Pencocokan Kurva (I): Regresi

Heri Purnawan

Disampaikan pada matakuliah **Metode Numerik**
Program Studi S-1 Teknik Elektro
Fakultas Sains dan Teknologi (FST)
Universitas Islam Lamongan (UNISLA)

November 18, 2024

Email: heripurnawan@unisla.ac.id

Pendahuluan ke Regresi

Bagaimana kita bisa memprediksi nilai suatu variabel ketika kita hanya memiliki data historis yang terbatas?

- **Tujuan Pembelajaran**: Menjelaskan konsep dasar regresi dan bagaimana teknik ini memungkinkan kita untuk memahami, memprediksi, dan mengendalikan berbagai sistem.

Apa itu Regresi?

- **Definisi Singkat**: "Regresi adalah teknik statistik untuk menemukan hubungan antara variabel input (independen) dan output (dependen)."
- **Contoh Nyata**: Misalnya, memprediksi tegangan keluaran berdasarkan arus masukan pada suatu rangkaian, atau memperkirakan konsumsi daya berdasarkan waktu.
- **Tujuan Regresi**: Memprediksi nilai variabel yang belum diketahui. Memahami pola dalam data untuk pengambilan keputusan.

Mengapa Regresi Penting dalam Teknik Elektro?

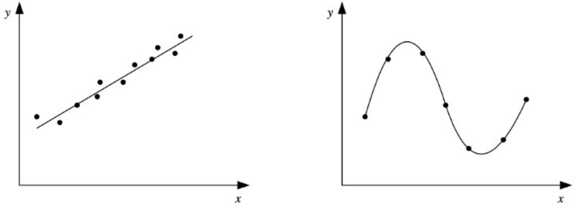
- Aplikasi Regresi:
  - **Kalibrasi Sensor**: Memperbaiki akurasi pengukuran yang sering kali dipengaruhi oleh faktor eksternal.
  - **Prediksi Beban dan Konsumsi Energi**: Untuk manajemen daya dan perencanaan beban listrik.
  - **Pemodelan Komponen Nonlinier**: Menyederhanakan pemodelan komponen kompleks seperti transistor atau motor.
  - **Pemeliharaan Prediktif**: Menggunakan data historis untuk memprediksi kapan peralatan akan membutuhkan perbaikan atau penggantian.
- **Manfaat**: Regresi membuka wawasan tentang bagaimana data digunakan dalam perancangan dan analisis sistem nyata.

Kapan Kita Membutuhkan Regresi?

- Data Historis yang tidak lengkap.
  Ketika data langsung sulit diperoleh atau terbatas.
- Hubungan yang kompleks antara variabel.
  Saat hubungan antara variabel tidak jelas atau tidak linier, regresi membantu kita membentuk pemahaman.
- Prediksi dan pengendalian.
  Untuk memprediksi hasil atau mengendalikan sistem yang bervariasi dari waktu ke waktu.
- Contoh Kasus
  - Analisis tren konsumsi energi dalam jaringan
  - Prediksi degradasi komponen
  - Model dinamis respon sistem.

Regresi dan Interpolasi

- Regresi
  - digunakan apabila sumber data yang digunakan mempunyai ketelitian yang cukup rendah.
  - kurva yang dibangun tidak perlu melalui semua titik data tersebut, tetapi cukup mengikuti kecendrungannya saja.
- Interpolasi
  - digunakan apabila sumber data yang digunakan mempunyai ketelitian yang sangat tinggi.
  - kurva yang dibangun harus melalui semua titik data.



(a) Regresi (b) Interpolasi

Bagaimana Regresi Bekerja?

- Ide Dasar: Mencari garis atau kurva terbaik yang “paling cocok” dengan data yang tersedia.
- Visualisasi Awal: Tampilkan grafik sederhana yang menunjukkan data titik dan garis yang mewakili model regresi.
- Intuisi: Semakin dekat model regresi dengan data aktual, semakin baik prediksi dan analisis yang dapat dihasilkan.
- Jenis-Jenis Regresi yang akan dipelajari
  - Regresi Linier
  - Regresi Eksponensial
  - Regresi Persamaan Berpangkat
  - Regresi Polinomial

Regresi linier

- Diberikan sekumpulan observasi berpasangan sebanyak n:

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

- Titik-titik data dihampiri oleh sebuah garis lurus (disebut kurva regresi) yang dinyatakan sebagai

$$y = f(x) = a_0 + a_1 x$$

- **Pertanyaan**: berapa nilai a_0 dan a_1 agar garis regresi tersebut sedekat mungkin dengan titik-titik data yang diberikan (meminimumkan galat)?
- Metode yang digunakan dalam regresi linier adalah Metode Kuadrat Terkecil (*least square method*). Tujuan dari metode ini adalah untuk membuat kesalahan yang terjadi sekecil mungkin.

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{data}} - y_{i,\text{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \qquad (1)$$

Penentuan nilai a_0 dan a_1

- Menentukan nilai a_0 dan a_1 dari Pers. (1):

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) = 0 \tag{2}$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n}[(y_i - a_0 - a_1 x_i)x_i] = 0 \tag{3}$$

- Pers. (2) dan (3) dapat dituliskan menjadi:

$$\begin{aligned} na_0 + a_1\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{4}$$

- Solusi dari SPL (4) untuk a_1 dan a_0 adalah

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

Regresi linier: Contoh 1

1. Tentukan regresi linier dari data yang diberikan berikut ini.

x_i	1	2	3	4	5	6	7
y_i	0,5	2,5	2	4	3,5	6	5,5

2. Tentukan galat dari garis regresi tersebut.

Solusi:

1. Dari data yang diberikan diperoleh:

$$n = 7 \ ; \quad \sum_{i=1}^{n} x_i = 28 \ ; \quad \sum_{i=1}^{n} y_i = 24 \ ; \sum_{i=1}^{n} x_i y_i = 119,5 \ ; \quad \sum_{i=1}^{n} x_i^2 = 140$$

sehingga,

$$a_1 = \frac{7(119,5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = \frac{24 - 0.8392857(28)}{7} = 0.07142857$$

Pers. regresi linier: $y = 0.07142857 + 0.8392857x$

2. Tugas anda untuk mengecek galatnya

Koefisien korelasi

- Koefisien korelasi (r) mempunyai interval dari 0 sampai 1. Semakin mendekati nilai 1 maka r nya semakin baik.

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

dimana:

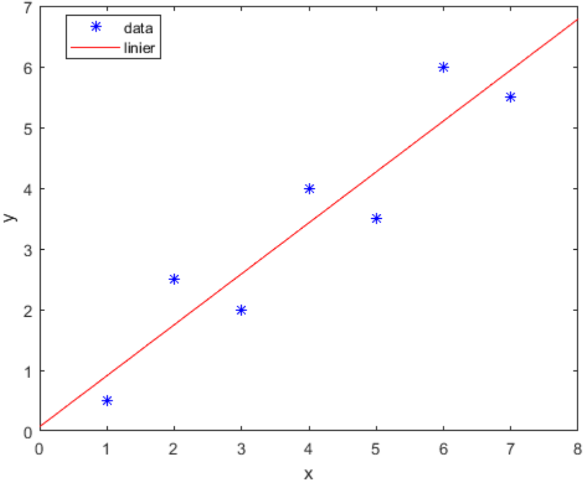
$$S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Dari Contoh 1, diperoleh:

$$r = \sqrt{\frac{22.7143 - 2.9911}{22.7143}} = \sqrt{0.868} = 0.932$$

Meskipun koefisien korelasi memberikan ukuran kesesuaian yang berguna, Anda harus berhati-hati untuk tidak menganggapnya lebih bermakna daripada yang seharusnya. Hanya karena r "mendekati" 1 tidak berarti bahwa kecocokan tersebut selalu "baik". Anda harus selalu memeriksa plot data beserta kurva regresi Anda.

Regresi linier: kurva vs data



Regresi Eksponensial

- Bentuk umum:

$$y = ae^{bx}$$

- Pelinieran:

$$\ln y = \ln(ae^{bx}) = \ln a + \ln e^{bx} = \ln a + bx$$

- Definisikan variabel baru:

$$\tilde{y} = \ln y; \quad a_0 = \ln a; \quad a_1 = b$$

- Diperoleh bentuk regresi linier:

$$\tilde{y} = a_0 + a_1 x$$

Nilai a_0 dan a_1 dapat dihitung seperti pada metode regresi linier yaitu

$$a_1 = \frac{n\sum_{i=1}^{n} x_i\tilde{y}_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \tilde{y}_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad \text{dan} \quad a_0 = \frac{\sum_{i=1}^{n} \tilde{y}_i - a_1 \sum_{i=1}^{n} x_i}{n}$$

Selanjutnya, hitung

$$a = e^{a_0} \quad \text{dan} \quad b = a_1$$

Regresi Eksponensial: Contoh 1

x_i	1	2	3	4	5	6	7
y_i	0,5	2,5	2	4	3,5	6	5,5
$\tilde{y}_i = \ln y_i$	-0,6931	0,9163	0,6931	1,3863	1,2528	1,7918	1,7047
$x_i\tilde{y}_i$	-0,6931	1,8326	2,0794	5,5452	6,2638	10,7506	11,9332

1. Dari tabel diperoleh:

$$\sum_{i=1}^{7} \tilde{y}_i = 7,0519; \; \sum_{i=1}^{7} x_i = 28; \; \sum_{i=1}^{7} x_i\tilde{y}_i = 37,7117$$

sehingga,

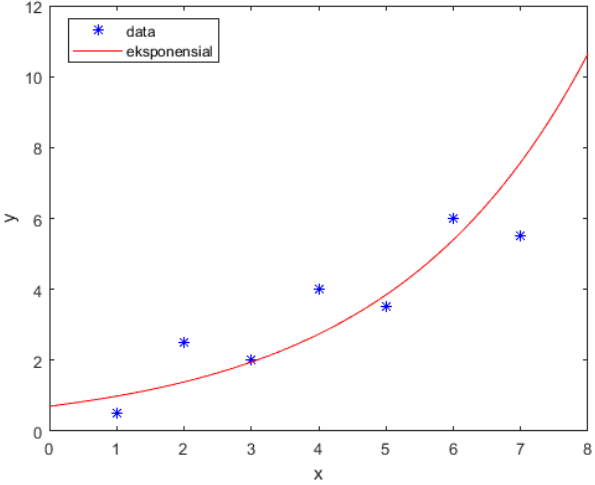
$$a_1 = \frac{7(37,7117) - 28(7,0519)}{7(140) - (28)^2} = 0,3394$$

$$a_0 = \frac{7,0519 - 0,3394(28)}{7} = -0,3502$$

Jadi, nilai $a = e^{a_0} = 0,7046$ dan $b = 0,3394$
Pers. regresi eksponensial: $y = 0,7046e^{0,3394x}$

2. Tugas anda untuk mengecek galatnya.

Regresi eksponensial: kurva vs data



Regresi Persamaan Berpangkat

- Bentuk umum:

$$y = ax^b$$

- Pelinieran:

$$\log y = \log(ax^b) = \log a + \log x^b = \log a + b \log x$$

- Definisikan variabel baru:

$$\tilde{y} = \log y; \quad a_0 = \log a; \quad a_1 = b; \quad \tilde{x} = \log x$$

- Diperoleh bentuk regresi linier:

$$\tilde{y} = a_0 + a_1 \tilde{x}$$

Nilai a_0 dan a_1 dapat dihitung seperti pada metode regresi linier. Selanjutnya, hitung

$$a = 10^{a_0} \quad \text{dan} \quad b = a_1$$

Regresi Persamaan Berpangkat: Contoh 1

x_i	1	2	3	4	5	6	7
y_i	0,5	2,5	2	4	3,5	6	5,5
$\tilde{x}_i = \log x_i$	0	0,3010	0,4771	0,6021	0,6990	0,7782	0,8451
$\tilde{x}_i^2$	0	0,0906	0,2276	0,3625	0,4886	0,6055	0,7142
$\tilde{y}_i = \log y_i$	-0,3010	0,3979	0,3010	0,6021	0,5441	0,7782	0,7404
$\tilde{x}_i\tilde{y}_i$	0	0,1198	0,1436	0,3625	0,3803	0,6055	0,6257

1. Dari tabel diperoleh:

$$\sum_{i=1}^{7} \tilde{y}_i = 3,0626; \sum_{i=1}^{7} \tilde{x}_i = 3,7024; \sum_{i=1}^{7} \tilde{x}_i\tilde{y}_i = 2,2374; \sum_{i=1}^{7} \tilde{x}_i^2 = 2,4890$$

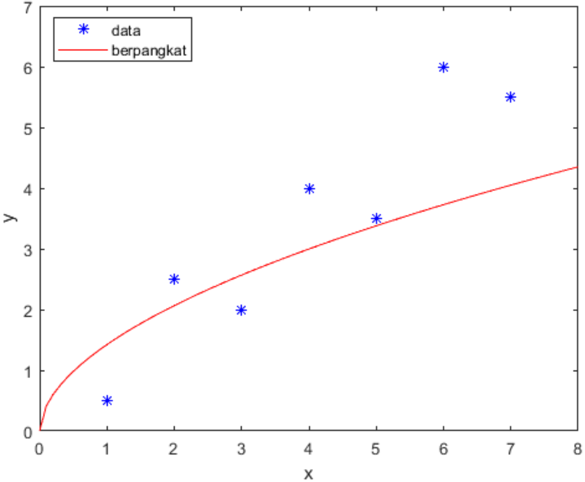
sehingga, $a_1 = \dfrac{7(2,2374) - 3,7024(3,0626)}{7(2,4890) - (3,0626)^2} = 0,5374 \rightarrow b = a_1 = 0,5374$

$$a_0 = \frac{3,0626 - 0,5374(3,7024)}{7} = 0,1533 \rightarrow a = 10^{a_0} = 1,4232$$

Pers. regresi persamaan berpangkat: $y = 1,4232x^{0,5374}$

2. Tugas anda untuk mengecek galatnya.

Regresi persamaan berpangkat: kurva vs data



Regresi Polinomial

- Perhatikan kembali n buah titik data

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Prosedur kuadrat terkecil pada regresi linier akan diperluas untuk membangun kurva regresi polinom berderajat m yang dinyatakan dengan

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

Catatan: nilai m dan n tidak ada hubungan tertentu.

- Terapkan rumus galat sebagai berikut:

$$E = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_m x_i^m)\right)^2}$$

Regresi Polinomial (lanj.)

- Akan ditentukan nilai $a_0, a_1, \ldots, a_m$ yang meminimumkan nilai $D = nE^2$. Hal ini tercapai ketika

$$\frac{\partial D}{\partial a_0} = 0, \frac{\partial D}{\partial a_1} = 0, \ldots, \frac{\partial D}{\partial a_m} = 0$$

- Langkah di atas menghasilkan persamaan normal berupa SPL dalam a_0, $a_1, \ldots, a_m$ sebagai berikut:

$$\begin{array}{ccccccccc}
na_0 & + & \sum_{i=1}^{n} x_i a_1 & + & \sum_{i=1}^{n} x_i^2 a_2 & + \cdots + & \sum_{i=1}^{n} x_i^m a_m & = & \sum_{i=1}^{n} y_i \\
\sum_{i=1}^{n} x_i a_0 & + & \sum_{i=1}^{n} x_i^2 a_1 & + & \sum_{i=1}^{n} x_i^3 a_2 & + \cdots + & \sum_{i=1}^{n} x_i^{m+1} a_m & = & \sum_{i=1}^{n} x_i y_i \\
\sum_{i=1}^{n} x_i^2 a_0 & + & \sum_{i=1}^{n} x_i^3 a_1 & + & \sum_{i=1}^{n} x_i^4 a_2 & + \cdots + & \sum_{i=1}^{n} x_i^{m+2} a_m & = & \sum_{i=1}^{n} x_i^2 y_i \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
\sum_{i=1}^{n} x_i^m a_0 & + & \sum_{i=1}^{n} x_i^{m+1} a_1 & + & \sum_{i=1}^{n} x_i^{m+2} a_2 & + \cdots + & \sum_{i=1}^{n} x_i^{m+m} a_m & = & \sum_{i=1}^{n} x_i^m y_i
\end{array}$$

- SPL di atas diselesaikan dengan salah satu teknik penyelesaian SPL yang sudah dibahas sebelumnya.

Regresi Polinomial (derajat 2): Contoh 1

Misalkan digunakan polinomial derajat 2 maka bentuk persamaan regresinya adalah

$$y = a_0 + a_1 x + a_2 x^2$$

diperoleh bentuk SPL sebagai berikut:

$$\begin{array}{ccccccc}
n a_0 & + & a_1 \sum_{i=1}^{n} x_i & a_2 \sum_{i=1}^{n} x_i^2 & = & \sum_{i=1}^{n} y_i \\
a_0 \sum_{i=1}^{n} x_i & + & a_1 \sum_{i=1}^{n} x_i^2 & a_2 \sum_{i=1}^{n} x_i^3 & = & \sum_{i=1}^{n} x_i y_i \\
a_0 \sum_{i=1}^{n} x_i^2 & + & a_1 \sum_{i=1}^{n} x_i^3 & a_2 \sum_{i=1}^{n} x_i^4 & = & \sum_{i=1}^{n} x_i^2 y_i
\end{array}$$

atau dalam bentuk matriks

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix}$$

Regresi Polinomial (derajat 2): Contoh 1 (lanj.)

x_i	1	2	3	4	5	6	7
y_i	0,5	2,5	2	4	3,5	6	5,5
x_i^2	1	4	9	16	25	36	49
x_i^3	1	8	27	64	125	216	343
x_i^4	1	16	81	256	625	1296	2401
$x_i y_i$	0,5	5,0	6,0	16,0	17,5	36,0	38,5
$x_i^2 y_i$	0,5	10	18	64	87,5	216	269,5

dari tabel diperoleh ($n = 7$):

$$\sum_{i=1}^{n} x_i = 28; \sum_{i=1}^{n} y_i = 24; \sum_{i=1}^{n} x_i^2 = 140; \sum_{i=1}^{n} x_i^3 = 784; \sum_{i=1}^{n} x_i^4 = 4676;$$

$$\sum_{i=1}^{n} x_i y_i = 119,5; \sum_{i=1}^{n} x_i^2 y_i = 665,5$$

sehingga,

$$\begin{bmatrix} 7 & 28 & 140 \\ 28 & 140 & 784 \\ 140 & 784 & 4676 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 24 \\ 119,5 \\ 665,5 \end{bmatrix} \rightarrow \begin{aligned} a_0 &= -0,2857 \\ a_1 &= 1,0774 \\ a_2 &= -0,0298 \end{aligned}$$

Regresi Polinomial (derajat 2): kurva vs data

Regresi polinomial: $y = -0,2857 + 1,0774x - 0,0298x^2$