http://istiarto.staff.ugm.ac.id

REGRESI DAN INTERPOLASI

Curve Fitting

Curve Fitting

- Acuan
  - Chapra, S.C., Canale R.P., 1990, *Numerical Methods for Engineers*, 2nd Ed., McGraw-Hill Book Co., New York.
    - Chapter 11 dan 12, pp. 319-398.

Curve Fitting

- Mencari garis/kurva yang mewakili serangkaian titik data
- Ada dua cara untuk melakukannya, yaitu
  - Regresi
  - Interpolasi
- Aplikasi di bidang enjiniring
  - Pola perilaku data (*trend analysis*)
  - Uji hipotesis (*hypothesis testing*)

Curve Fitting

- Pemakaian regresi
  - Apabila data menunjukkan tingkat kesalahan yang cukup signifikan atau menunjukkan adanya noise
  - Untuk mencari satu kurva tunggal yang mewakili pola umum perilaku data
  - Kurva yang dicari tidak perlu melewati setiap titik data

Curve Fitting

- Interpolasi
  - Diketahui bahwa data sangat akurat
  - Untuk mencari satu atau serangkaian kurva yang melewati setiap titik data
  - Untuk memperkirakan nilai-nilai di antara titik-titik data
- Extrapolasi
  - Mirip dengan interpolasi, tetapi untuk memperkirakan nilai-nilai di luar range titik-titik data

Curve Fitting terhadap Data Pengukuran

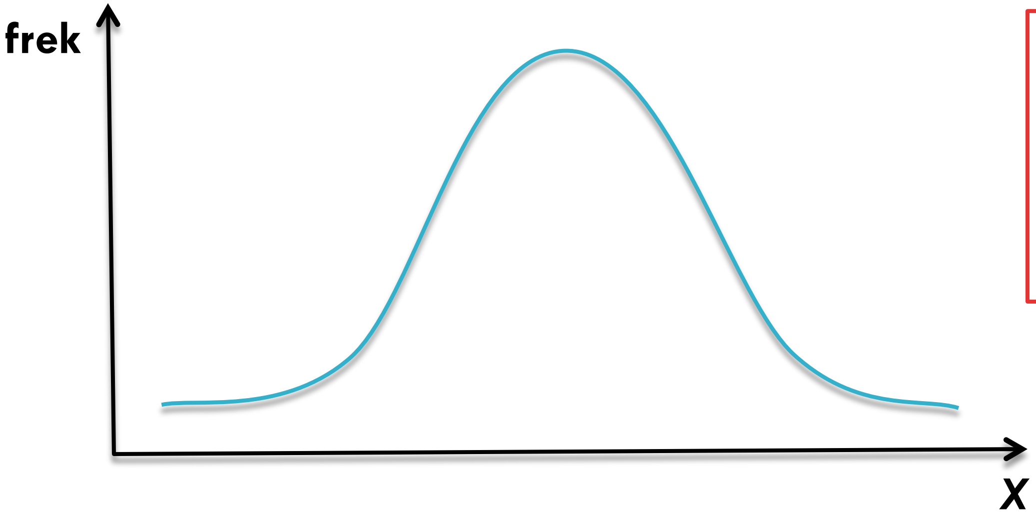
- Analisis pola perilaku data
  - Pemanfaatan pola data (pengukuran, experimen) untuk melakukan perkiraan
  - Apabila data persis (akurat): interpolasi
  - Apabila data tak persis (tak akurat): regresi
- Uji hipotesis
  - Pembandingan antara hasil teori atau hasil hitungan dengan hasil pengukuran

Beberapa Parameter Statistik

merepresentasikan sebaran data

- Rata-rata aritmatik, *mean* $\Rightarrow$ $\bar{y} = \frac{1}{n}\sum y_i$
- Simpangan baku, *standard deviation*, deviasi standar $\Rightarrow$ $s_y = \sqrt{\frac{S_t}{n-1}}$ $\quad S_t = \sum \left(y_i - \bar{y}\right)^2$
- Varian ('ragam'), *variance* $\Rightarrow$ $s_y{}^2 = \frac{S_t}{n-1}$
- *Coefficient of variation* $\Rightarrow$ $\text{c.v.} = \frac{s_y}{\bar{y}} 100\%$

Distribusi Probabilitas



Distribusi Normal

salah satu distribusi/sebaran data yang sering dijumpai adalah distribusi normal

Regresi

Regresi linear

Regresi non-linear

Regresi: Metode Kuadrat Terkecil

- Mencari suatu kurva atau suatu fungsi (pendekatan) yang sesuai dengan pola umum yang ditunjukkan oleh data
  - Datanya menunjukkan kesalahan yang cukup signifikan
  - Kurva tidak perlu memotong setiap titik data
- Regresi linear
- Regresi persamaan-persamaan tak-linear yang dilinearkan
- Regresi tak-linear

Regresi: Metode Kuadrat Terkecil

- Bagaimana caranya?
  - Program komputer
  - Spreadsheet (Microsoft Excel)
  - SciLab

Regresi Linear

- Mencari suatu kurva lurus yang cocok menggambarkan pola serangkaian titik data: (x_1,y_1), (x_2,y_2) … (x_n,y_n)

$$y = a_0 + a_1 x + e$$

a_0 : intercept

a_1 : slope

e : error

- Microsoft Excel
  - INTERCEPT(y_1:y_n;x_1:x_n)
  - SLOPE(y_1:y_n;x_1:x_n)

Regresi Linear

- Kesalahan atau residu (e) adalah perbedaan antara nilai y sesungguhnya dan y nilai pendekatan menurut persamaan linear $a_0 + a_1 x$.

$$e = y - a_0 - a_1 x$$

- Minimumkan jumlah kuadrat residu tersebut

$$\min\left[S_r\right] = \min\left[\sum e_i^2\right] = \min\left[\sum \left(y_i - a_0 - a_1 x_i\right)^2\right]$$

Regresi Linear

- Bagaimana cara mencari koefisien a_0 dan a_1?
  - Diferensialkan persamaan tersebut dua kali, masing-masing terhadap a_0 dan a_1.
  - Samakan kedua persamaan hasil diferensiasi tersebut dengan nol.

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1 x_i\right) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum\left(y_i - a_0 - a_1 x_i\right)x_i = 0$$

Regresi Linear

- Selesaikan persamaan yang didapat untuk mencari a_0 dan a_1

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^{\,2} - \left(\sum x_i\right)^2}$$

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

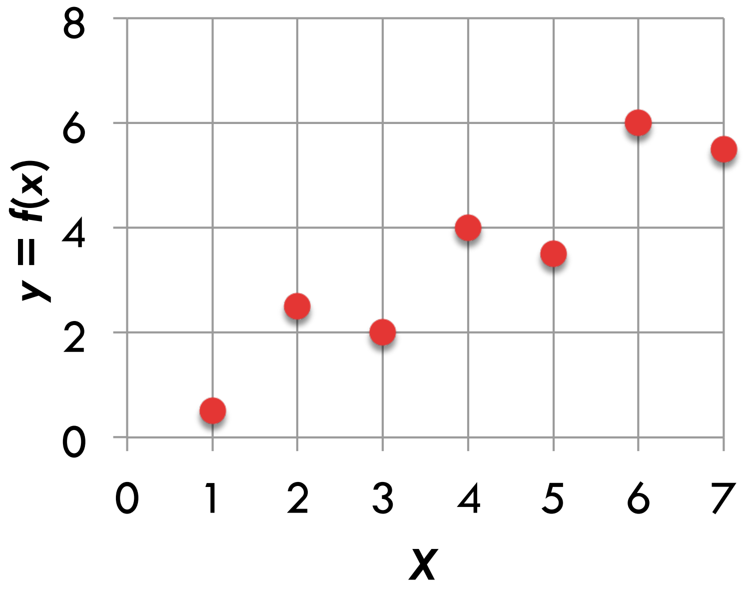
- dalam hal ini, $\bar{y}$ dan $\bar{x}$ masing-masing adalah nilai *y* rata-rata *x* rata-rata

Contoh Regresi Linear

Tabel data

i	x_i	$y_i = f(x_i)$
0	1	0.5
1	2	2.5
2	3	2
3	4	4
4	5	3.5
5	6	6
6	7	5.5

Grafik/kurva data



Hitungan regresi linear

i	x_i	y_i	$x_i\ y_i$	x_i^2	y_{reg}	$(y_i - y_{reg})^2$	$(y_i - y_{mean})^2$
0	1	0.5	0.5	1	0.910714	0.168686	8.576531
1	2	2.5	5	4	1.75	0.5625	0.862245
2	3	2.0	6	9	2.589286	0.347258	2.040816
3	4	4.0	16	16	3.428571	0.326531	0.326531
4	5	3.5	17.5	25	4.267857	0.589605	0.005102
5	6	6.0	36	36	5.107143	0.797194	6.612245
6	7	5.5	38.5	49	5.946429	0.199298	4.290816
$\sum =$	**28**	**24.0**	**119.5**	**140**	$\sum =$	**2.991071**	**22.71429**

Hitungan regresi linear

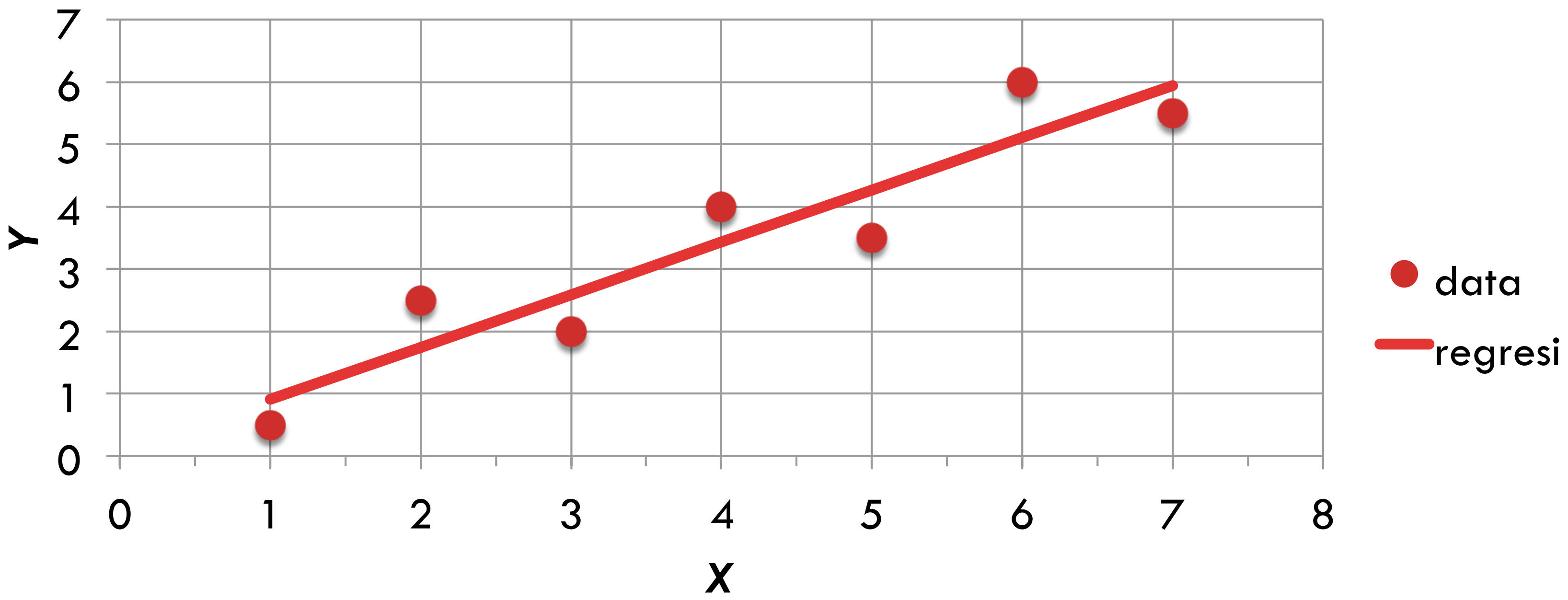
$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^{\,2} - \left(\sum x_i\right)^2} = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.839286$$

$$\bar{y} = \frac{24}{7} = 3.4$$

$$\bar{x} = \frac{28}{7} = 4$$

$$a_0 = 3.4 - 0.839286(4) = 0.071429$$

Hitungan regresi linear



Regresi Linear

- Kuantifikasi kesalahan
  - Kesalahan standar

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} \qquad S_r = \sum\left(y_i - a_0 - a_1 x_i\right)^2$$

  - Perhatikan kemiripannya dengan simpangan baku (*standard deviation*)

$$s_y = \sqrt{\frac{S_t}{n-1}} \qquad S_t = \sum\left(y_i - \bar{y}\right)^2$$

Regresi Linear

- Beda antara kedua kesalahan tersebut menunjukkan perbaikan atau pengurangan kesalahan

$$r^2 = \frac{S_t - S_r}{S_t}$$

→ koefisien determinasi (*coefficient of determination*)

$$r = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}$$

→ koefisien korelasi (*correlation coefficient*)

Hitungan regresi linear

$$S_r = \sum \left(y_i - a_0 - a_1 x_i\right)^2 = 2.991071$$

$$S_t = \sum \left(y_i - \bar{y}\right)^2 = 22.71429$$

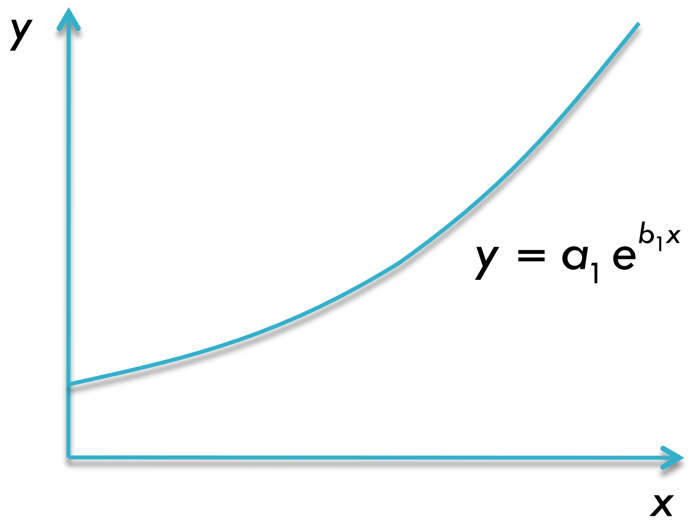
$$r^2 = \frac{S_t - S_r}{S_t} = \frac{22.71429 - 2.991071}{22.71429} = 0.868318$$

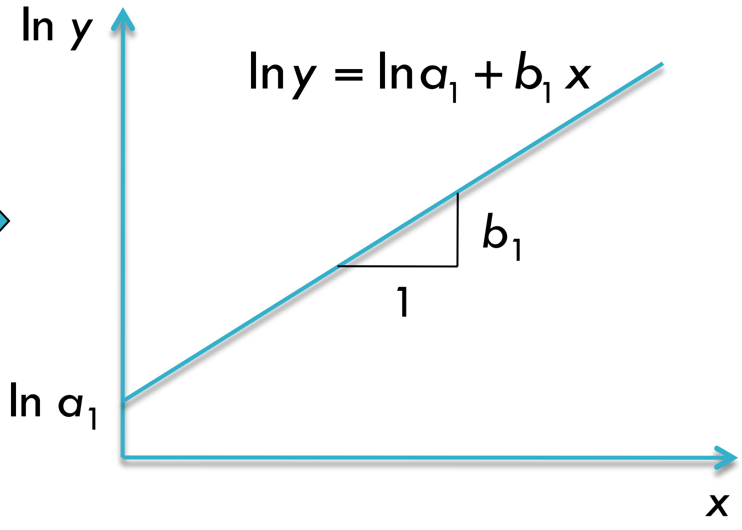
$$r = 0.931836$$

Regresi Linear

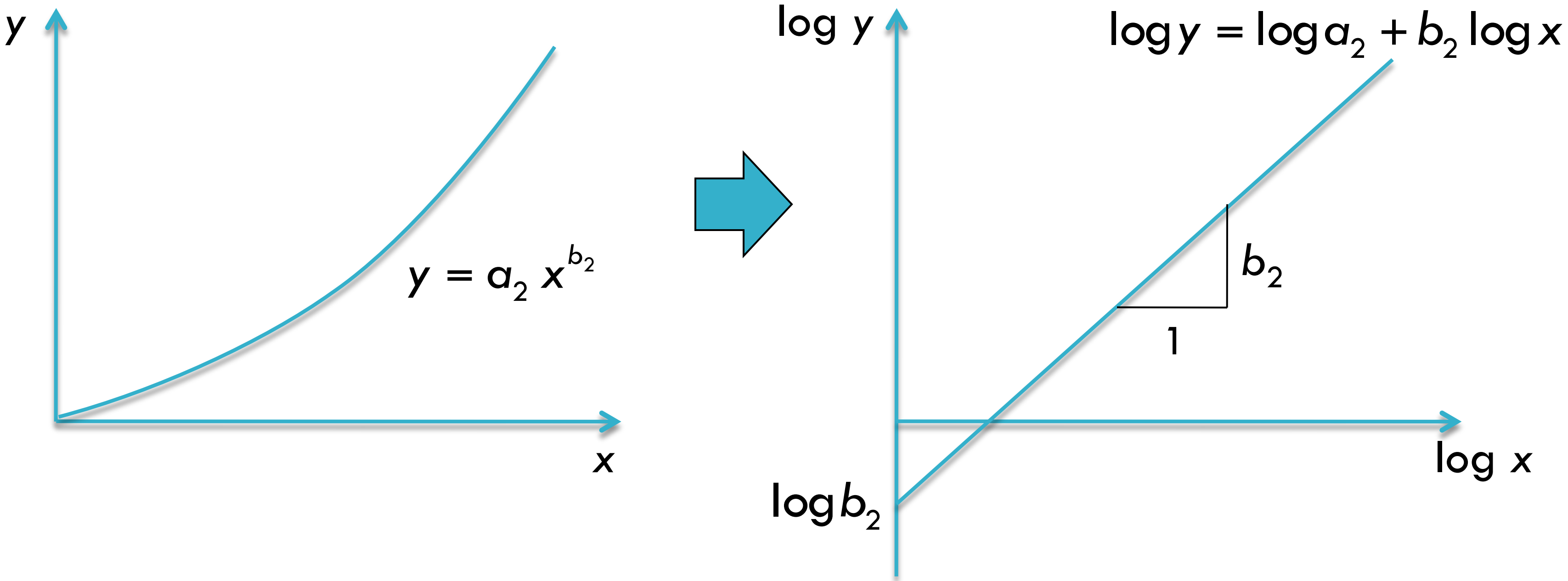
- Linearisasi persamaan-persamaan tak-linear
  - Logaritmik menjadi linear
  - Eksponensial menjadi linear
  - Pangkat (polinomial tingkat n > 1) menjadi linear (polinomial tingkat 1)
  - Dll.

Linearisasi persamaan non-linear (1)



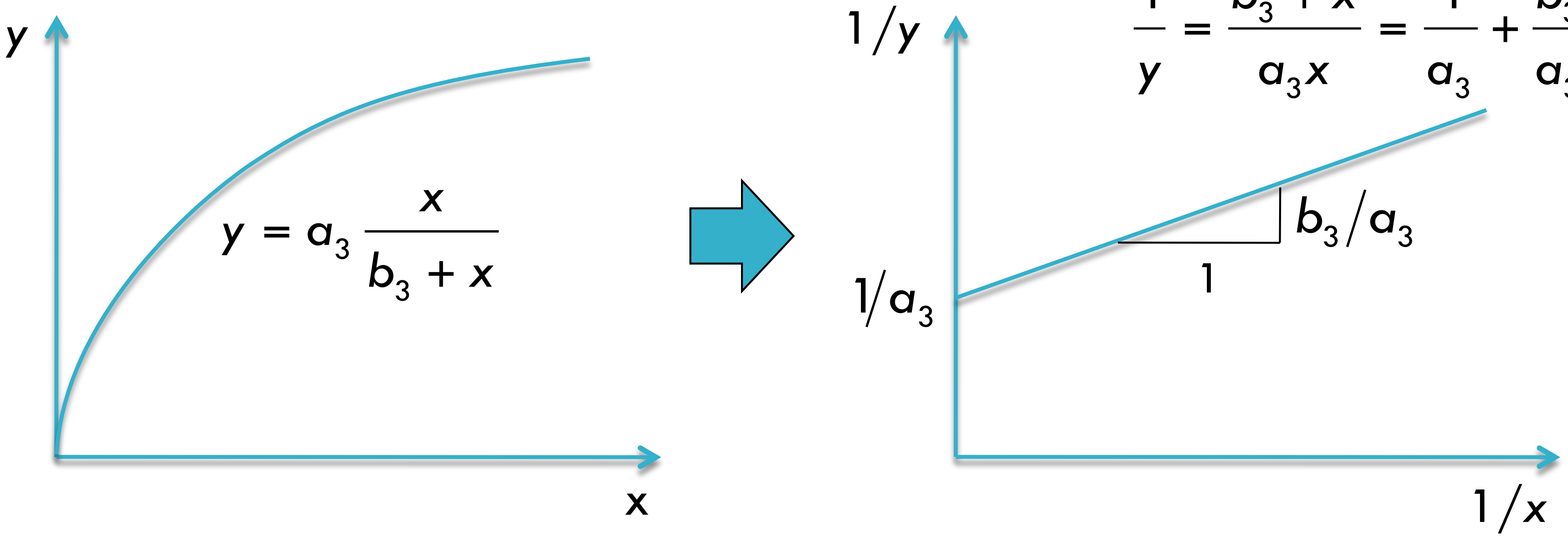


Linearisasi persamaan non-linear (2)



Linearisasi persamaan non-linear (3)

$$\frac{1}{y} = \frac{b_3 + x}{a_3 x} = \frac{1}{a_3} + \frac{b_3}{a_3}\frac{1}{x}$$

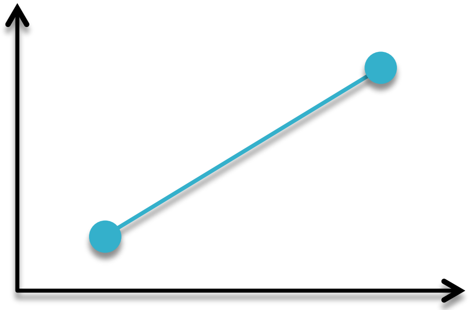


Interpolasi

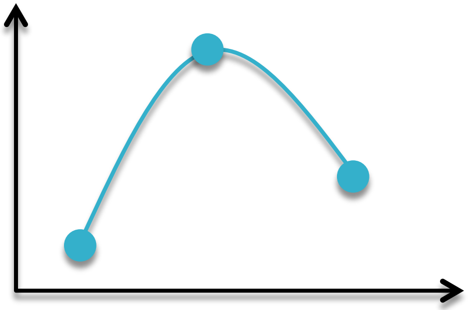
Metode Newton

Metode Lagrange

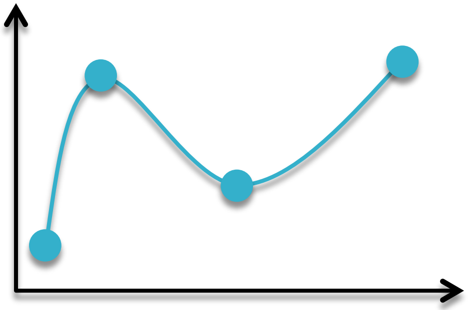
Interpolasi



linear



kuadratik

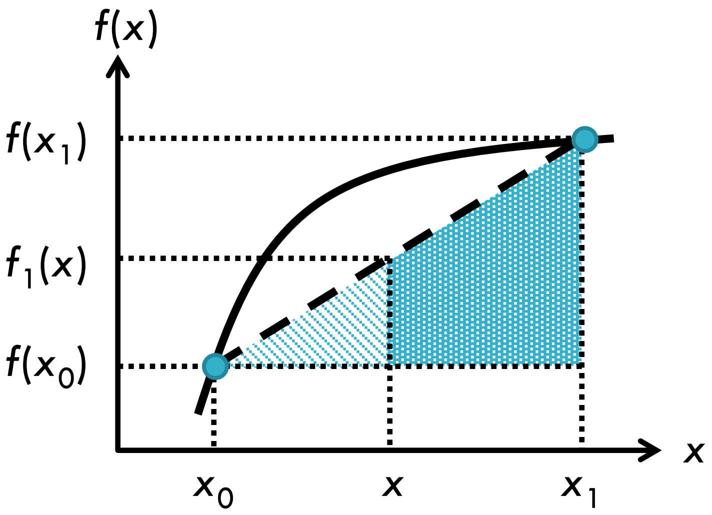


kubik

Interpolasi

- Penyelesaian persamaan polinomial tingkat n membutuhkan sejumlah $n + 1$ titik data
- Metode untuk mencari polinomial tingkat n yang merupakan interpolasi sejumlah $n + 1$ titik data:
  - Metode Newton
  - Metode Lagrange

Interpolasi Linear: Metode Newton



$$\frac{f_1(x)-f(x_0)}{x-x_0}=\frac{f(x_1)-f(x_0)}{x_1-x_0}$$

$$f_1(x)=f(x_0)+\frac{f(x_1)-f(x_0)}{x_1-x_0}(x-x_0)$$

Interpolasi Kuadratik: Metode Newton

$$
\begin{aligned}
f_2(x) &= b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) \\
&= b_0 + b_1 x - b_1 x_0 + b_2 x^2 + b_2 x_0 x_1 - b_2 x x_0 - b_2 x x_1 \\
&= \underbrace{(b_0 - b_1 x_0 + b_2 x_0 x_1)}_{a_0} + \underbrace{(b_1 - b_2 x_0 - b_2 x_1)}_{a_1} x + \underbrace{(b_2)}_{a_2} x^2
\end{aligned}
$$

$$
f_2(x) = a_0 + a_1 x + a_2 x^2 \quad
\begin{cases}
a_0 = b_0 - b_1 x_0 + b_2 x_0 x_1 \\
a_1 = b_1 - b_2 x_0 - b_2 x_1 \\
a_2 = b_2
\end{cases}
$$

Interpolasi Kuadratik: Metode Newton

$$b_0 = f\left(x_0\right)$$

$$b_1 = \frac{f\left(x_1\right) - f\left(x_0\right)}{x_1 - x_0} = f\left[x_1, x_0\right]$$

$$b_2 = \frac{\dfrac{f\left(x_2\right) - f\left(x_1\right)}{x_2 - x_1} - \dfrac{f\left(x_1\right) - f\left(x_0\right)}{x_1 - x_0}}{x_2 - x_1} = f\left[x_2, x_1, x_0\right] = \frac{f\left[x_2, x_1\right] - f\left[x_1, x_0\right]}{x_2 - x_1}$$

Interpolasi Polinomial: Metode Newton

$$f_n(x) = b_0 + b_1(x - x_0) + \ldots + b_n(x - x_0)(x - x_1)\ldots(x - x_{n-1})$$

$$\begin{aligned}
b_0 &= f(x_0) \\
b_1 &= f[x_1, x_0] \\
b_2 &= f[x_2, x_1, x_0] \\
&\;\;\vdots \\
b_n &= f[x_n, x_{n-1}, \ldots, x_1, x_0]
\end{aligned}$$

Interpolasi Polinomial: Metode Newton

$$f\left[x_i, x_j\right] = \frac{f\left(x_i\right) - f\left(x_j\right)}{x_i - x_j}$$

$$f\left[x_i, x_j, x_k\right] = \frac{f\left[x_i, x_j\right] - f\left[x_j, x_k\right]}{x_i - x_k}$$

$$f\left[x_n, x_{n-1}, \ldots, x_1, x_0\right] = \frac{f\left[x_n, x_{n-1}, \ldots, x_1\right] - f\left[x_{n-1}, n_{n-2}, \ldots, x_0\right]}{x_n - x_0}$$

Interpolasi Polinomial: Metode Newton

$$f_n(x) = f(x_0) + (x - x_0)f[x_1, x_0] + (x - x_0)(x - x_1)f[x_2, x_1, x_0] + \ldots + (x - x_0)(x - x_1)\ldots(x - x_{n-1})f[x_n, x_{n-1}, \ldots, x_0]$$

Interpolasi Polinomial: Metode Newton

i	x_i	$f(x_i)$	langkah hitungan		
			ke-1	ke-2	ke-3
0	x_0	$f(x_0)$	$f[x_1,x_0]$	$f[x_2,x_1,x_0]$	$f[x_3,x_2,x_1,x_0]$
1	x_1	$f(x_1)$	$f[x_2,x_1]$	$f[x_3,x_2,x_1]$	
2	x_2	$f(x_2)$	$f[x_3,x_2]$		
3	x_3	$f(x_3)$			

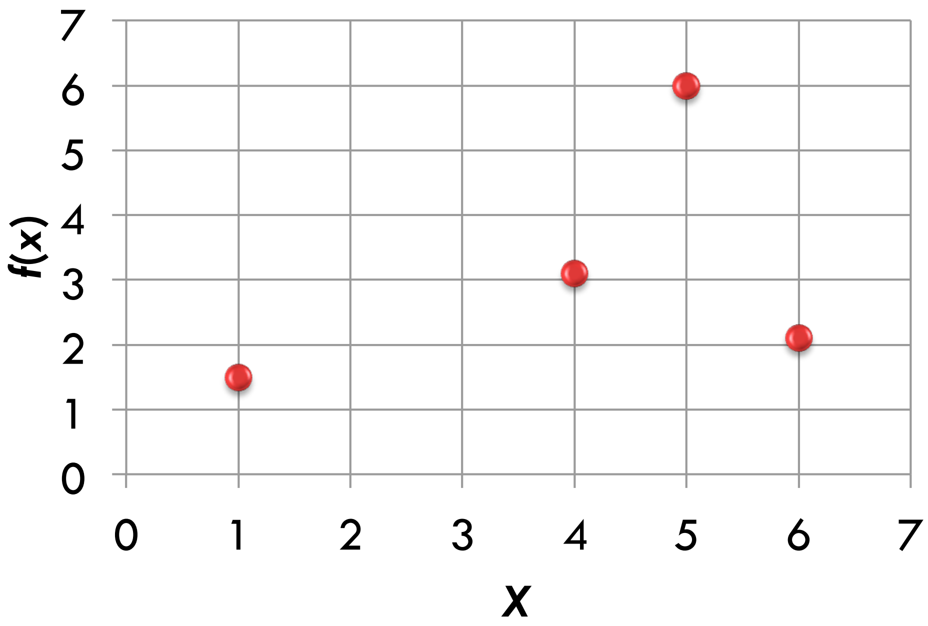
Interpolasi Polinomial: Metode Lagrange

$$f_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i)$$

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}$$

Contoh interpolasi

i	x_i	$f(x_i)$
0	1	1.5
1	4	3.1
2	5	6
3	6	2.1



Spline

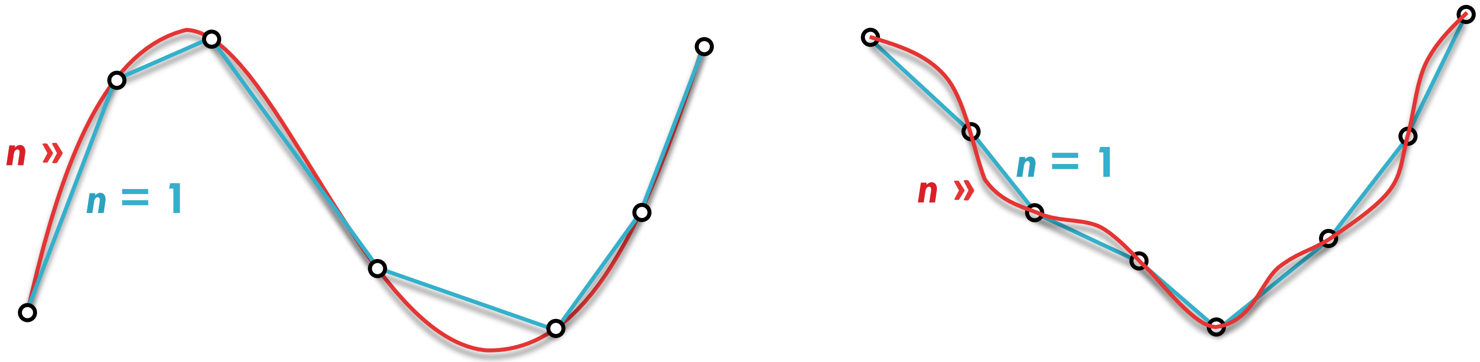
- Linear
- Kuadratik
- Kubik

Interpolasi: Spline

- Jumlah titik data $n + 1$ → interpolasi polinomial tingkat n
  - Tingkat besar, n >>, mengalami kesulitan apabila titik-titik data menunjukkan adanya perubahan tiba-tiba di suatu titik tertentu (perubahan gradien secara tiba-tiba).
  - Dalam situasi tsb, polinomial bertingkat kecil, n «, dapat lebih representatif untuk mewakili pola data.
  - *Spline*
    - *Cubic splines* ($n = 3$)
    - *Quadratic splines*
    - *Linear splines*

Interpolasi Polinomial vs *Spline*

- polinomial tingkat *n*



Linear Splines

- *First-order spline* : garis lurus
- Data urut : $x_0, x_1, x_2, \ldots, x_n$

$$f(x) = f(x_0) + m_0(x - x_0) \qquad x_0 \le x \le x_1$$

$$f(x) = f(x_1) + m_1(x - x_1) \qquad x_1 \le x \le x_2$$

.

.

.

$$f(x) = f(x_{n-1}) + m_{n-1}(x - x_{n-1}) \qquad x_{n-1} \le x \le x_n$$

Linear Splines

- *Slope*/gradien/kemiringan garis lurus, m_i:

$$m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$

- *Linear spline*, dengan demikian, adalah sama dengan interpolasi linear.
- Kekurangan *linear spline* adalah ketidak-mulusan kurva interpolasi.
- Terdapat perubahan *slope* yang sangat tajam di titik-titik data atau di titik-titik pertemuan kurva *spline* (*knot*).

Linear Splines

- Dengan kata lain, terdapat diskontinuiti/ketidak-mulusan diferensial pertama (kemiringan) fungsi *spline* di titik-titik *knot*.
- Perlu dicari suatu fungsi *spline* (bertingkat *n* lebih tinggi) dengan menyamakan diferensial pertamanya (kemiringannya) di titik-titik *knot*.

Quadratic Splines

- Untuk mendapatkan kurva yang memiliki diferensial/laju-perubahan ke-m kontinu di titik knot, maka diperlukan kurva *spline* yang bertingkat paling kecil $m + 1$.
- Yang paling banyak dipakai adalah *spline* tingkat 3 (*cubic spline*): diferensial pertama dan kedua kontinu di titik-titik knot.
- Ketidak-mulusan diferensial ketiga, keempat, dst. umumnya tidak begitu tampak secara visual.
- Sebelum membahas *cubic spline*, berikut dipaparkan *quadratic spline* terlebih dulu.

Quadratic Splines

- Tujuan: mencari polinomial tingkat 2 untuk setiap interval titik-titik data.
- Polinomial tingkat 2 tsb harus memiliki diferensial pertama (laju perubahan) yang kontinu di titik-titik data.
- Polinomial tingkat 2:

$$f(x) = a_i x^2 + b_i x + c_i$$

- Untuk $(n+1)$ titik data ($i = 0, 1, 2, \ldots, n$), terdapat n interval, sehingga terdapat $3n$ koefisien yang harus dicari (a_i, b_i, c_i), $i = 1, 2, \ldots, n$.
- Perlu persamaan sejumlah $3n$.

Quadratic Splines

- Ke-3*n* persamaan tsb adalah sbb.
  - Kurva *spline* memotong titik-titik data (*knot*): interval $i - 1$ dan i bertemu di titik data $\{x_{i-1}, f(x_{i-1})\}$.

$$a_{i-1}x_{i-1}^2 + b_{i-1}x_{i-1} + c_{i-1} = f(x_{i-1})$$

$$a_i x_{i-1}^2 + b_i x_{i-1} + c_i = f(x_{i-1})$$

$i = 2, 3, \ldots, n$

$2(n - 1)$ pers.

Quadratic Splines

- Kurva *spline* di selang (interval) pertama memotong titik data pertama ($i = 1$) dan kurva *spline* di interval terakhir memotong titik data terakhir ($i = n$).

$$a_1 x_0^2 + b_1 x_0 + c_1 = f(x_0)$$
$$a_n x_n^2 + b_n x_n + c_n = f(x_n)$$

2 pers.

Quadratic Splines

- Diferensial (gradien) kurva *spline* di dua interval berurutan adalah sama di titik data yang bersangkutan.
  Gradien:

$$f'(x) = 2ax + b$$

$$\Downarrow$$

$$2a_{i-1}x_{i-1} + b_{i-1} = 2a_i x_{i-1} + b_i$$

$i = 2, 3, \ldots, n$
$(n - 1)$ pers.

Quadratic Splines

- Diferensial kedua (laju perubahan gradien) kurva *spline* di titik data pertama sama dengan nol.

$a_i = 0$ 1 pers.

- Konsekuensi: 2 titik data pertama (i = 0 dan i = 1) dihubungkan dengan garis lurus.

Quadratic Splines

- Dengan demikian, jumlah persamaan seluruhnya adalah:

$$2(n - 1) + 2 + (n - 1) + 1 = 3n$$

Cubic Splines

- Tujuan: mencari polinomial tingkat 3 untuk setiap interval titik-titik data.
- Polinomial tingkat 3 tsb harus memiliki diferensial pertama (gradien) dan diferensial kedua (laju perubahan gradien) yang kontinu di titik-titik data.
- Polinomial tingkat 3:

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i$$

- Untuk (n+1) titik data (i = 0, 1, 2, …, n), terdapat n interval, shg. terdapat 4n koefisien yang harus dicari (a_i,b_i,c_i,d_i), i = 1, 2, ..., n.
- Perlu persamaan sejumlah 4n.

Cubic Splines

- Ke-4n persamaan tsb adalah sbb.
  - Kurva *spline* memotong titik-titik data (*knot*): interval $i - 1$ dan i bertemu di titik data $\{x_{i-1}, f(x_{i-1})\}$ → $(2n - 2)$ pers.
  - Kurva *spline* di interval pertama memotong titik data pertama dan kurva *spline* terakhir memotong titik data terakhir → 2 pers.
  - Diferensial pertama kurva *spline* di dua interval berurutan adalah sama di titik data ybs. → $(n - 1)$ pers.

Cubic *Splines*

- Diferensial kedua kurva spline di dua interval berurutan adalah sama di titik data ybs. → $(n-1)$ pers.
- Diferensial kedua kurva *spline* di titik data pertama dan terakhir sama dengan nol → 2 pers.

Konsekuensi: kurva *spline* berupa garis lurus di titik data pertama dan di titik data terakhir.

Alternatif: diferensial kedua kurva *spline* di 2 titik data tsb diketahui.

Cubic Splines

- Diperoleh *4n* persamaan yang harus diselesaikan untuk mencari *4n* koefisien, a_i, b_i, c_i, d_i.
- Dimungkinkan untuk melakukan manipulasi matematis shg diperoleh suatu teknik *cubic splines* yang hanya memerlukan *n* – 1 penyelesaian (lihat uraian di buku acuan):
  - Chapra, S.P., Canale, R.P., 1985, *Numerical Methods for Engineers*, McGraw-Hill Book Co., New York, hlm. 395-396).

Cubic Splines

2 *unknowns* di setiap interval:

$f''(x_{i-1})$ dan $f''(x_i)$

$$
\begin{aligned}
f_i(x) = {} & \frac{f''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f''(x_{i-1})}{6(x_i - x_{i-1})}(x - x_{i-1})^3 \\
& + \left[\frac{f(x_{i-1})}{(x_i - x_{i-1})} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) \\
& + \left[\frac{f(x_i)}{(x_i - x_{i-1})} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x - x_{i-1})
\end{aligned}
$$

$$
\begin{aligned}
& (x_i - x_{i-1}) f''(x_{i-1}) + 2(x_{i+1} - x_{i-1}) f''(x_i) + (x_{i+1} - x_i) f''(x_{i+1}) = \\
& \qquad \frac{6}{(x_{i+1} - x_i)}\left[f(x_{i+1}) - f(x_i)\right] + \frac{6}{(x_i - x_{i-1})}\left[f(x_{i-1}) - f(x_i)\right]
\end{aligned}
$$

$$
\left.\begin{array}{l}
n \text{ interval} \\
f''(x_0) = 0 \\
f''(x_n) = 0
\end{array}\right\} \Rightarrow (n-1)\,\text{persamaan}
$$

Sekian