CS321-002

Introduction to Numerical Methods

Lecture 7

Least Squares and Curve Fitting

Professor Jun Zhang

Department of Computer Science
University of Kentucky
Lexington, KY 40506-0046

November 28, 2000

Method of Least Squares

computer aided data collections have produced tremendous amount of data that are impossible to understand without some sort of post-processing

given a set of data

x	x_0	x_1	$\cdots$	x_m
y	y_0	y_1	$\cdots$	y_m

if we assume that the data form a linear functional relation, we can write the function as

$$y = a\,x + b$$

with the coefficients a and b to be determined

for each pairs (x_i, y_i) we can request

$$y_i = ax_i + b$$

for $i = 0, 1, \ldots, m$. Note that, if $m > 1$, we have more than two linear equations to determine just two unknowns

Over-Determined System

in general, we have more equations than unknowns. There is no *exact solution* to the problem. However, we could determine a solution that minimizes the total error

suppose the linear equation is given as

$$y = a\,x + b$$

if the point (x_i, y_i) is on the straight line defined by the function, we have

$$y_i - (a\,x_i + b) = 0$$

in most cases, a point (x_i, y_i) is not on the line, we have

$$r_i = y_i - (a\,x_i + b) \neq 0$$

for $i = 0, 1, 2, \ldots, m$. $|r_i|$ is the curve fitting error

Minimizing Total Error

one can reason that if the sum of the errors is minimized, the data should fit the line as best as it can

the total error can be represented as

$$\sum_{k=0}^{m} |r_i| = \sum_{k=0}^{m} |ax_k + b - y_k|$$

we can minimize the above functional to select the coefficients a and b. This can be solved by the techniques of linear programming

this is an l_1 approximation (1-norm approximation)

the shortcoming of this formulation is that the function of the total error is not differentiable. Many tools in calculus may not be used

Method of Least Squares

an alternative is to minimize a different error function

$$\phi(a,b) = \sum_{k=0}^{m} (a\,x_k + b - y_k)^2$$

which is continuously differentiable

this is also called an l_2 approximation. It is a special case of the l_p approximation with the l_p norm is defined as

$$\|x\|_p = \left\{ \sum_{i=1}^{n} |x_i|^p \right\}^{1/p} \qquad (1 \le p < \infty)$$

where $x = (x_1, x_2, \ldots, x_n)^T$ is an n vector

from statistical considerations, if the errors follow a normal probability distribution, the minimization of ϕ produces a best estimate of a and b

How to Compute a Minimum

we use a technique in calculus to determine the extreme points of a function

$$\frac{\partial \phi}{\partial a} = 0 \qquad \frac{\partial \phi}{\partial b} = 0$$

this gives us two equations

$$\begin{aligned} \sum_{k=0}^{m} 2(a\,x_k + b - y_k)x_k &= 0 \\ \sum_{k=0}^{m} 2(a\,x_k + b - y_k) &= 0 \end{aligned}$$

they are called the *normal equations* and can be written explicitly as

$$\begin{aligned} \left(\sum_{k=0}^{m} x_k^2\right) a + \left(\sum_{k=0}^{m} x_k\right) b &= \sum_{k=0}^{m} y_k\, x_k \\ \left(\sum_{k=0}^{m} x_k\right) a + (m+1) b &= \sum_{k=0}^{m} y_k \end{aligned}$$

which can be solved for a and b

Least Squares Solution

the solution of the previous two-by-two linear system can be solved as

$$a = \frac{1}{d}\left[(m+1)\sum_{k=0}^{m} x_k y_k - \left(\sum_{k=0}^{m} x_k\right)\left(\sum_{k=0}^{m} y_k\right)\right]$$

$$b = \frac{1}{d}\left[\left(\sum_{k=0}^{m} x_k^2\right)\left(\sum_{k=0}^{m} y_k\right) - \left(\sum_{k=0}^{m} x_k\right)\left(\sum_{k=0}^{m} x_k y_k\right)\right]$$

where

$$d = (m+1)\sum_{k=0}^{m} x_k^2 - \left(\sum_{k=0}^{m} x_k\right)^2$$

lots of simple computations

Linear Example

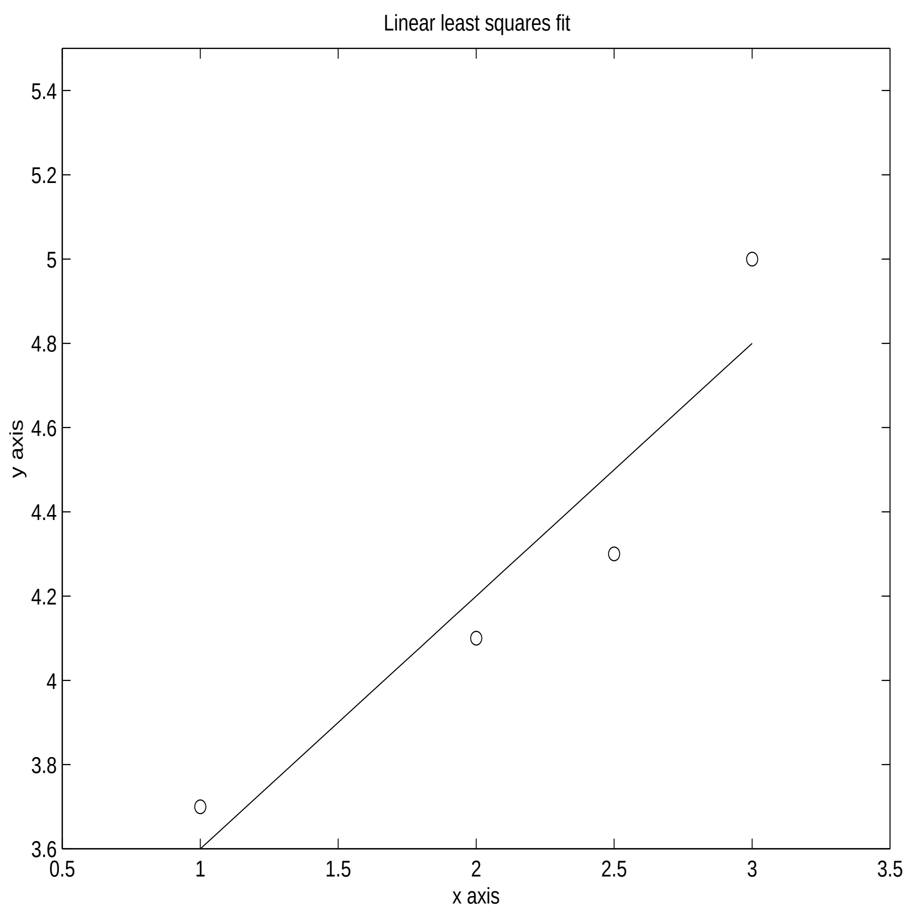
x	1.0	2.0	2.5	3.0
y	3.7	4.1	4.3	5.0

Least Squares Fit

the data in the previous slide lead to

$$\begin{aligned} 20.25a + 8.5b &= 37.65 \\ 8.5a + 4.0b &= 17.1 \end{aligned}$$

which can be solved for $a = 0.6$ and $b = 3.0$

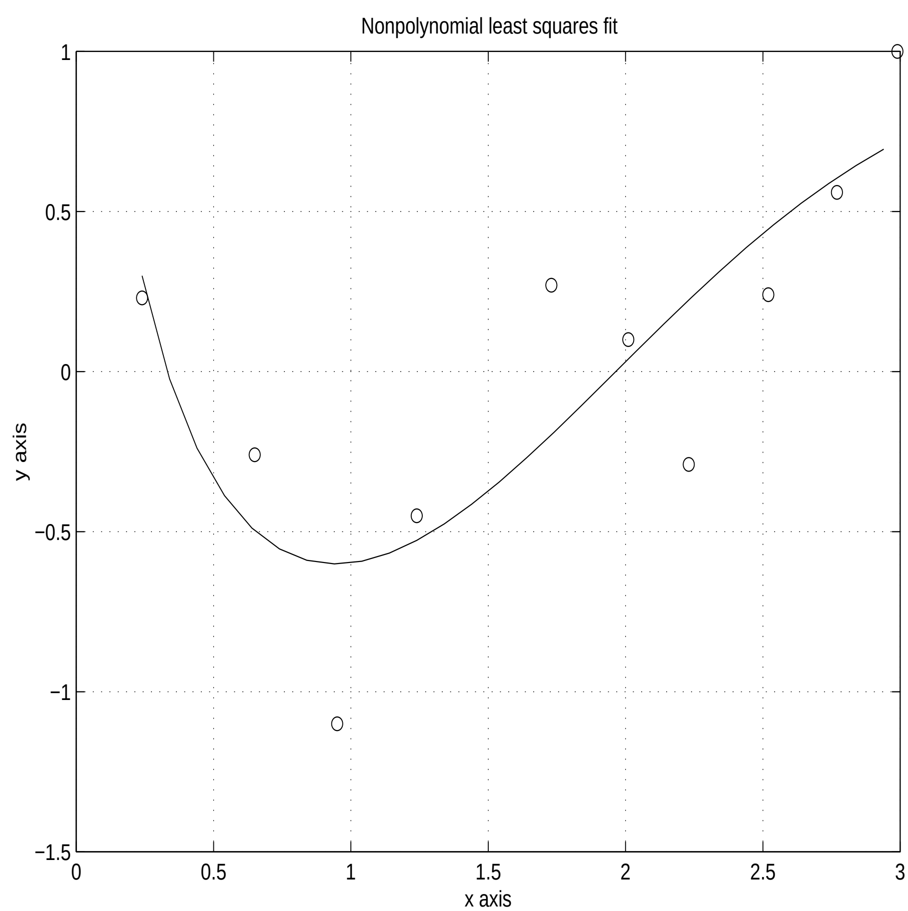


Nonpolynomial Example

we can fit a table by a nonpolynomial function

x	0.24	0.56	0.95	1.24	$\cdots$	2.77	2.99
y	0.23	-0.26	-1.10	-0.45	$\cdots$	0.56	1.00

$$a \ln x + b \cos x + ce^x$$



Basis Functions

a general least squares fitting can be written as

$$y = \sum_{j=0}^{n} c_j \, g_j(x)$$

in which the functions $g_0(x), g_1(x), \ldots, g_n(x)$ are called basis functions. They are known and kept fixed

given a set of data (x_k, y_k), we want to find the values of c_j to minimize the total error as

$$\phi(c_0, c_1, \ldots, c_n) = \sum_{k=0}^{m} \left[\sum_{j=0}^{n} c_j \, g_j(x_k) - y_k \right]^2$$

we again set the partial derivatives to be zero

$$\frac{\partial \phi}{\partial c_i} = 0 \qquad (0 \le i \le n)$$

Basis Functions (II)

the partial derivatives are for $0 \le i \le n$

$$\frac{\partial \phi}{\partial c_i} = \sum_{k=0}^{m} 2 \left[\sum_{j=0}^{n} c_j \, g_j(x_k) - y_k \right] g_i(x_k)$$

setting them simultaneously zero, we have

$$\sum_{j=0}^{n} \left[\sum_{k=0}^{m} g_i(x_k) \, g_j(x_k) \right] c_j = \sum_{k=0}^{m} y_k \, g_i(x_k)$$

this is the normal equation, which is a system of linear equations with unknowns $c_0, c_1, \ldots, c_n$

the coefficients of the linear system are

$$a_{ij} = \sum_{k=0}^{m} g_i(x_k) \, g_j(x_k)$$

the coefficient matrix is nonsingular if the basis function $g_0(x), g_1(x), \ldots, g_n(x)$ are linearly independent. The basis functions should be appropriate for the problem in question and make the resulting coefficient matrix well conditioned