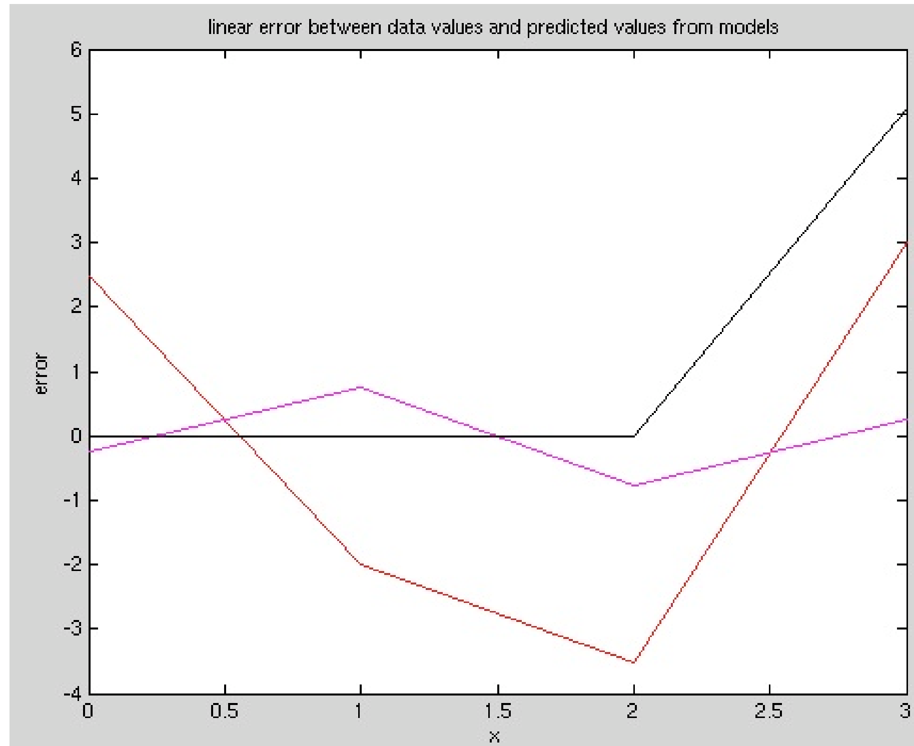
Example from the last few classes of computing error

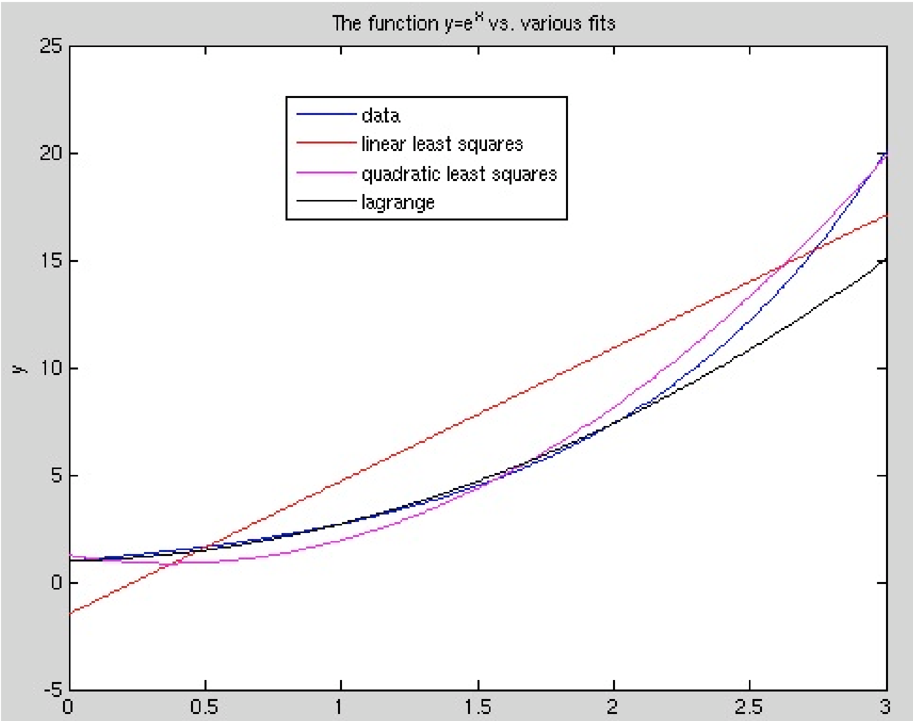
- Approximation of $y = e^x$ using the various methods we know already
- I generated simulated data by computing y=exp(x) for a domain of [0, 1]at three points (0.0, 0.5, 1.0)
  - **Then I created a linear least squares fit, a quadratic least squares fit, and Lagrange fits**

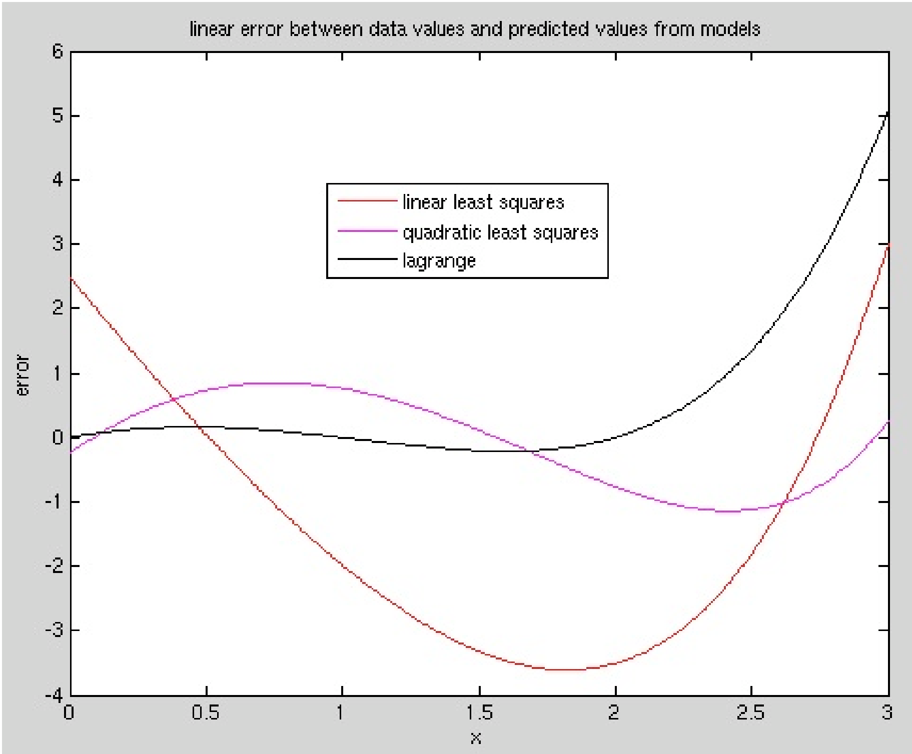
Assessing the models

- I assess how well each model fit does by first plotting the error between the data and the different methods
- Then I plot the real function (or data) vs. the different methods along a continuous curve


Plot of actual data and various data fits
data
linear least squares
quadratic least squares
lagrange
y
x
25
20
15
10
5
0
-5
0
0.5
1
1.5
2
2.5
3


linear error between data values and predicted values from models
error
x
6
5
4
3
2
1
0
-1
-2
-3
-4
0
0.5
1
1.5
2
2.5
3


The function y=e^x vs. various fits
data
linear least squares
quadratic least squares
lagrange
y
x


linear error between data values and predicted values from models
linear least squares
quadratic least squares
lagrange
error
x

We can also compute the error as a single quantity

- e2lls = 125.7192

- e2nlls = 10.9367

- e2lag = 86.3331
- From this we see that over this interval, the nonlinear least squares polynomial fits the data the best if we're trying to minimize this error as a criterion for goodness of fit
- Again it depends on our criterion, as the lagrange has the lowest error over the domain of data used for computing the fit
  - **It doesn't extrapolate the future points as well in this case**