



Bab 5

Prediksi Beban Maksimum Mempergunakan Regresi Linier

Ref. [LILJA] Chap 8

Dr. Yeffry Handoko Putra , M.T



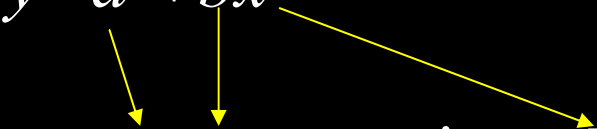
MODEL REGRESI LINIER

Model matematis yang digunakan untuk menyatakan hubungan kinerja komputer / variabel output (y) dengan metrik kinerja / variabel input (x) dan hubungannya adalah linier

Contoh Model Regresi Linier:

$$y = a + bx$$

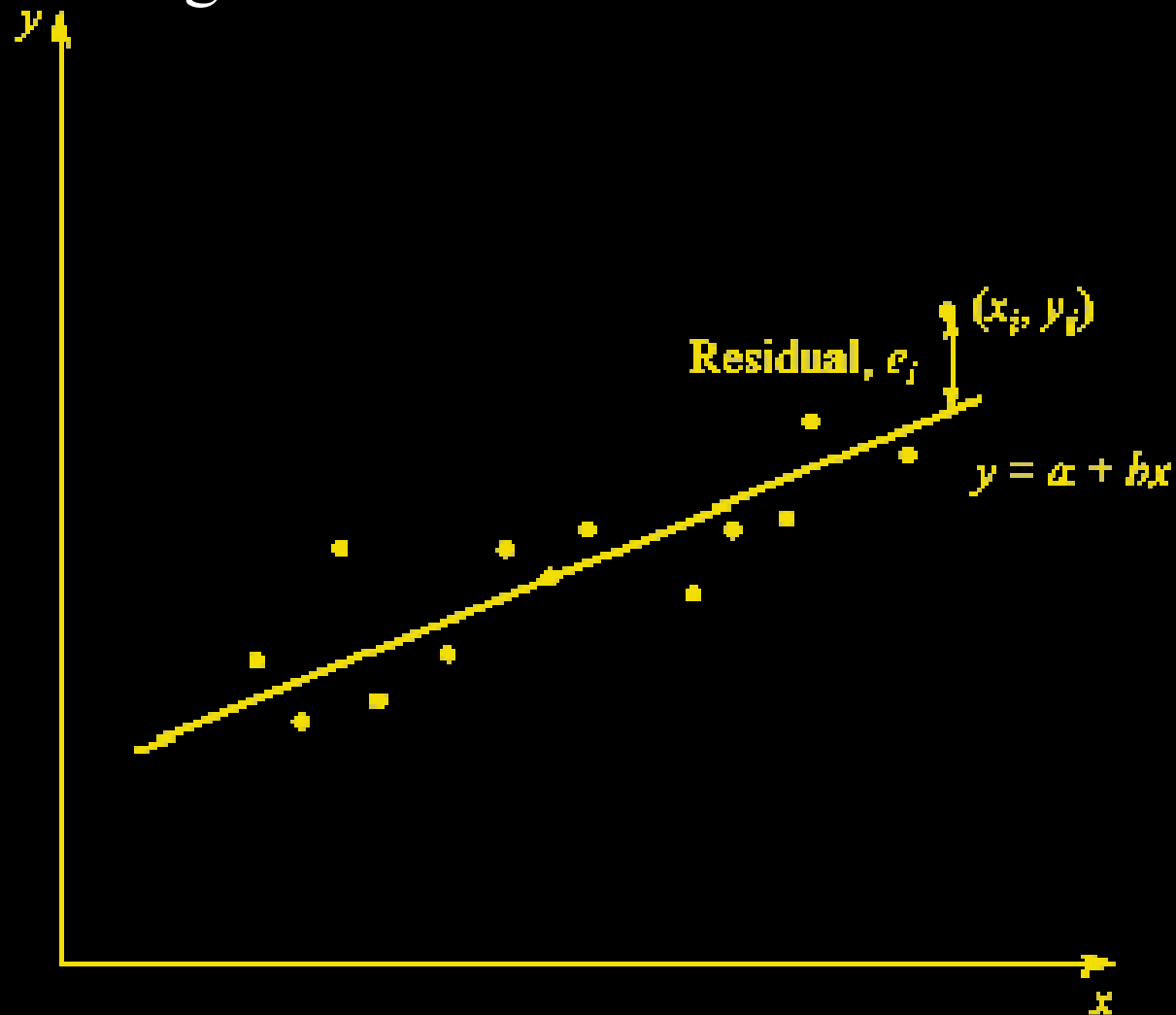
parameter regresi Input



Dalam pemodelan regresi linier diperoleh pasangan (x_i, y_i) sehingga tugas dalam pemodelan adalah mencari nilai a dan b



Plotting data terukur



Residual (e_i) adalah perbedaan antara garis model regresi linier dengan titik terukur

$$y_i = a + bx_i + e_i$$

untuk mencari garis linier yang paling mendekati pasangan titik ukur diperoleh dengan meminimalkan penjumlahan kuadrat (sum square) dari residu:

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

Minimasi diperoleh dengan turunan parsial dari terhadap a dan b dengan menyamakan dengan nol, diperoleh:



$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

Diselesaikan lebih lanjut diperoleh:

$$a + b\bar{x} = \bar{y}$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$



Contoh : waktu terukur yang diperlukan untuk membaca beberapa file dengan variasi ukuran:

Ukuran File (byte) (x_i)	Waktu (μs) (y_i)
10	3,8
50	8,1
100	11,9
500	55,6
1.000	99,6
5.000	500,2
10.000	1.006,1

Dapat dihitung



Contoh : waktu terukur yang diperlukan untuk membaca beberapa file dengan variasi ukuran:

Ukuran File (byte) (x_i)	Waktu (μ s) (y_i)
10	3,8
50	8,1
100	11,9
500	55,6
1.000	99,6
5.000	500,2
10.000	1.006,1
$\sum_{i=1}^7 x_i = 16.660$	$\sum_{i=1}^7 y_i = 1.685,3$



$$\sum_{i=1}^7 x_i y_i = 12.691,033;$$

$$\sum_{i=1}^7 x_i^2 = 126.262.600$$

$$\bar{x} = 2.380$$

$$\bar{y} = 240,76$$



Dapat dihitung:

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b = \frac{(7)(12.691,033) - (16.660)(1685,3)}{(7)(126.262.600) - (16.600)^2} = 0,1002$$

$$a = 240,76 - (0,1002)(2.380) = 2,24$$

Jadi Model Regresi Linier:

$$y = 2,24 + 0,1002x$$



Dari model dapat diprediksi berapa waktu untuk membaca
Beban maksimum, yaitu saat file berukuran 3000 byte
 $y = 2,24 + (0,1002)(3000) = 303\mu s$



Mencari Regresi Linier Menggunakan Matriks



Regresi Linier

$$y = a + bx$$

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

Regresi Parabola (polinomial regression)

$$y = a_0 + a_1x + a_2x^2$$

$$\begin{bmatrix} N & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Pelajari mencari Matriks Invers

Mencari Invers matriks 2x2

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Mencari Invers matriks 3x3

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = adj \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$$

dimana

$$adj \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} = \begin{bmatrix} + \begin{vmatrix} e & h \\ f & i \end{vmatrix} & - \begin{vmatrix} b & h \\ c & i \end{vmatrix} & + \begin{vmatrix} b & e \\ c & f \end{vmatrix} \\ - \begin{vmatrix} d & g \\ f & i \end{vmatrix} & + \begin{vmatrix} a & g \\ c & i \end{vmatrix} & - \begin{vmatrix} a & d \\ c & f \end{vmatrix} \\ + \begin{vmatrix} d & g \\ e & h \end{vmatrix} & - \begin{vmatrix} a & g \\ b & h \end{vmatrix} & + \begin{vmatrix} a & d \\ b & e \end{vmatrix} \end{bmatrix}$$

Interval Kepercayaan (Confidence Interval) dari parameter regresi linier (a dan b)

Diperoleh dengan

- Membagi jumlah kuadrat residu dengan derajat kebebasan:

$$s^2 = \frac{SSE}{n - 2}.$$

jika dikembangkan:

$$\begin{aligned} SSE &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 = \sum_{i=1}^n [(y_i - \bar{y}) - b(x_i - \bar{x})]^2 \\ &= S_{yy} - 2bS_{xy} + b^2S_{xx} = S_{yy} - bS_{xy} \end{aligned}$$

Dengan:



$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

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

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

Diperoleh varian regresi linier:

$$s^2 = \frac{SSE}{n-2} = \frac{S_{yy} - bS_{xy}}{n-2}$$

Diperoleh interval kepercayaan untuk b:

$$(b_1, b_2) = b \mp \frac{t_{[1-\alpha/2; n-2]} S}{\sqrt{S_{xx}}};$$

dengan $t_{[1-\alpha/2; n-2]}$ adalah distribusi t untuk derajat kebebasan n-2

Interval kepercayaan untuk a:

$$(a_1, a_2) = a \mp \frac{t_{[1-\alpha/2; n-2]} S \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n S_{xx}}}.$$

Model prediksi dengan interval kepercayaan menjadi :

$$(y_{p1}, y_{p2}) = y_p \mp t_{[1-\alpha/2; n-2]} s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}}.$$



Contoh : dari contoh sebelumnya hitung interval kepercayaan masing-masing parameter:

Jawab:

$$S_{xx} = 126,262,600 - \frac{(16,660)^2}{7} = 86,611,800$$

$$S_{yy} = 1,275,670.43 - \frac{(1,685.3)^2}{7} = 869,922.42$$

$$S_{xy} = 12,691,033 - \frac{(16,660)(1,685.3)}{7} = 8,680,019.$$

Varian parameter regresi:

$$s^2 = \frac{869,922.42 - 0.1002(8,680,019)}{7 - 2} = 36.9027.$$

atau



Standar deviasinya:

$$s = \sqrt{36.9027} = 6.0748.$$

90% Interval kepercayaan (Confidence level) :

$$(b_1, b_2) = 0.1002 \mp \frac{2.015(6.0748)}{\sqrt{86,611,800}} = 0.1002 \mp 0.0013 = (0.099, 0.102)$$

$$(a_1, a_2) = 2.24 \mp \frac{2.015(6.0748)\sqrt{126,262,600}}{\sqrt{7(8,680,019)}} = 2.24 \mp 5.59 = (-3.35, 7.83)$$

Jadi untuk pembacaan file berukuran 3000 byte diprediksi:

$$\begin{aligned}(y_{3000;1}, y_{3000;2}) &= 303 \mp 2.015(6.0748)\sqrt{1 + \frac{1}{7} + \frac{(3,000 - 2,380)^2}{86,611,800}} \\ &= 303 \mp 13.11 = (290, 316).\end{aligned}$$

Jadi dengan 90% kepercayaan untuk file berukuran 3000 byte diperlukan waktu antara 290-316 μ s

