305233
Numerical Methods

Curve Fitting

Assoc. Prof. Tanit Malakorn, PhD
Department of Electrical and Computer Engineering
Faculty of Engineering
Naresuan University

1

Curve Fitting

- ในการวิเคราะห์ข้อมูล วิธีหนึ่งคือการหาฟังก์ชันหรือสมการที่สามารถใช้บรรยายข้อมูลเหล่านั้นซึ่งวิธีดังกล่าวนี้เรียกว่า Curve Fitting
- Curve Fitting มีใช้อยู่ 2 กลุ่มขึ้นกับลักษณะของข้อมูล
  - ถ้าข้อมูลมีการกระจายตัว มีความผิดพลาดในการวัด หรือมีสิ่งรบกวนปนมากับข้อมูลที่วัดได้ นิยมเลือกสร้างฟังก์ชันหรือสมการที่ใช้บรรยายพฤติกรรมของข้อมูลในภาพรวมเพื่อดูแนวโน้มของข้อมูล เมื่อนำสมการมาวาดกราฟ กราฟที่ได้จะผ่านกลางย่านของข้อมูลโดยไม่จำเป็นต้องผ่านทุกจุด วิธีการหนึ่งที่นิยมใช้คือ Least-squares regression
  - ถ้าข้อมูลที่ได้มีความถูกต้อง อาจจะเป็นค่าที่เกิดจากผลการทดลองที่แม่นยำสูง และต้องการฟังก์ชันหรือสมการที่เมื่อนำมาวาดกราฟแล้ว กราฟที่ได้จะผ่านทุกจุดของข้อมูล วิธีการสร้างฟังก์ชันในลักษณะนี้เรียกว่า Interpolation

2

Least-squares regression

วิธีการของ Least-squares regression ที่ง่ายที่สุดคือการสร้างสมการเชิงเส้นที่อยู่ในรูปแบบของ $y_{approx} = a_1x + a_0$ หากกำหนดให้ e คือค่าความผิดพลาดระหว่างข้อมูลจริง y และค่าที่ได้จากสมการ y_{approx} นั่นคือ

$$e = y - y_{approx} = y - (a_1x + a_0)$$

เพื่อให้การประมาณมีค่าดีที่สุด จึงต้องออกแบบหาค่า a_0 และ a_1 ที่ทำให้ e มีค่าน้อยที่สุดในความหมายที่ว่า ผลรวมของค่าผิดพลาด (ที่แต่ละจุดข้อมูล) กำลังสองเฉลี่ยน้อยที่สุด นั่นคือ

$$\min \sum_{i=1}^{n} e_i^2 = \min \sum_{i=1}^{n} (y_i - a_1x_i - a_0)^2$$

Tanit MALAKORN, PhD

3

Least-squares regression

เนื่องจากฟังก์ชันที่ต้องการหาค่าต่ำสุดมี 2 ตัวแปรที่ไม่ทราบค่า นั่นคือ a_0 และ a_1 ดังนั้นจึงต้องหาอนุพันธ์เทียบตัวแปรแต่ละตัว นั่นคือ

$$\frac{\partial S_r}{\partial a_0} = \frac{\partial \sum (y_i - a_0 - a_1x_i)^2}{\partial (y_i - a_0 - a_1x_i)} \frac{\partial (y_i - a_0 - a_1x_i)}{\partial a_0} = -2\sum (y_i - a_0 - a_1x_i)$$

$$\frac{\partial S_r}{\partial a_1} = \frac{\partial \sum (y_i - a_0 - a_1x_i)^2}{\partial (y_i - a_0 - a_1x_i)} \frac{\partial (y_i - a_0 - a_1x_i)}{\partial a_1} = -2\sum [(y_i - a_0 - a_1x_i)x_i]$$

เมื่อให้ค่าอนุพันธ์ที่ได้มีค่าเป็นศูนย์ และเนื่องจาก $\sum a_0 = na_0$ จึงได้ว่า

$$na_0 + \left(\sum x_i\right)a_1 = \sum y_i$$

$$\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 = \sum x_i y_i$$

Tanit MALAKORN, PhD

4

Least-squares regression

จึงได้ว่า

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad \& \quad a_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

นิสิตสามารถพิสูจน์ได้ว่า

$$a_0 = \bar{y} - a_1\bar{x}$$

เมื่อ $\bar{y}, \bar{x}$ คือค่าเฉลี่ยของ y, x ตามลำดับ

Tanit MALAKORN, PhD

Department of Electrical and Computer Engineering
Naresuan University, THAILAND

5

Least-squares regression

นำ a_0 และ a_1 มาแทนลงในสมการ regression : $y_{approx} = a_1 x + a_0$

- ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) ของข้อมูลเมื่อเทียบกับค่าเฉลี่ยของข้อมูล คือ

$$S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) ของข้อมูลเมื่อเทียบกับค่าที่ประมาณได้จาก LSq คือ

$$S_r = \sum_{i=1}^{n} [y_i - (a_0 + a_1 x_i)]^2$$

- สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) คำนวณจากสูตร

$$r = \sqrt{\frac{S_t - S_r}{S_t}}$$

Tanit MALAKORN, PhD

Department of Electrical and Computer Engineering
Naresuan University, THAILAND

6

Least-squares regression

สัมประสิทธิ์สหสัมพันธ์ (r) บ่งบอกถึงความใกล้เคียงกันระหว่าง 2 ชุดข้อมูล โดยค่า r อยู่ระหว่าง 0 และ 1

- r = 1 แสดงว่าชุดข้อมูลทั้งสองเหมือนกัน
- r = 0 แสดงว่าชุดข้อมูลทั้งสองต่างกัน

ตัวอย่าง จากชุดข้อมูลที่กำหนดให้ จงหาสมการเชิงเส้นที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression พร้อมหาค่า corr. coeff.

i	1	2	3	4	5	6	7	8
x_i	10	20	30	40	50	60	70	80
y_i	25	70	380	550	610	1,220	830	1,450

Tanit MALAKORN, PhD
Department of Electrical and Computer Engineering
Naresuan University, THAILAND

Least-squares regression

วิธีทำ จากชุดข้อมูลที่กำหนดให้ นำมาคำนวณหา x^2 และ xy เพื่อนำไปแทนในสูตรของ a_1

i	1	2	3	4	5	6	7	8	SUM
x_i	10	20	30	40	50	60	70	80	360
y_i	25	70	380	550	610	1,220	830	1,450	5,135
$(x_i)^2$	100	400	900	1,600	2,500	3,600	4,900	6,400	20,400
x_iy_i	250	1,400	11,400	22,000	30,500	73,200	58,100	116,000	312,850

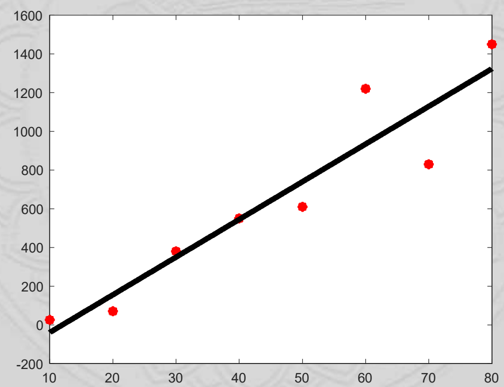
$$a_1 = \frac{8(312{,}850) - (360)(5{,}135)}{8(20{,}400) - (360)^2} = 19.47024,$$

$$\bar{x} = \frac{360}{8} = 45,\ \bar{y} = \frac{5{,}135}{8} = 641.875 \text{ และ } a_0 = \bar{y} - a_1\bar{x} = -234.2857$$

Tanit MALAKORN, PhD
Department of Electrical and Computer Engineering
Naresuan University, THAILAND

Least-squares regression

สมการเชิงเส้นที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression คือ $y_{approx} = a_1x + a_0 = (19.47024)x - 234.2857$



Tanit MALAKORN, PhD

Department of Electrical and Computer Engineering
Naresuan University, THAILAND

Least-squares regression

สำหรับค่า corr. coeff. ให้นำข้อมูลของ x แต่ละค่ามาแทนลงในสมการ regression ที่ประมาณมาได้เพื่อนำมาคำนวณหา y_{approx} แต่ละตัว แล้วนำมาแทนค่าเพื่อหา S_r

i	1	2	3	4	5	6	7	8	SUM
x_i	10	20	30	40	50	60	70	80	360
y_i	25	70	380	550	610	1,220	830	1,450	5,135
y_{approx}	-39.58	155.12	349.82	544.52	739.23	933.93	1,128.63	1,323.33	
$(y_i - \bar{y})^2$	380,535	327,041	68,579	8,441	1,016	334,229	35,391	653,066	1,808,297
$(y_i - y_{ap})^2$	4,171	7,245	911	30	16,699	81,837	89,180	16,044	216,118

$$r = \sqrt{\frac{S_t - S_r}{S_t}} = \sqrt{\frac{1{,}808{,}297 \ -216{,}118}{1{,}808{,}297}} = \sqrt{0.8804853} = 0.93834$$

Tanit MALAKORN, PhD

Department of Electrical and Computer Engineering
Naresuan University, THAILAND

Least-squares regression

ตัวอย่าง จากชุดข้อมูลที่กำหนดให้ จงหาสมการเชิงเส้นที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression พร้อมหาค่า corr. coeff.

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	0.7	2.2	2.8	4.4	4.9

วิธีทำ จากชุดข้อมูลที่กำหนดให้ นำมาคำนวณหา x^2 และ xy เพื่อนำไปแทนในสูตรของ a_1

i	1	2	3	4	5	SUM
x_i	1	2	3	4	5	15
y_i	0.7	2.2	2.8	4.4	4.9	15
$(x_i)^2$	1	4	9	16	25	55
x_iy_i	0.7	4.4	8.4	17.6	24.5	55.6

MALAKORN, PhD

Department of Electrical and Computer Engineering
Naresuan University, THAILAND

11

11

Least-squares regression

$a_1 = \dfrac{5(55.6) - (15)(15)}{5(55) - (55)^2} = 1.06$, $\bar{x} = \dfrac{15}{5} = 3$, $\bar{y} = \dfrac{15}{5} = 3$ และ $a_0 = \bar{y} - a_1\bar{x} = -0.18$

สมการเชิงเส้นที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression คือ

$y_{approx} = a_1x + a_0 = (1.06)x - 0.18$

สำหรับค่า corr. coeff. ให้นำข้อมูลของ x แต่ละค่ามาแทนลงในสมการ regression ที่ประมาณมาได้เพื่อนำมาคำนวณหา y_{approx} แต่ละตัว แล้วนำมาแทนค่าเพื่อหา S_r

i	1	2	3	4	5	SUM
x_i	1	2	3	4	5	15
y_i	0.7	2.2	2.8	4.4	4.9	15
y_{approx}	0.88	1.94	3.00	4.06	5.12	
$(y_i - \bar{y})^2$	5.29	0.64	0.04	1.96	3.61	11.54
$(y_i - y_{ap})^2$	0.0324	0.0676	0.04	0.1156	0.0484	0.304

MALAKORN, PhD

Dep
Nare

12

12

Least-squares regression

$$r = \sqrt{\frac{S_t - S_r}{S_t}} = \sqrt{\frac{11.54 \quad -0.304}{11.54}} = \sqrt{0.973657} = 0.98674$$

บ่อยครั้งพบว่า ข้อมูลที่ได้มานั้นไม่ได้แปรผันตรงกับตัวแปรต้น ดังนั้นเมื่อประมาณด้วยสมการเชิงเส้นจึงทำให้เกิดความผิดพลาดสูง จึงจำเป็นต้องศึกษาการประมาณแบบ Least-squares ในรูปแบบอื่นที่ไม่ได้อยู่ในรูปเชิงเส้น ได้แก่

- Polynomial function นั่นคือเขียนสมการให้อยู่ในรูปของฟังก์ชันพนุนามกำลังมากกว่า 1
- Trigonometric function นั่นคือเขียนสมการอยู่ในรูปของผลรวมเชิงเส้นของ sin และ cos
- Exponential function
- ฯลฯ

ในการเลือกรูปแบบของสมการที่เหมาะสมนั้น จำเป็นนำชุดข้อมูลมา Plot จุดเพื่อดูแนวโน้มว่า สมการที่ควรนำมา fit curve ควรอยู่ในรูปแบบใด

Least-squares regression

สมการไม่เชิงเส้นบางรูปแบบสามารถเปลี่ยนมาให้อยู่ในรูปของสมการเชิงเส้นได้ เมื่อสร้างฟังก์ชันการแปลงที่เหมาะสมบ่อยครั้ง ยกตัวอย่างเช่น

- $y = Ax^B$ สามารถแปลงได้เป็น $\log_{10} y = \log_{10} A + B(\log_{10} x)$ ซึ่งอยู่ในรูปแบบของเชิงเส้น $y = a_1x + a_0$ ในการทำ Linear Least-squares ให้เลือกใช้ $\log_{10} y$ แทน y เลือก $\log_{10} x$ แทน x และเมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์ $B = a_1$ และ $A = 10^{a_0}$ (นิสิตสามารถเลือกฟังก์ชัน logarithm ฐานอื่นได้ตามความชอบใจ เช่น ฐาน 2, ฐาน 3 หรือ ฐาน e)
- $y = Ae^{Bx}$ สามารถแปลงได้เป็น $\log y = \log A + Bx$ ซึ่งอยู่ในรูปแบบของเชิงเส้น $y = a_1x + a_0$ ในการทำ Linear Least-squares ให้เลือกใช้ log y แทน y ส่วนข้อมูล x ไม่ต้องแปลง เมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์ $B = a_1$ และ $A = e^{a_0}$ (ในกรณีนี้ นิสิตไม่สามารถเลือกฟังก์ชัน logarithm ฐานอื่นได้ ต้องเป็น ฐาน e เท่านั้น)

Least-squares regression

- $y = \dfrac{Ax}{B + x}$ ในกรณีนี้จะสามารถแปลงได้เมื่อ $Ax \neq 0$ และ $y \neq 0$ จะได้ว่า

$$\frac{1}{y} = \frac{B+x}{Ax} = \frac{B}{Ax} + \frac{1}{A} = \frac{B}{A}\left(\frac{1}{x}\right) + \frac{1}{A}$$

ซึ่งอยู่ในรูปแบบของเชิงเส้น $y = a_1x + a_0$ ในการทำ Linear Least-squares ให้เลือกใช้ $1/y$ แทน y เลือก $1/x$ แทน x และเมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์ $B = a_1/a_0$ และ $A = 1/a_0$

- $y = \dfrac{A}{B + x}$ ในกรณีนี้จะสามารถแปลงได้เมื่อ $A \neq 0$ และ $y \neq 0$ จะได้ว่า

$$\frac{1}{y} = \frac{B+x}{A} = \frac{1}{A}x + \frac{B}{A}$$

ซึ่งอยู่ในรูปแบบของเชิงเส้น $y = a_1x + a_0$ ในการทำ Linear Least-squares ให้เลือกใช้ $1/y$ แทน y ส่วนข้อมูล x ไม่ต้องแปลง เมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์ $B = a_0/a_1$ และ $A = 1/a_1$

15

Least-squares regression

ตัวอย่าง จากชุดข้อมูลที่กำหนดให้ จงหาสมการที่อยู่ในรูปของ $y = Ae^{-Bx}$ ที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression

i	1	2	3	4
x_i	0	2	4	6
y_i	5.0	3.7	2.7	2.0

วิธีทำ โจทย์ต้องการ fit curve กับสมการไม่เชิงเส้นที่อยู่ในรูปของ $y = Ae^{-Bx}$ ดังนั้นจึงต้องทำการแปลงสมการดังกล่าวให้เป็นสมการเชิงเส้นด้วยการใช้ฟังก์ชันลอการิทึม ซึ่งจะได้ว่า

$$\log y = \log A - Bx$$

ในการทำ Linear Least-squares ให้เลือกใช้ $\log y$ แทน y ส่วนข้อมูล x ไม่ต้องแปลง เมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์

$B = -a_1$ และ $A = e^{a_0}$

16

Least-squares regression

i	1	2	3	4	SUM
x_i	0	2	4	6	12
y_i	5.0	3.7	2.7	2.0	
$\log(y_i)$	1.6094	1.3083	0.9933	0.6931	4.6041
$(x_i)^2$	0	4	16	36	56
$x_i \log(y_i)$	0	2.6166	3.9732	4.1586	10.7484

$$a_1 = \frac{4(10.7484) - (12)(4.6041)}{4(56) - (12)^2} = -0.1532,$$

$\bar{x} = \frac{12}{4} = 3$, $\overline{\log(y)} = \frac{4.6041}{4} = 1.151$ และ $a_0 = \overline{\log(y)} - a_1\bar{x} = 1.610625$

สมการเชิงเส้นที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression คือ

$\log(y)_{approx} = a_1x + a_0 = (-0.1532)x + 1.610625$

Tanit MALAKORN, PhD
Department of Electrical and Computer Engineering
Naresuan University, THAILAND

Least-squares regression

เมื่อคำนวณหา a_0 และ a_1 ได้แล้ว ให้แปลงกลับไปเป็น A และ B ด้วยความสัมพันธ์

$$B = -a_1 \text{ และ } A = e^{a_0}$$

จึงได้ว่า B = 0.1532 และ $A = e^{1.610625} = 5.0059$

ดังนั้นสมการที่ประมาณชุดข้อมูลดังกล่าวโดยใช้วิธี Least-squares regression คือ

$y_{approx} = Ae^{-Bx} = 5.0059e^{-0.1532x}$

Tanit MALAKORN, PhD
Department of Electrical and Computer Engineering
Naresuan University, THAILAND