Regresi

Umi Sa'adah
Politeknik Elektronika Negeri Surabaya
2008

Topik

- Regresi Linier
- Regresi Non Linier

Regresi & Pencocokan Kurva

Pada pembahasan interpolasi, data *dianggap telah diperoleh dengan teliti tanpa kesalahan yang signifikan. Kurva* polinom penginterpolasi *melintasi setiap titik/simpul data secara langsung.*

Pada data dengan tingkat kesalahan yang signifikan, kurva dikehendaki menyatakan kecenderungan umum dari data (mengikuti pola titik-titik data yang diperoleh). Pendekatan ini dinamakan **regresi kuadrat terkecil.**

1 Regresi Linear

Penerapan paling sederhana prinsip kuadrat terkecil (*least square*) adalah mencocokkan kurva linear pada sekumpulan n data pengamatan/percobaan yang terdiri atas: (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), dimana diasumsikan setiap x_i terbebas dari kesalahan.

Regresi vs Interpolasi

REGRESI KUADRAT TERKECIL	INTERPOLASI POLINOM
1. Data berasal dari hasil pengukuran	1. Data berasal dari fungsi yang ingin disederhanakan dengan polinom, dari tabel di literatur, atau dari hasil pengukuran
2. Data berketelitian rendah (mengandung error)	2. Data berketelitian tinggi
3. Fungsi kuadrat terkecil tidak perlu melalui setiap titik data. Kurva fungsinya dirancang mengikuti pola titik-titik sebagai suatu kelompok	3. Fungsi polinom interpolasi harus melalui semua titik data. Semakin banyak datanya, semakin tinggi derajat polinom, dan semakin besar error pembulatannya

Manfaat pencocokan kurva untuk data hasil pengukuran :

1. Bagi ahli sains/rekayasa: mengembangkan formula empirik untuk sistem yang diteliti
2. Bagi ahli ekonomi: menentukan kurva kecenderungan ekonomi untuk “meramalkan” kecenderungan masa depan

REGRESI LINIER

- Misalkan (xi, yi) adalah data hasil pengukuran, kita akan menghampiri titik tsb dengan sebuah garis lurus
- Garis lurus tsb dibuat sedemikian sehingga errornya sekecil mungkin dengan titik-titik data

Regresi Linier

- Karena data mengandung error, maka nilai data sebenarnya, g(xi) dapat ditulis sbb :

  g(xi) = yi + ei   i = 1, 2, 3, .., n

  Yang dalam hal ini ei adalah error setiap data
- Diinginkan untuk membentuk fungsi linier :

  f(xi) = a + bx
- Yang mencocokkan data sedemikian sehingga deviasinya

  $$r_i = y_i - f(x_i) = y_i - (a + bx_i)$$
- Total kuadrat deviasinya adalah

$$R = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Regresi Linier

- Agar R minimum, maka haruslah

$$\frac{\partial R}{\partial a} = -2\sum (y_i - a - bx_i) = 0$$
$$\frac{\partial R}{\partial b} = -2\sum x_i(y_i - a - bx_i) = 0$$

- Masing-masing ruas kedua persamaan dibagi dengan -2 :

$$\sum (y_i - a - bx_i) = 0 \Rightarrow \sum y_i - \sum a - \sum bx_i = 0$$
$$\sum x_i(y_i - a - bx_i) = 0 \Rightarrow \sum x_i y_i - \sum ax_i - \sum bx_i^2 = 0$$

- Selanjutnya

$$\sum a + \sum bx_i = \sum y_i$$
$$\sum ax_i + \sum bx_i^2 = \sum x_i y_i$$

atau

$$na + b\sum x_i = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i$$

Dalam bentuk matriks didapat bentuk sbb:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Penghitungan Error

- Untuk menentukan seberapa bagus fungsi hampiran mencocokkan data dapat diukur dengan error RMS (*Root-Mean-Square* error)

$$E_{RMS} = \left(\frac{1}{n} \sum_{i=1}^{n} | f(x_i) - y_i |^2 \right)^{1/2}$$

- Semakin kecil nilai E_{RMS} semakin bagus fungsi hampiran mencocokkan titik-titik data

Contoh Soal :

- Tentukan persamaan garis lurus yang mencocokkan data pada tabel dibawah ini.
- Kemudian perkirakan nilai y untuk x = 1.0
- Penyelesaian :

i	**xi**	**yi**	**(xi)2**	**xiyi**
1	0.1	0.61	0.01	0.061
2	0.4	0.92	0.16	0.368
3	0.5	0.99	0.25	0.495
4	0.7	1.52	0.49	1.064
5	0.7	1.47	0.49	1.029
6	0.9	2.03	0.81	1.827
	3.3	**7.54**	**2.21**	**4.844**

Contoh Soal :

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

- Diperoleh Sistem Persamaan Linier

$$\begin{bmatrix} 6 & 3.3 \\ 3.3 & 2.21 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 7.54 \\ 4.844 \end{bmatrix}$$

- a = 0.28616 b = 1.76456
- Pers garis regresi f(x) = 0.28616 + 1.76256x

Contoh Soal

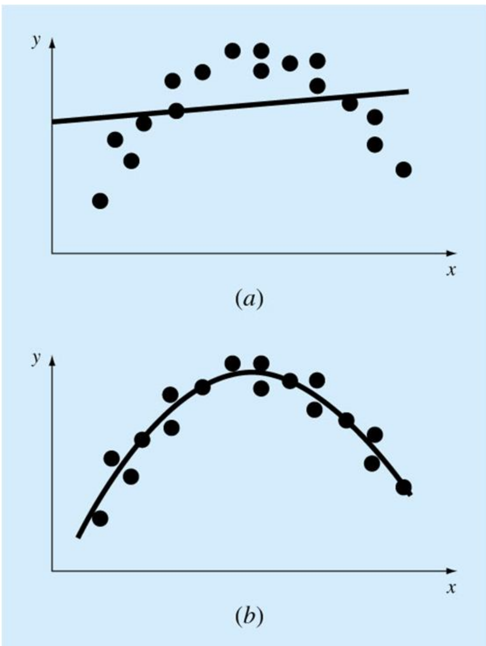
- Perbandingan antara nilai yi dan f(xi)

i	xi	yi	f(xi)	Deviasi |f(xi)-yi|	(deviasi)2
1	0.1	0.61	0.46265	0.14735	0.021712
2	0.4	0.92	0.992	0.072	0.005184
3	0.5	0.99	1.16845	0.17845	0.031844
4	0.7	1.52	1.52135	0.00135	1.82E-06
5	0.7	1.47	1.52135	0.05135	0.002637
6	0.9	2.03	1.87425	0.15575	0.024258
					0.085637

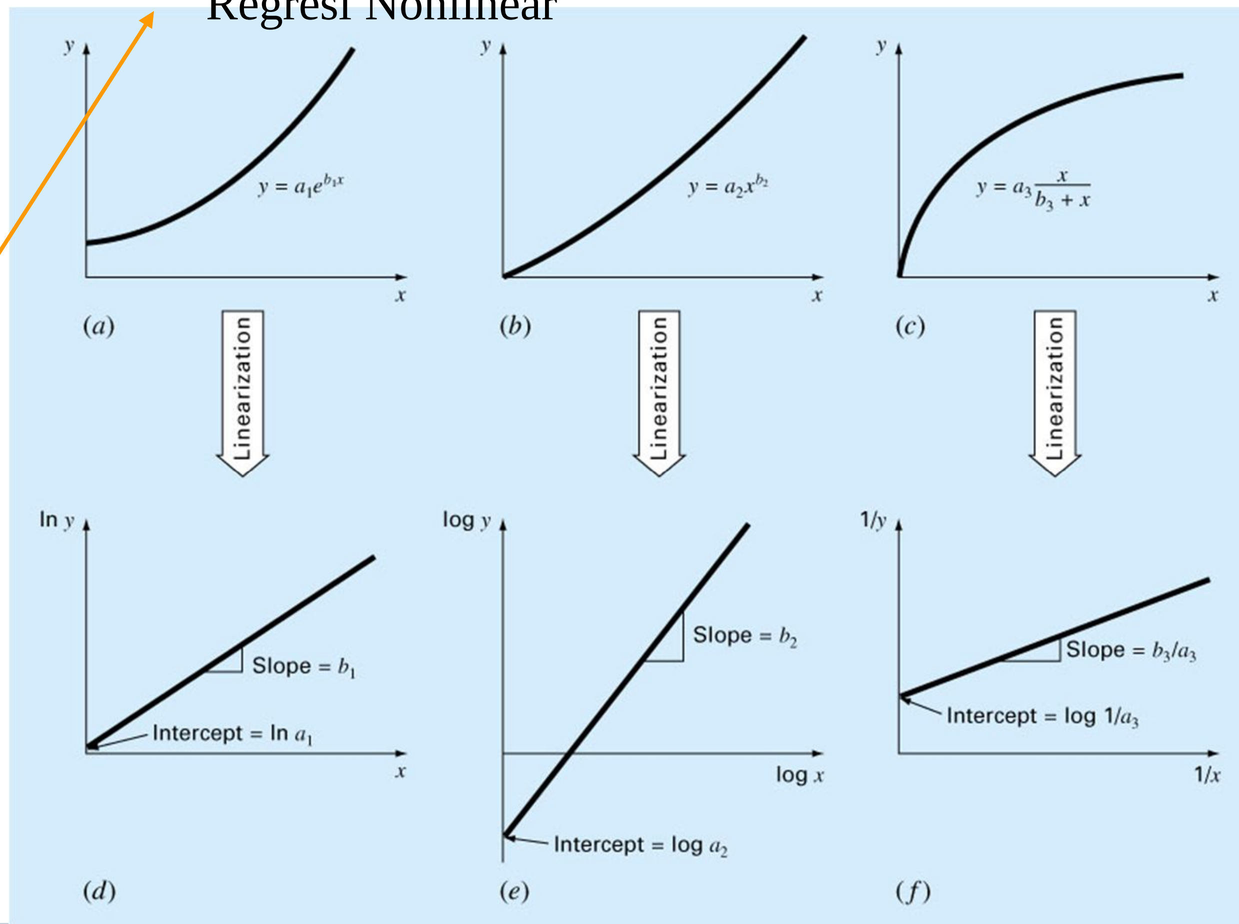
- Taksiran nilai y untuk x = 1.0 adalah 2.0507
- $E_{RMS} = \frac{(0.085637/6)}{6}^{1/2} = 0.119464$

Linearisasi Persamaan Nonlinear

Regresi Nonlinear



Data yang tidak cocok dengan bentuk linear



Transformasi Linear (jika mungkin)

Linearisasi Persamaan Nonlinear

- Meskipun fungsi hampiran berbentuk non linier, namun pencocokan kurva dengan fungsi non linier tersebut dapat diselesaikan juga diselesaikan dengan cara regresi linier
- Misalnya 2 macam fungsi non linear berikut ini :

  1. Persamaan pangkat sederhana

     $y = Cx^b$, C dan b konstanta

  2. Model Eksponensial

     $y = Ce^{bx}$, C dan b konstanta

     Contoh : - model pertumbuhan populasi

     - model peluruhan zat radioaktif

Pelinearan Pers Pangkat Sederhana

- Misalkan kita akan mencocokkan data dengan fungsi :

$$y = Cx^b \Leftrightarrow \ln(y) = \ln(C) + b\ln(x)$$

- Definisikan : $Y = \ln(y)$

  $a = \ln(C)$

  $X = \ln(x)$

  Sehingga persamaan regresi liniearnya adalah :

$$Y = a + bX$$

- Lakukan pengubahan dari (x_i,y_i) menjadi $(\ln(x_i),\ln(y_i))$ lalu hitung a dan b dengan cara regresi linier.
- Dari pers a = ln(C) maka kita dapat menghitung nilai $C = e^a$.
- Masukkan nilai b dan C ke dalam pangkat $y=Cx^b$

Contoh Soal :

- Cocokkan data berikut dengan $f(x) = Cx^b$

i	xi	yi	Xi=ln(xi)	Yi=ln(yi)	Xi2	XiYi
1	0.15	4.4964	-1.89712	1.503277	3.599064	-2.8519
2	0.4	5.1284	-0.91629	1.634794	0.839589	-1.49795
3	0.6	5.6931	-0.51083	1.739255	0.260943	-0.88846
4	1.01	6.2884	0.00995	1.838707	9.9E-05	0.018296
5	1.5	7.0989	0.405465	1.95994	0.164402	0.794687
6	2.2	7.5507	0.788457	2.02164	0.621665	1.593977
7	2.4	7.5106	0.875469	2.016315	0.766446	1.765221
			-1.24489	**12.71393**	**6.252207**	**-1.06612**

- Diperoleh sistem persamaan linier

$$\begin{bmatrix} 7 & -1.2447 \\ -1.2447 & 6.2522 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 12.7139 \\ -1.0659 \end{bmatrix}$$

Contoh Soal :

- $a = 1.8515 \qquad b = 0.1981$
- Hitung $C = e^a = e^{1.8515} = 6.369366$
- Jadi $f(x) = 6.369366x^{0.1981}$

Pelinieran Model Eksponensial $y = Ce^{bx}$

- Misalkan kita akan mencocokkan data dg fungsi : $y = Ce^{bx}$
- Lakukan pelinieran sbb :

  $y = Ce^{bx}$

  $\ln(y) = \ln(C) + bx \ln(e)$

  $\ln(y) = \ln(C) + bx$ → $\ln(e) = 1$
- Definisikan :
  - $Y = \ln(y)$
  - $a = \ln(C)$ → $C = e^a$
  - $X = x$
- Persamaan Regresi Liniernya : $Y = a + bX$
- Lakukan pengubahan (x_i, y_i) → $(x_i, \ln(y_i))$ lalu hitung a dan b dengan cara regresi linier
- Dari persamaan $a = \ln(C)$ di dapat $C = e^a$
- Masukkan nilai b dan C dalam persamaan eksponensial $y = Ce^{bx}$