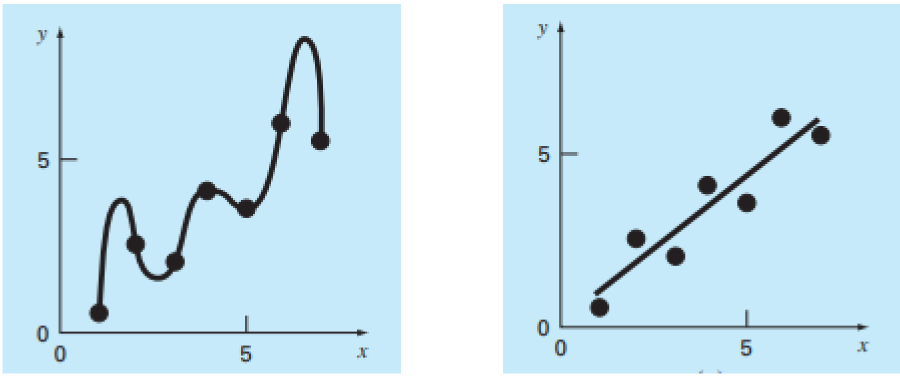
Modul 4

Curve Fitting

A. Teori

Pada eksperimen atau pengamatan, biasanya diperoleh data dengan kesalahan pengukuran yang dapat bersifat sistematis maupun acak. Hubungan variabel-variabel fisis pada data dapat diperoleh de ngan pencocokan kurva atau regresi. Berbeda dengan interpolasi, regresi lebih tahan terhadap kesalahan pada pengukuran data yang memiliki noise, sedangkan interpolasi cenderung digunakan untuk mendapatkan data yang lebih banyak diantara titik-titik data yang diketahui. Dengan metode regresi, kesalahan pengukuran dapat secara rata-rata dikurangi. Orde polinom atau fungsi yang biasa digunakan untuk regresi jauh lebih kecil dari jumlah data. Khususnya hubungan yang paling sederhana menggunakan polinom order satu atau regresi linier selalu digunakan sebelum fungsi nonlinier diterapkan. Hal ini karena apabila orde yang digunakan dalam regresi sama dengan jumlah data, maka sifatnya akan mirip dengan interpolasi dimana fungsi yang dibangun akan melalui semua titik data yang ada, inilah yang disebut sebagai *overfitting.*



Gambar 1. Gambar kiri menunjukkan interpolasi gambar kanan menunjukkan regresi

1. Regresi Linear

Regresi linear adalah suatu teknik analisis statistik yang digunakan untuk menentukan hubungan antar dua variabel dengan bentuk pendekatan garis lurus. Dimana bentuk matematis dari garis lurus adalah

$$f(x) = A + Bx$$

Misalnya diperoleh suatu data pengukuran dengan suatu sensor (x_0, y_0), (x_1, y_1), (x_2, y_2), ... (x_n, y_n). Data pengukuran tersebut tercampur dengan noise, sehingga diperlukan suatu fungsi pendekatan yang dapat merepresentasikan fungsi sesungguhnya dengan metode regresi linear. Pada dasarnya untuk melakukan hal tersebut, fungsi yang akan dibangun harus memiliki error terkecil dari masing-masing titik data. Dimana error tersebut dapat dicari dengan persamaan

$$S_r = \sum_{i=1}^{n} \left(y_{i\,pengukuran} - y_{i\,model}\right)^2 = \sum_{i=1}^{n} \left(y_{i\,pengukuran} - (A + Bx_i)\right)^2$$

Nilai minimum dari error dapat dicari dengan melakukan diferensial. Dimana titik minimum dapat dicari apabila turunan pertama dari fungsi (fungsi error) sama dengan 0. Sehingga persamaannya menjadi

$$\frac{\partial S_r}{\partial A} = -2\sum_{i=1}^{n}\left(y_{i\,pengukuran} - A - Bx_i\right) = 0$$

$$\frac{\partial S_r}{\partial B} = -2\sum_{i=1}^{n}\left[\left(y_{i\,pengukuran} - A - Bx_i\right)x_i\right] = 0$$

Dalam beberapa buku atau sumber biasanya nilai 2 dalam penurunan nilai error dihilangkan karena fungsi error dapat ditulis juga sebagai berikut

$$S_r = \frac{1}{2}\sum_{i=1}^{n}\left(y_{i\,pengukuran} - y_{i\,model}\right)^2$$

Penambahan nilai $\frac{1}{2}$ di atas biasanya digunakan untuk mempermudah perhitungan saat dicari nilai minimumnya, agar nilai 2 dari turunan error bisa hilang. Sehingga fungsi turunannya dapat ditulis kembali sebagai berikut

$$\sum_{i=1}^{n}\left(y_{i\,pengukuran} - A - Bx_i\right) = 0$$

$$\sum y_i - \sum A - \sum Bx_i = 0$$

$$NA + B\sum x_i = \sum y_i$$

Dan

$$\sum_{i=1}^{n}\left[\left(y_{i\,pengukuran} - A - Bx_i\right)x_i\right] = 0$$

$$\sum y_i x_i - \sum Ax_i - \sum Bx_i^2 = 0$$

$$A\sum x_i + B\sum x_i^2 = \sum y_i x_i$$

Dengan begitu kita dapat menyelesaikan secara langsung maupun dengan pendekatan matriks.

$$\begin{pmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix}$$

Kemudian, koefisien nilai A dan B dapat diketahui dengan melakukan operasi penyelesaian matriks

2. Regresi Polinomial

Regresi polinomial adalah pendekatan yang sama dengan linear, hanya saja derajat dari fungsi yang dibangun bernilai lebih dari 1. Bentuk umum dari fungsi polinomial adalah

$$f(x) = A + Bx + Cx^2 + \cdots + Dx^n$$

Sebagai contoh, fungsi yang dibangun adalah fungsi polinomial berderajat dua atau kuadratik. Maka persamaan errornya adalah

$$\frac{\partial S_r}{\partial A} = -2\sum_{i=1}^{n}\left(y_{i\,pengukuran} - A - Bx_i - Cx_i^2\right) = 0$$

$$\frac{\partial S_r}{\partial B} = -2\sum_{i=1}^{n}\left[\left(y_{i\,pengukuran} - A - Bx_i - Cx_i^2\right)x_i\right] = 0$$

$$\frac{\partial S_r}{\partial C} = -2\sum_{i=1}^{n}\left[\left(y_{i\ pengukuran} - A - Bx_i\ - Cx_i^2\right)x_i^2\right] = 0$$

Jadi matriks sistem persamaan linear yang dibentuk adalah sebagai berikut

$$NA + B\sum x_i + C\sum x_i^2 = \sum y_i$$
$$A\sum x_i +\ B\sum x_i^2 + C\sum x_i^3 = \sum x_i y_i$$
$$A\sum x_i^2 +\ B\sum x_i^3 + C\sum x_i^4 = \sum x_i^2 y_i$$

Sehingga bentuk umum dari regresi polinomial dapat ditulis sebagai berikut

$$Na_0 + a_1\sum x_i + a_2\sum x_i^2 + \cdots + a_n\sum x_i^n = \sum y_i$$
$$a_0\sum x_i +\ a_1\sum x_i^2 + a_2\sum x_i^3 + \cdots + a_n\sum x_i^{n+1} = \sum x_i y_i$$
$$a_0\sum x_i^2 +\ a_1\sum x_i^3 + a_2\sum x_i^4 + \cdots + a_n\sum x_i^{n+2} = \sum x_i^2 y_i$$
$$\ldots$$
$$a_0\sum x_i^n +\ a_1\sum x_i^{n+1} + a_2\sum x_i^{n+2} + \cdots + a_n\sum x_i^{n+2} = \sum x_i^n y_i$$

3. **Regresi Eksponensial**

Regresi eksponensial adalah teknik yang digunakan untuk menentukan hubungan dua variabel yang kemungkinan memiliki hubungan eksponensial. Regresi eksponensial ini mengharuskan untuk mengubah fungsi eksponensial ke dalam bentuk linear terlebih dahulu. Kemudian dapat diselesaikan dengan metode regresi linear. Bentuk fungsi eksponensial secara umum adalah sebagai berikut:

$$f(x) = ae^{bx}$$

Fungsi eksponensial tersebut dapat diubah ke dalam bentuk persamaan linear menjadi seperti berikut

$$F(x) =\ \ln\left(f(x)\right) = \ln(a) + bx$$

Dengan begitu, dapat diperoleh nilai residu/error dari fungsi aproksimasi logaritma yang dibangun seperti berikut

$$R_i = \ln(y_i) -\ F(x_i)$$

Dimana $\ln(y_i)$ adalah data yang diperoleh dan $F(x_i)$ adalah fungsi yang dibangun. Untuk mencari nilai $\ln(a)$ dan b dapat menggunakan cara yang sama dengan regresi linear yang telah dijelaskan sebelumnya.

4. **Regresi Linear Multivariabel (Regresi Linear Berganda)**

Regresi Linear Berganda atau Multivariabel adalah regresi linear dengan variabel lebih dari satu. Namun semua variabelnya memiliki derajat 1. Sehingga bentuk persamaan linear berganda ditulis sebagai berikut

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Sebagai contoh terdapat dua variabel, maka fungsi linear berganda tersebut memiliki fungsi SSE yang dapat ditulis seperti berikut

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_{i\ Data} - a_0 - a_1x_{1i} - a_2x_{2i}) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n}\left[x_{1_i}(y_{i\,Data} - a_0 - a_1 x_{1i} - a_2 x_{2i})\right] = 0$$
$$\frac{\partial S_r}{\partial a_2} = -2\sum_{i=1}^{n}\left[x_{2_i}(y_{i\,Data} - a_0 - a_1 x_{1i} - a_2 x_{2i})\right] = 0$$

Dengan melihat bentuk persamaan di atas, maka matriks yang dapat dibangun dapat ditulis sebagai berikut

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1_i} x_{2_i} \\ \sum x_{2i} & \sum x_{1i} x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

5. **Jenis Kesalahan**

a. Kesalahan penjumlahan kuadrat (SSE) dapat dihitung dengan persamaan sebagai berikut

$$SSE = S_r = \sum_{i=1}^{n}\left(y_{i\,Data} - y_{i\,Approksimasi}\right)^2$$

b. Kesalahan rata-rata kuadrat (MSE) dapat dihitung dengan persamaan berikut

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i\,Data} - y_{i\,Approksimasi}\right)^2$$

c. Kesalahan total kuadrat (SST) dapat dihitung dengan persamaan berikut

$$SST = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i\,Data} - \mu_{y_{aproksimasi}}\right)^2$$

Dimana $\mu_{y_{aproksimasi}}$ = rata-rata nilai aproksimasi

d. Koefisien Determinasi (R^2) dapat dihitung dengan persamaan berikut

$$R^2 = 1 - \frac{SSE}{SST}$$

e. Standar Deviasi (σ)

$$\sigma = \sqrt{\frac{SSE}{n-2}}$$

Dengan n adalah jumlah data

B. Persoalan

Penemuan Dark Matter dimulai pada akhir abad ke 19 ketika para ilmuwan mulai memperhatikan bahwa ada sesuatu yang aneh dengan pergerakan benda-benda. Dimana rotasi galaksi Bima Sakti dan galaksi-galaksi spiral lain umumnya menunjukkan profile seperti benda tegar pada radius kecil dan mendata pada radius besar. Berikut adalah data pengukuran laju rotasi bintang-bintang di suatu galaksi dan jarak bintang ke pusat galaksi tersebut.

Radius dari Pusat (kpc)	Laju Rotasi (km/s)
0.5	40.226
1.43	103.33

2.86	134.14
4.29	134.25
5.71	107.57
7.14	112.66
8.57	108.78
11.42	130.05
12.86	126.37
14.29	112.81
15.71	120.29
17.14	110.44
18.57	116.30
20.00	118.56

Dari data yang diperoleh di atas :

a. Tentukan fungsi aproksimasi dengan regresi polinomial (silahkan pilih 3 orde secara bebas, kemudian bandingkan ketiga fungsi aproksimasi tersebut dengan menganalisis poin b dan c)
b. Tentukan *errornya*, gunakan SSE dan SST, kemudian cari nilai standar deviasinya dari tiga fungsi aproksimasi yang dibuat.
c. Buat grafik dari ketiga fungsi aproksimasi dengan grafik *errorbar*nya (sebagai contoh bisa menggunakan plt.errorbar() dari modul matplotlib)

C. Bentuk Laporan

[Markdown/Text] 1. Problem Statement

[Markdown/Text] 2. Persamaan Matematika

[Markdown/Text] 3. Algoritma Metode Regresi yang digunakan

[Markdown/Text] 4. Diagram Alir Regresi yang digunakan

[Code] 5. Program Metode Regresi yang digunakan

[Markdown/Text/Code] 6. Analisis Hasil Regresi yang digunakan (diperbolehkan untuk menambahkan program yang mendukung pernyataan analisis semisal analisis error, konsumsi waktu, dan beban memori)