Math 636 – Mathematical Modeling

Lecture Notes – Least Squares

Joseph M. Mahaffy,
⟨jmahaffy@mail.sdsu.edu⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

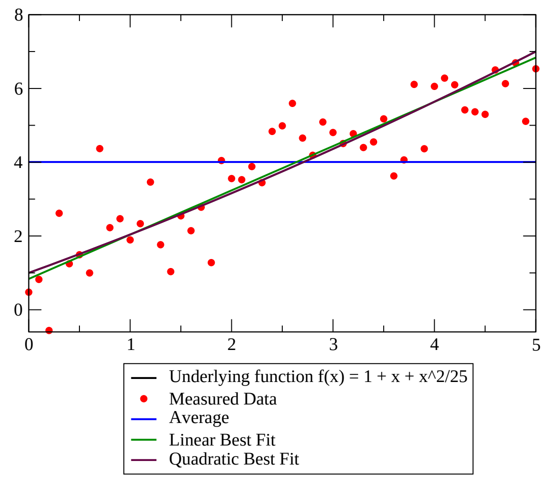
http://www-rohan.sdsu.edu/~jmahaffy

Fall 2014

Outline

Introduction: Matching a Few Parameters to a Lot of Data.

Sometimes we get a lot of data, ***many observations***, and want to fit it to a simple model.



PDF-link: code.

Why a Low Dimensional Model?

Low dimensional models (*e.g.* low degree polynomials) are **easy to work with**, and are quite **well behaved** (high degree polynomials can be quite oscillatory.)

All measurements are **noisy**, to some degree. Often, we want to use a large number of measurements in order to "average out" random noise.

Approximation Theory looks at two problems:

[1] Given a data set, find the best fit for a model (*i.e.* in a class of functions, find the one that best represents the data.)

[2] Find a simpler model approximating a given function.

Discrete Least Squares: Linear Approximation.

The form of Least Squares you are most likely to see: **Find the Linear Function, $p_1(x) = a_0 + a_1 x$, that best fits the data.** The error $E(a_0, a_1)$ we need to minimize is:

$$E(a_0, a_1) = \sum_{i=0}^{n} \left[(a_0 + a_1 x_i) - y_i\right]^2.$$

The first partial derivatives with respect to a_0 and a_1 better be zero at the minimum:

$$\begin{aligned} \frac{\partial}{\partial a_0} E(a_0, a_1) &= 2\sum_{i=0}^{n} \left[(a_0 + a_1 x_i) - y_i\right] &= 0 \\ \frac{\partial}{\partial a_1} E(a_0, a_1) &= 2\sum_{i=0}^{n} x_i \left[(a_0 + a_1 x_i) - y_i\right] &= 0. \end{aligned}$$

We "massage" these expressions to get the **Normal Equations**...

Linear Approximation: The Normal Equations $p_1(x)$

$$\begin{cases} \mathbf{a_0}(n+1) + \mathbf{a_1}\sum_{i=0}^{n} x_i = \sum_{i=0}^{n} y_i \\ \mathbf{a_0}\sum_{i=0}^{n} x_i + \mathbf{a_1}\sum_{i=0}^{n} x_i^2 = \sum_{i=0}^{n} x_i y_i. \end{cases}$$

Since **everything except** $\mathbf{a_0}$ and $\mathbf{a_1}$ is known, this is a 2-by-2 system of equations.

$$\begin{bmatrix} (n+1) & \sum_{i=0}^{n} x_i \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \mathbf{a_0} \\ \mathbf{a_1} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \end{bmatrix}.$$

Quadratic Model, $p_2(x)$

For the quadratic polynomial $p_2(x) = a_0 + a_1 x + a_2 x^2$, the error is given by

$$E(a_0, a_1, a_2) = \sum_{i=0}^{n} \left[a_0 + a_1 x_i + a_2 x_i^2 - y_i\right]^2$$

At the minimum (best model) we must have

$$\begin{aligned} \frac{\partial}{\partial a_0} E(a_0, a_1, a_2) &= 2\sum_{i=0}^{n} \left[(a_0 + a_1 x_i + a_2 x_i^2) - y_i\right] &= 0 \\ \frac{\partial}{\partial a_1} E(a_0, a_1, a_2) &= 2\sum_{i=0}^{n} x_i \left[(a_0 + a_1 x_i + a_2 x_i^2) - y_i\right] &= 0 \\ \frac{\partial}{\partial a_2} E(a_0, a_1, a_2) &= 2\sum_{i=0}^{n} x_i^2 \left[(a_0 + a_1 x_i + a_2 x_i^2) - y_i\right] &= 0. \end{aligned}$$

Quadratic Model: The Normal Equations $p_2(x)$

Similarly for the quadratic polynomial $p_2(x) = a_0 + a_1 x + a_2 x^2$, the normal equations are:

$$\begin{cases} \mathbf{a_0}(n+1) + \mathbf{a_1}\sum_{i=0}^{n} x_i + \mathbf{a_2}\sum_{i=0}^{n} x_i^2 = \sum_{i=0}^{n} y_i \\ \mathbf{a_0}\sum_{i=0}^{n} x_i + \mathbf{a_1}\sum_{i=0}^{n} x_i^2 + \mathbf{a_2}\sum_{i=0}^{n} x_i^3 = \sum_{i=0}^{n} x_i y_i. \\ \mathbf{a_0}\sum_{i=0}^{n} x_i^2 + \mathbf{a_1}\sum_{i=0}^{n} x_i^3 + \mathbf{a_2}\sum_{i=0}^{n} x_i^4 = \sum_{i=0}^{n} x_i^2 y_i. \end{cases}$$

Note: Even though the model is quadratic, the resulting (normal) equations are **linear**. — The model is linear in its parameters, a_0, a_1, and a_2.

The Normal Equations — As Matrix Equations.

We rewrite the Normal Equations as:

$$\begin{bmatrix} (n+1) & \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 \\ \sum_{i=0}^n x_i^2 & \sum_{i=0}^n x_i^3 & \sum_{i=0}^n x_i^4 \end{bmatrix} \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n x_i y_i. \\ \sum_{i=0}^n x_i^2 y_i. \end{bmatrix}.$$

It is not immediately obvious, but this expression can be written in the form $\mathbf{A^T A \tilde{a} = A^T \tilde{y}}$. Where the matrix A is very easy to write in terms of x_i. [Jump Forward].

SDSU

Discrete Least Squares: A Simple, Powerful Method.

Given the data set $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, where $\tilde{\mathbf{x}} = \{x_0, x_1, \ldots, x_n\}^T$ and $\tilde{\mathbf{y}} = \{y_0, y_1, \ldots, y_n\}^T$, we can quickly find the best polynomial fit for **any** specified polynomial degree!

Notation: Let $\tilde{\mathbf{x}}^j$ be the vector $\{x_0^j, x_1^j, \ldots, x_n^j\}^T$.

E.g. to compute the best fitting polynomial of degree 3, $p_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$, define:

$$A = \begin{bmatrix} | & | & | & | \\ \tilde{\mathbf{1}} & \tilde{\mathbf{x}} & \tilde{\mathbf{x}}^2 & \tilde{\mathbf{x}}^3 \\ | & | & | & | \end{bmatrix}, \quad \text{and compute} \quad \tilde{\mathbf{a}} = (A^T A)^{-1}(A^T \tilde{\mathbf{y}}).$$

See Numerical Analysis (Math 541) to solve this equation

This is solvable in MatLab (See **polyfit**)

SDSU

Snowy Tree Cricket

SDSU

Chirping Crickets and Temperature

- Folk method for finding temperature (Fahrenheit)
  Count the number of chirps in a minute and divide by 4, then add 40
- In 1898, A. E. Dolbear [3] noted that
  "crickets in a field [chirp] synchronously, keeping time as if led by the wand of a conductor"
- He wrote down a formula in a scientific publication (first?)

$$T = 50 + \frac{N - 40}{4}$$

[3] A. E. Dolbear, The cricket as a thermometer, American Naturalist (1897) 31, 970-971

SDSU

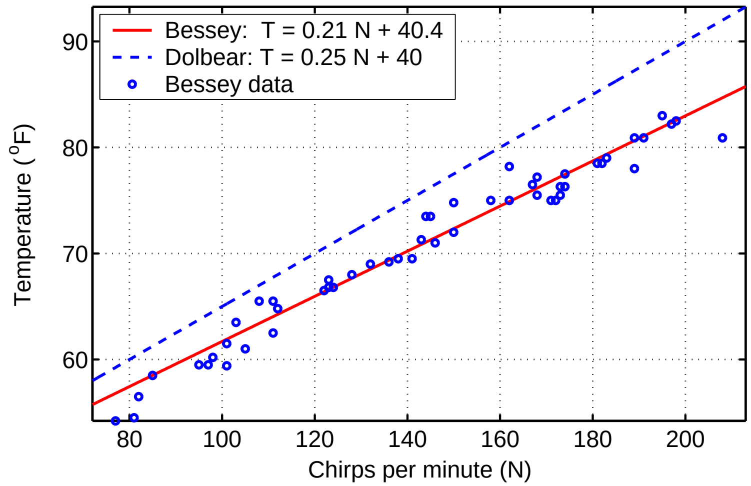
Data Fitting Linear Model

- Mathematical models for chirping of snowy tree crickets, *Oecanthulus fultoni*, are **Linear Models**
- Data from C. A. Bessey and E. A. Bessey [2] (8 crickets) from Lincoln, Nebraska during August and September, 1897 (shown on next slide)
- The least squares best fit line to the data is

$$T = 60 + \frac{N - 92}{4.7}$$

[2] C. A. Bessey and E. A. Bessey, Further notes on thermometer crickets, American Naturalist (1898) 32, 263-264

Bessey Data and Linear Models



Model

Biological Questions – Cricket Model 1

How well does the line fitting the Bessey & Bessey data agree with the Dolbear model given above?

- Graph shows Linear model fits the data well
- Data predominantly below **Folk/Dolbear** model
- Possible discrepancies
  - Different cricket species
  - Regional variation
  - Folk only an approximation
- Graph shows only a few °F difference between models

Biological Questions – Cricket Model 2

When can this model be applied from a practical perspective?

- Biological thermometer has limited use
- Snowy tree crickets only chirp for a couple months of the year and mostly at night
- Temperature needs to be above 50°F

Mathematical Questions – Cricket Model 1

Over what range of temperatures is this model valid?

- Biologically, observations are mostly between 50°F and 85°F
- Thus, limited **range** of temperatures, so limited **range** on the **Linear Model**
- **Range** of **Linear functions** affects its **Domain**
- From the graph, **Domain** is approximately 50–200 **Chirps/min**

Mathematical Questions – Cricket Model 2

How accurate is the model and how might the accuracy be improved?

- Data closely surrounds **Bessey Model** – No more than about 3°F away fom line
- **Dolbear Model** is fairly close though not as accurate – Sufficient for rapid temperature estimate
- Observe that the temperature data trends lower at higher chirp rates – compared against linear model
- Better fit with **Quadratic function** – Is this really significant?

Cricket Data Analysis

C. A. Bessey and E. A. Bessey collected data on eight different crickets that they observed in Lincoln, Nebraska during August and September, 1897. The number of chirps/min was N and the temperature was T.

Create matrices

$$A_1 = \begin{pmatrix} 1 & N_1 \\ 1 & N_2 \\ \vdots & \vdots \end{pmatrix} \qquad A_2 = \begin{pmatrix} 1 & N_1 & N_1^2 \\ 1 & N_2 & N_2^2 \\ \vdots & \vdots & \vdots \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 1 & N_1 & N_1^2 & N_1^3 \\ 1 & N_2 & N_2^2 & N_2^3 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \qquad A_4 = \begin{pmatrix} 1 & N_1 & N_1^2 & N_1^3 & N_1^4 \\ 1 & N_2 & N_2^2 & N_2^3 & N_2^4 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Cricket Linear Model

If you compute the matrix *which you never should!*

$$A_1^T A_1 = \begin{pmatrix} 52 & 7447 \\ 7447 & 1133259 \end{pmatrix},$$

it has eigenvalues

$$\lambda_1 = 3.0633 \quad \text{and} \quad \lambda_2 = 1,133,308,$$

which gives the condition number

$$cond(A_1^T A_1) = 3.6996 \times 10^5.$$

Whereas

$$\texttt{cond}(A_1) = 608.2462.$$

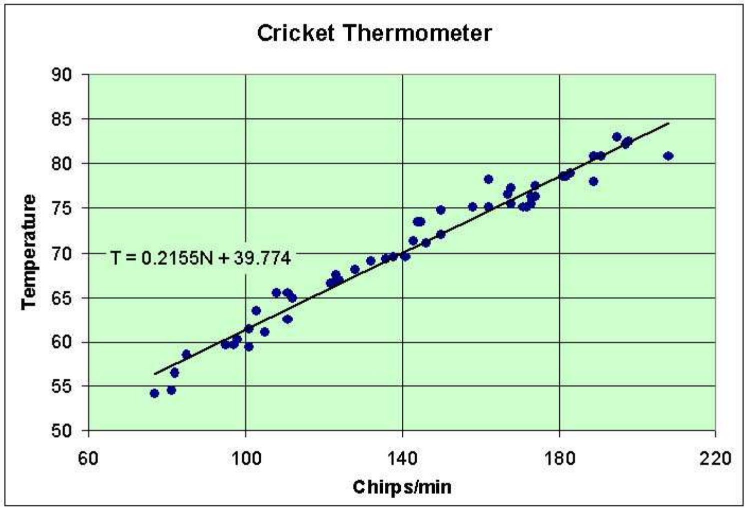
In Matlab

```
A1\T
```

gives the parameters for best linear model

$$T_1(N) = 0.2155\,N + 39.7441.$$

Polynomial Fits to the Data: Linear



Linear Fit

Cricket Quadratic Model

Similarly, the matrix

$$A_2^T A_2 = \begin{pmatrix} 52 & 7447 & 1133259 \\ 7447 & 1133259 & 1.8113\times 10^8 \\ 1133259 & 1.8113\times 10^8 & 3.0084\times 10^1 0 \end{pmatrix},$$

has eigenvalues

$$\lambda_1 = 0.1957, \quad \lambda_2 = 42,706, \quad \lambda_3 = 3.00853\times 10^{10}$$

which gives the condition number

$$cond(A_2^T A_2) = 1.5371\times 10^{11}.$$

Whereas,

$$\texttt{cond}(A_2) = 3.9206\times 10^5,$$

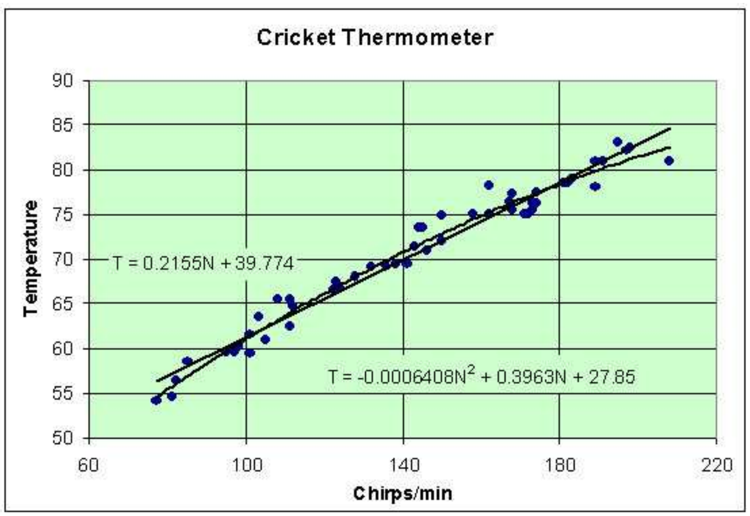
and

$$\texttt{A}_2\backslash\texttt{T},$$

gives the parameters for best quadratic model

$$T_2(N) = -0.00064076\,N^2 + 0.39625\,N + 27.8489.$$

Polynomial Fits to the Data: Quadratic



Quadratic Fit

Cricket Cubic and Quartic Models

The condition numbers for the cubic and quartic rapidly get larger with

$$cond(A_3^T A_3) = 6.3648\times 10^{16} \quad \text{and} \quad cond(A_4^T A_4) = 1.1218\times 10^{23}$$

These last two condition numbers suggest that any coefficients obtained are highly suspect.

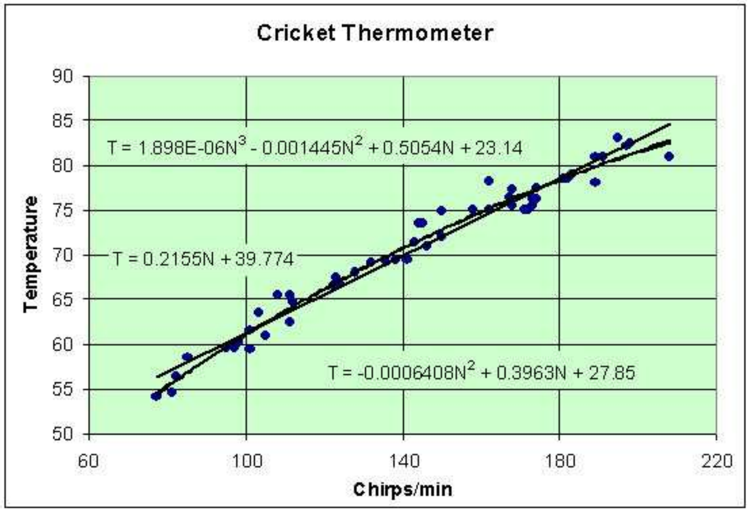
However, if done right, we are "only" subject to the conditon numbers

$$\texttt{cond}(A_3) = 2.522\times 10^8, \qquad \texttt{cond}(A_4) = 1.738\times 10^{11}.$$

The best cubic and quartic models are given by

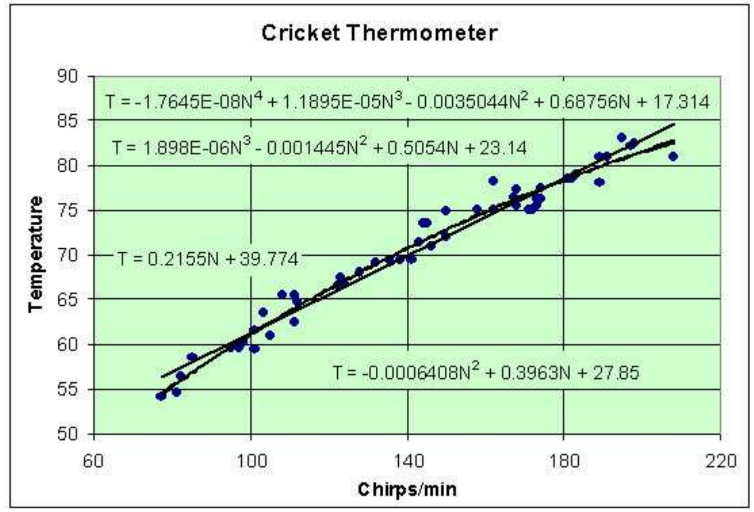
$$\begin{aligned} T_3(N) &= 0.0000018977\,N^3 - 0.001445\,N^2 + 0.50540\,N + 23.138 \\ T_4(N) &= -0.00000001765\,N^4 + 0.00001190\,N^3 - 0.003504\,N^2 \\ &= +0.6876\,N + 17.314 \end{aligned}$$

Polynomial Fits to the Data: Cubic



Cubic Fit

Polynomial Fits to the Data: Quartic



Quartic Fit

Best Cricket Model

So how does one select the best model?

Visually, one can see that the linear model does a very good job, and one only obtains a slight improvement with a quadratic. Is it worth the added complication for the slight improvement.

It is clear that the sum of square errors (SSE) will improve as the number of parameters increase.

From statistics, it is hotly debated how much penalty one should pay for adding parameters.

Best Cricket Model - Analysis

Bayesian Information Criterion

Let n be the number of data points, SSE be the sum of square errors, and let k be the number of parameters in the model.

$$BIC = n\ln(SSE/n) + k\ln(n).$$

Akaike Information Criterion

$$AIC = 2k + n(\ln(2\pi SSE/n) + 1).$$

Best Cricket Model - Analysis Continued

The table below shows the by the Akaike information criterion that we should take a quadratic model, while using a Bayesian Information Criterion we should use a cubic model.

	Linear	Quadratic	Cubic	Quartic
SSE	108.8	79.08	78.74	**78.70**
BIC	46.3	33.65	**33.43**	37.35
AIC	189.97	**175.37**	177.14	179.12

Returning to the original statement, we do fairly well by using the folk formula, despite the rest of this analysis!