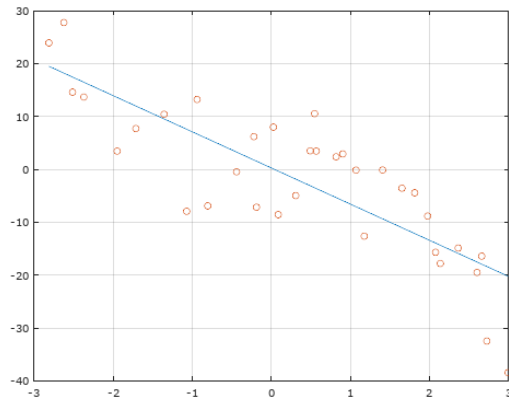


Nama : Aditya Alif Nugraha
NIM : 1301154183
Kelas : IF 39-01

1. Diberikan data training 1 (x_i, y_i) terdapat pada file 'trainset1-[NomorUrutAndaDiDaftarHadirKelas].csv'. Juga diberikan data test 1 (x_i) terdapat pada file 'testset1-[NomorUrutAndaDiDaftarHadirKelas].csv'.

(a) (15 points) Bangunlah model regresi univariate linear menggunakan data training 1.



Dari model yang telah dibangun, maka dihasilkan persamaan untuk garis linier sebagai berikut:

$$y = -6.8400x + 0.26142$$

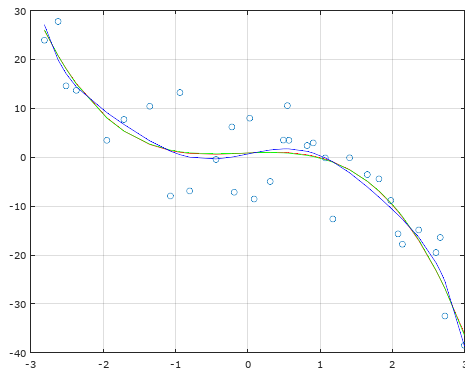
Dengan nilai **Sum Squared Error (SSE)**: 2214.7

(b) (5 points) Lakukan prediksi terhadap data test 1 menggunakan model regresi univariate linear yang telah dibangun pada 1(a).

Jawaban terurut dari baris pertama hingga akhir:

18.8525
6.1402
-3.3237
-11.8111
-18.8836

(c) (15 points) Bangunlah model regresi univariate non-linear menggunakan data training 1. (Petunjuk: gunakan polinomial pangkat/derajat p , besarnya p berdasarkan analisis anda)



Garis merah menunjukkan polinomial 3, hijau menunjukkan polinomial 5, dan biru polinomial 7. Polinomial yang dibandingkan hanya polinomial ganjil karena bentuk dari persebaran data seperti huruf S.

Walaupun nilai error yang paling minimum ada di polinomial 7, dikhawatirkan akan menjadi overfitting karena semakin besar polinomial, garis yang terbentuk akan semakin mendekati titik data. Sehingga dipilih polinomial 5 yang memiliki nilai error lebih kecil dari polinomial 3 dan garis yang terbentuk tidak mendekati titik data sehingga tidak overfit.

Sehingga didapatkan persamaan garis sebagai berikut:

$$y_i = f(x_i) = 8.6168e-001 + 4.8842e-001x_i - 3.3682e-001x_i^2 + -1.2211e+000x_i^3 + -5.9576e-004x_i^4 + -1.1948e-002x_i^5$$

Dengan nilai **Sum Squared Error (SSE)**: 1164.5

(d) (5 points) Lakukan prediksi terhadap data test 1 menggunakan model regresi univariate non-linear yang telah dibangun pada 1(c).

Jawaban terurut dari baris pertama hingga akhir:

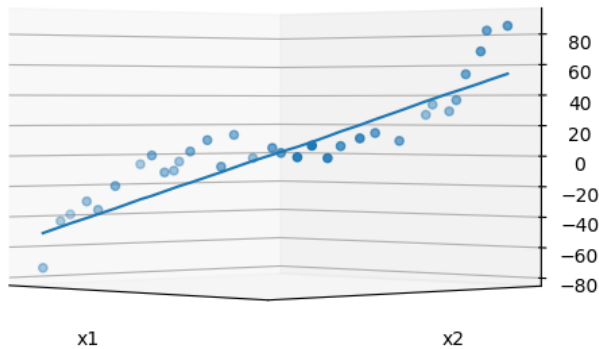
23.30443
0.97364
0.84880
-6.24997
-29.27593

(e) (10 points) Diantara dua model regresi yang dihasilkan pada 1(a) dan 1(c), menurut anda model mana yang memberikan prediksi lebih baik untuk data test 1? Berikan penjelasan anda secara lengkap.

- Dari kedua model diatas, **saya memilih model univariate non-linear**. Karena memiliki nilai SSE yang lebih kecil dan sesuai dengan bentuk persebaran data pada data train. Sehingga saat dilakukan prediksi akan lebih akurat.

2. Diberikan data training 2 (x_{i1} , x_{i2} , y_i) terdapat pada file 'trainset2 [NomorUrutAndaDiDaftarHadirKelas].csv'. Juga diberikan data test 2 (x_{i1} , x_{i2}) terdapat pada file 'testset2- [NomorUrutAndaDiDaftarHadirKelas].csv'.

(a) (15 points) Bangunlah model regresi multivariate linear menggunakan data training 2.



Dari model yang telah dibangun, maka dihasilkan persamaan untuk garis linier sebagai berikut:

$$y = 1.6441x_1 + 2.0457x_2$$

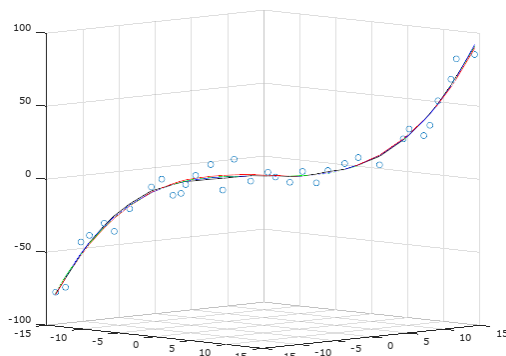
Dengan **Sum Squared Error (SSE)**: 8526.4

(b) (5 points) Lakukan prediksi terhadap data test 2 menggunakan model regresi multivariate linear yang telah dibangun pada 2(a).

Jawaban terurut dari baris pertama hingga akhir:

-53.0395
-24.4854
-2.3297
26.5717
52.4667

(c) (15 points) Bangunlah model regresi multivariate non-linear menggunakan data training 2. (Petunjuk: gunakan analisis anda untuk membangun model regresi dengan melibatkan interaksi antar atribut input. Salah satu contohnya dijelaskan pada slide Regresi halaman 17. Anda dapat membuat interaksi yang lain antar atribut input)



Dari model yang dibangun, didapat persamaan garis sebagai berikut:

$$y_i = f(x_i) = 2.2451e+000 + 3.6389e-001 x_{i1} + -3.3711e-001 x_{i2} + 1.0094e-002 x_{i1}x_{i2} + 3.0397e-003 (x_{i1} + x_{i2})^2 + 5.2254e-007 (x_{i1} + x_{i2})^3$$

Dengan nilai **Sum Squared Error (SSE)**: 1303.4

Model tersebut dipilih karena memiliki nilai error yang terkecil dan perbedaan error terhadap polynomial yang lainnya tidak terlalu signifikan. Dan polynomial yang terlalu tinggi akan membuat model terpacu terhadap data training.

(d) (5 points) Lakukan prediksi terhadap data test 2 menggunakan model regresi multivariate non-linear yang telah dibangun pada 2(c).

Jawaban terurut dari baris pertama hingga akhir:

-78.3205
-4.4721
2.0628
12.7692
83.4084

(e) (10 points) Diantara dua model regresi yang dihasilkan pada 2(a) dan 2(c), menurut anda model mana yang memberikan prediksi lebih baik untuk data test 2? Berikan penjelasan anda secara lengkap.

- Dari kedua model diatas, **saya memilih model multivariate non-linear**. Karena memiliki nilai SSE yang lebih kecil dan sesuai dengan bentuk persebaran data pada data train. Sehingga saat dilakukan prediksi akan lebih akurat.