Machine Learning
Assignment 2
CLO2

Ida Bagus Dwi Satria Kusuma
1301140297

October 25, 2017

1. Given data set 1 (x_i, y_i) shown in Table 1.

   (a) (20 points) Build a univariate linear regression model using data set 1 (except p4).
   Untuk membuat regresi linear [2]

   $$f(x) = \omega_1 x + \omega_0 \ ,$$

   kita perlu mengetahui nilai ω_1 dan ω_0 terlebih dahulu. Untuk mencari nilai ω_1 dan ω_0, kita dapat menggunakan *normal equation* [2]

   $$\begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

   yang kemudian disesuaikan untuk mencari nilai ω_1 dan ω_0 ,

   $$\begin{pmatrix} \hat{\omega_0} \\ \hat{\omega_1} \end{pmatrix} = \begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

   Setelah melakukan perhitungan, diketahui fakta berikut :

   $$N = 6 \ , \ \sum_i x_i = 33.2 \ , \ \sum_i x_i^2 = 209.74 \ , \ \sum_i y_i = 19.2 \ , \ \sum_i x_i y_i = 102.5$$

Kemudian, dengan menggunakan persamaan normal equation[2], kita memperoleh

$$\begin{pmatrix} \hat{\omega_0} \\ \hat{\omega_1} \end{pmatrix} = \begin{pmatrix} 6 & 33.2 \\ 33.2 & 209.74 \end{pmatrix}^{-1} \begin{pmatrix} 19.2 \\ 102.5 \end{pmatrix}$$

Diketahui :

$$\begin{pmatrix} 6 & 33.2 \\ 33.2 & 209.74 \end{pmatrix}^{-1} = \frac{1}{1258.44 - 1102.24} \begin{pmatrix} 209.74 & -33.2 \\ -33.2 & 6 \end{pmatrix}$$

$$= \frac{1}{156.2} \begin{pmatrix} 209.74 & -33.2 \\ -33.2 & 6 \end{pmatrix}$$

$$= \begin{pmatrix} 1.343 & -0.213 \\ -0.213 & 0.038 \end{pmatrix}$$

Maka :

$$\begin{pmatrix} \hat{\omega_0} \\ \hat{\omega_1} \end{pmatrix} = \begin{pmatrix} 6 & 33.2 \\ 33.2 & 209.74 \end{pmatrix}^{-1} \begin{pmatrix} 19.2 \\ 102.5 \end{pmatrix}$$

$$= \begin{pmatrix} 1.343 & -0.213 \\ -0.213 & 0.038 \end{pmatrix} \begin{pmatrix} 19.2 \\ 102.5 \end{pmatrix}$$

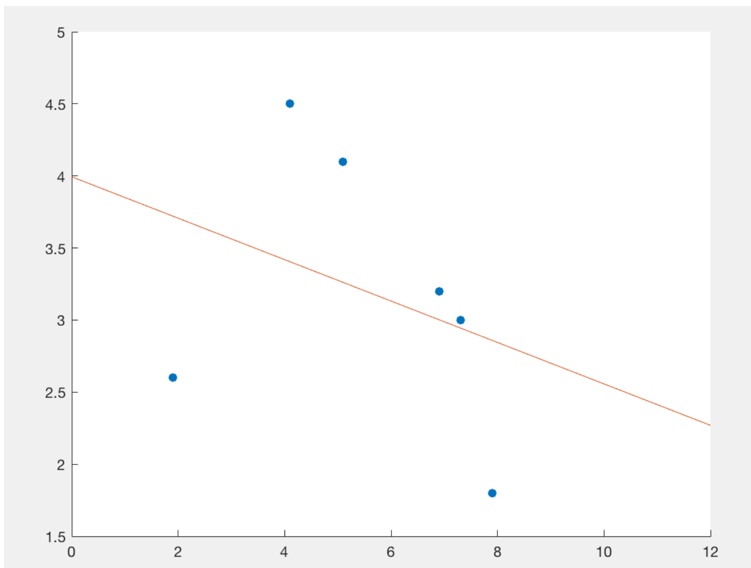
$$= \begin{pmatrix} 3.994 \\ -0.1437 \end{pmatrix}$$

$$\hat{\omega_0} = 3.994$$
$$\hat{\omega_1} = -0.1437$$

Dari perhitungan tersebut, kita telah mendapatkan nilai ω_0 dan ω_1. Maka model regersi linearnya adalah

$$f(x) = -0.1437x + 3.994$$

Jika digambarkan pada grafik, maka akan seperti ini



(b) (5 points) Predict the $y_4$that is the y value of p4 using univariate linear regression model in 15(a).
Diketahui nilai x adalah 6.0, maka

$$\begin{aligned} f(x) &= -0.1437(6.0) + 3.994 \\ &= 3.1318 \end{aligned}$$

Jadi nilai y_4 adalah 3.1318

(c) (20 points) Build a univariate non-linear regression model using data set 1 (except p4). (Hints: use m degree polynomial. It is up to you in selecting m value.)
Diketahui fungsi polynomial dengan derajat m = 7 adalah

$$y = \omega_0 + \omega_1 x + \omega_2 x^2 + \omega_3 x^3 + \omega_4 x^4 + \omega_5 x^5 + \omega_6 x^6 + \omega_7 x^7$$

Untuk menghitung masing-masing nilai a, kita dapat menggunakan kembali *normal equation* dengan menyesuaikan matrix inputan, menjadi

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} \\ \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} \\ \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} \\ \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} & \sum x_i^{13} \\ \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} & \sum x_i^{13} & \sum x_i^{14} \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \omega_7 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \\ \sum x_i^4 y_i \\ \sum x_i^5 y_i \\ \sum x_i^6 y_i \\ \sum x_i^7 y_i \end{pmatrix}$$

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} \\ \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} \\ \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} \\ \sum x_i^6 & \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} & \sum x_i^{13} \\ \sum x_i^7 & \sum x_i^8 & \sum x_i^9 & \sum x_i^{10} & \sum x_i^{11} & \sum x_i^{12} & \sum x_i^{13} & \sum x_i^{14} \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \\ \sum x_i^4 y_i \\ \sum x_i^5 y_i \\ \sum x_i^6 y_i \\ \sum x_i^7 y_i \end{pmatrix} = \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \omega_7 \end{pmatrix}$$

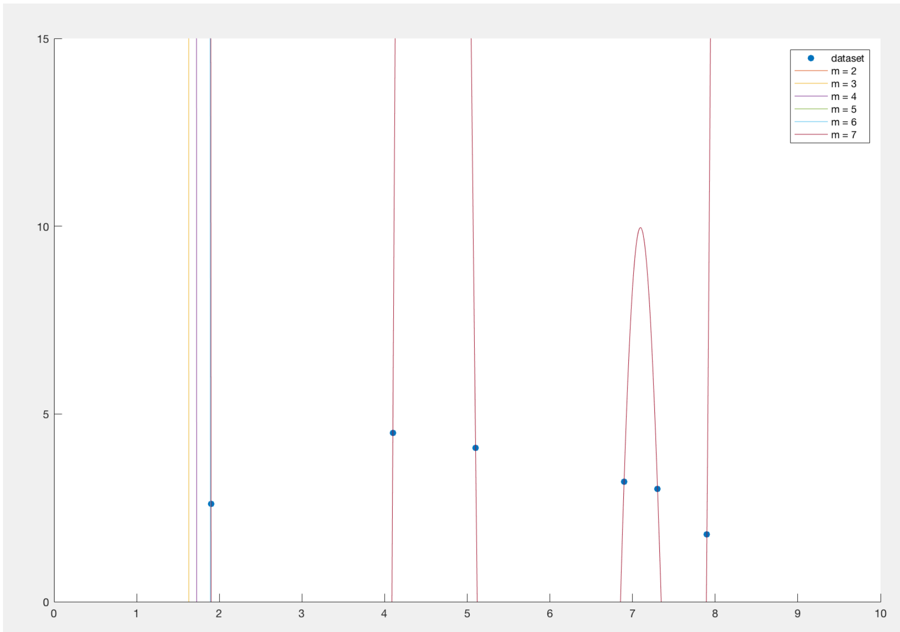
$$\begin{pmatrix} 6 & 33.20 & 209.74 & 1418.996 & 9973.6726 & 71775.16892 & 524733.28 & 3879071.96 \\ 33.20 & 209.74 & 1418.996 & 9973.67 & 71775.169 & 524733.28 & 3879071.96 & 28911371.296 \\ 209.7400 & 1418.996 & 9973.67 & 71775.169 & 524733.28 & 3879071.96 & 28911371.296 & 216837140.3 \\ 1418.996 & 9973.67 & 71775.169 & 524733.28 & 3879071.96 & 28911371.3 & 216837140.3 & 1634456718.7 \\ 9973.67 & 71775.169 & 524733.28 & 3879071.96 & 28911371.3 & 216837140.3 & 1634456718.72 & 12371294954.18 \\ 71775.17 & 524733.28 & 3879071.96 & 28911371.3 & 216837140.32 & 1634456718.72 & 12371294954.18 & 93972084048.93 \\ 524733.28 & 3879071.96 & 28911371.3 & 216837140.3 & 1634456718.72 & 12371294954.18 & 93972084048.93 & 716039188528.61 \\ 3879071.96 & 28911371.3 & 216837140.3 & 1634456718.72 & 12371294954.18 & 93972084048.93 & 716039188528.61 & 5471265871016.91 \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} 19.20 \\ 102.50 \\ 616.23 \\ 3977.60 \\ 26863.17 \\ 187052.11 \\ 1330540.79 \\ 9609566.42 \end{pmatrix} = \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \omega_7 \end{pmatrix}$$

Sehingga kita mendapatkan nilai-nilai ω yaitu

$$\omega_0 = 54727.64, \omega_1 = -68068.79, \omega_2 = 30076.49,$$
$$\omega_3 = -5467.20, \omega_4 = 143.43, \omega_5 = 87.41, \omega_6 = -11.59, \omega_7 = 0.45$$

Berikut adalah gambar grafik regresi menggunakan polynomial dengan derajat m = 2 sampai m = 7. Jika diperhatikan baik-baik, semakin tinggi nilai m, semakin akurat garis dengan titik-titik pada grafik.



(d) (5 points) Predict the y4 that is the y value of p4 using univariate non-linear regression model in 15(c).

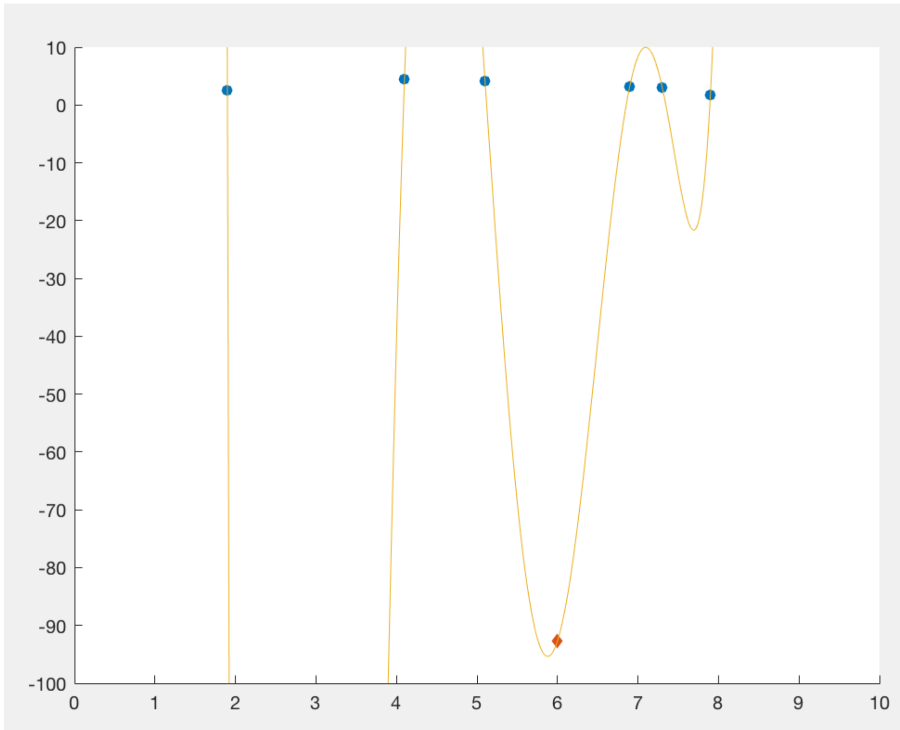
Menggunakan persamaan

$$y = \omega_0 + \omega_1 x + \omega_2 x^2 + \omega_3 x^3 + \omega_4 x^4 + \omega_5 x^5 + \omega_6 x^6 + \omega_7 x^7$$

dengan masing-masing nilai ω

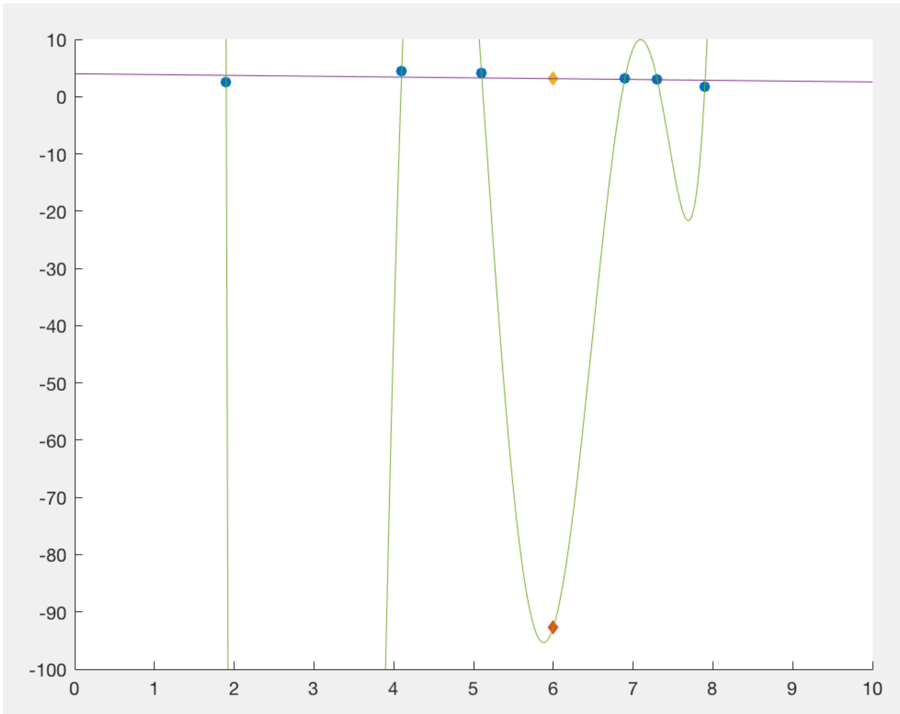
$$\omega_0 = 54727.64, \omega_1 = -68068.79, \omega_2 = 30076.49,$$
$$\omega_3 = -5467.20, \omega_4 = 143.43, \omega_5 = 87.41, \omega_6 = -11.59, \omega_7 = 0.45$$

didapatkan hasil y_4 = -92.6961.



(e) (5 points) Between two models resulted from 15(a) and 15(b), which model that gives better prediction to p4? Why? Give your explanation.

Gambar di bawah menampilkan prediksi dengan regresi linear dan regres non-linear menggunakan fungsi polinomial berderajat m = 7. Dari gambar, kita dapat melihat bahwa hasil prediksi regresi linear lebih dekat dengan *data set* atau *data latih*, dibandingkan dengan hasil prediksi regresi non-linear. Maka dari itu, menurut saya pada kasus di atas, prediksi menggunakan regresi linear lebih baik daripada regresi non-linear menggunakan fungsi polinomial berderajat m = 7.



2. Given data set 2 (xi, yi), where xi = (x1, x2), shown in Table 2.

   (a) (20 points) Build a multivariate linear regression model using data set 2 (except p4).

   Untuk membuat model linear, kita harus mencari nilai Ω terlebih dahulu. Nilai Ω dapat dicari menggunakan persamaan [2]

$$\Omega = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T y$$

   di mana Ω adalah

$$\Omega = (\omega_0, \omega_1)^T$$

   dan diketahui

$$\mathrm{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1d} \\ 1 & x_{21} & x_{22} & \ldots & x_{2d} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{N1} & x_{N2} & \ldots & x_{Nd} \end{pmatrix}$$

Dengan memasukkan data latih (x_1, x_2, dan y) yang ada di dataset, maka kita akan mendapatkan matriks

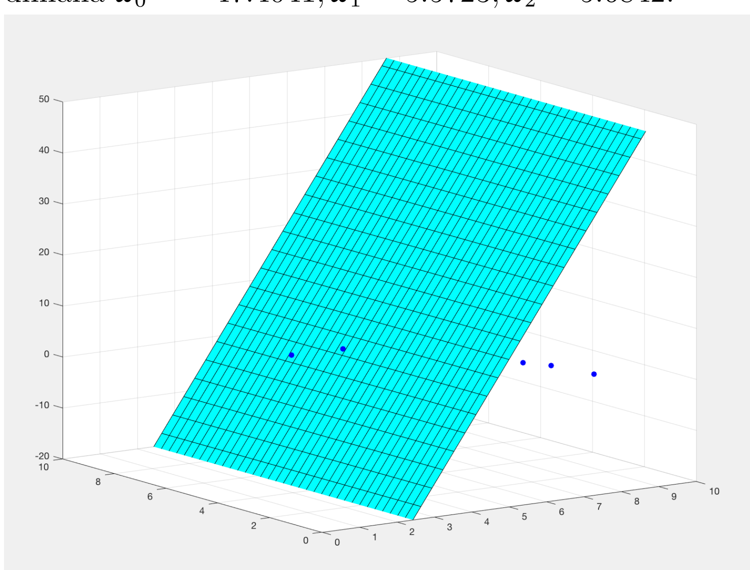
$$\mathrm{X} = \begin{pmatrix} 1 & 3.2 & 5.8 \\ 1 & 4.1 & 5.1 \\ 1 & 5.1 & 4.0 \\ 1 & 6.9 & 2.2 \\ 1 & 7.3 & 1.7 \\ 1 & 7.9 & 0.9 \end{pmatrix}, \mathrm{y} = \begin{pmatrix} 3.2 \\ 4.5 \\ 4.1 \\ 3.2 \\ 3 \\ 1.8 \end{pmatrix}$$

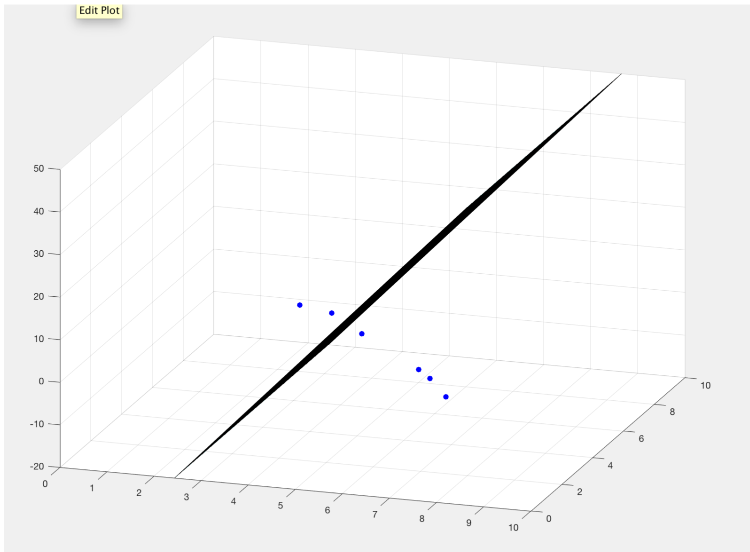
Menggunakan persamaan di atas,

$$\Omega = \left(\begin{pmatrix} 1 & 3.2 & 5.8 \\ 1 & 4.1 & 5.1 \\ 1 & 5.1 & 4.0 \\ 1 & 6.9 & 2.2 \\ 1 & 7.3 & 1.7 \\ 1 & 7.9 & 0.9 \end{pmatrix}^T \begin{pmatrix} 1 & 3.2 & 5.8 \\ 1 & 4.1 & 5.1 \\ 1 & 5.1 & 4.0 \\ 1 & 6.9 & 2.2 \\ 1 & 7.3 & 1.7 \\ 1 & 7.9 & 0.9 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 3.2 & 5.8 \\ 1 & 4.1 & 5.1 \\ 1 & 5.1 & 4.0 \\ 1 & 6.9 & 2.2 \\ 1 & 7.3 & 1.7 \\ 1 & 7.9 & 0.9 \end{pmatrix}^T \begin{pmatrix} 3.2 \\ 4.5 \\ 4.1 \\ 3.2 \\ 3 \\ 1.8 \end{pmatrix}$$

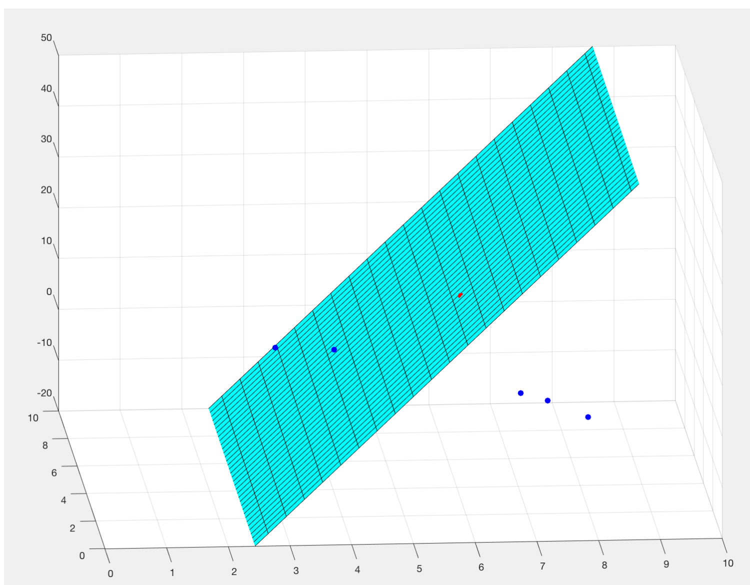
$$= \begin{pmatrix} -47.4041 \\ 5.5723 \\ 5.6842 \end{pmatrix}$$

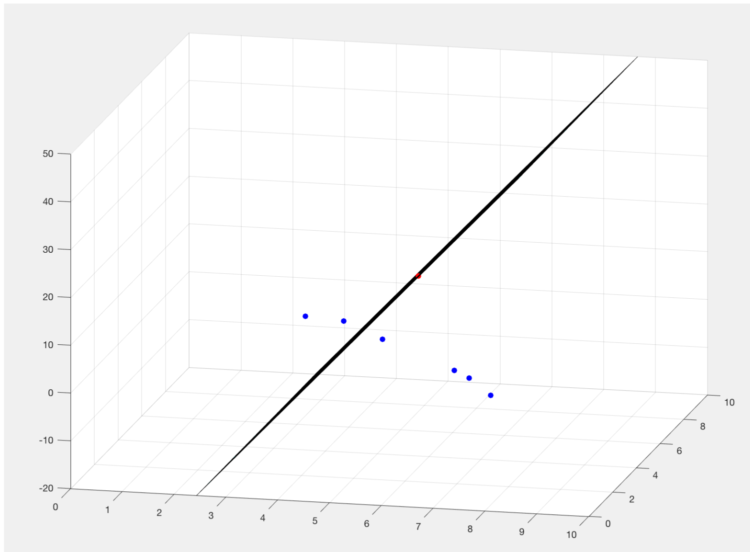
dimana $\hat{\omega_0} = -47.4041, \hat{\omega_1} = 5.5723, \hat{\omega_2} = 5.6842$.





(b) (5 points) Predict the y value of p4 using model in 16(a).
dengan menggunakan model di 16(a), nilai y_4 adalah 20.1353.





(c) (20 points) Build a multivariate non-linear regression model using data set 2 (except p4). (Hints: use some interaction terms as given in the slide of Regression page 17.)

Dengan menggunakan rumus pada [5]

$$y_i = f(\mathrm{x}_i) = \omega_0 + \omega_1 x_{i1} + \omega_1 x_{i2} + \omega_3 x_{i1} x_{i2}$$

dan menggunakan matriks masukan

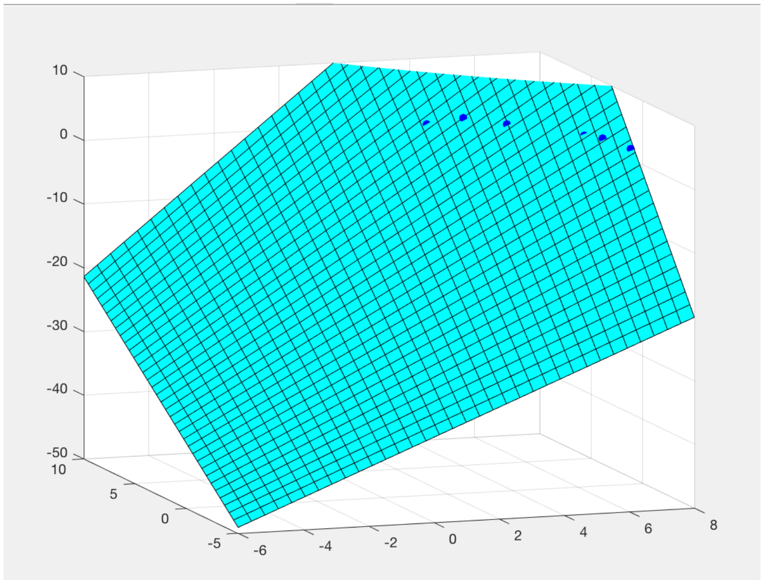
$$\mathrm{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ 1 & x_{31} & x_{32} & x_{31}x_{32} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \mathrm{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix}$$

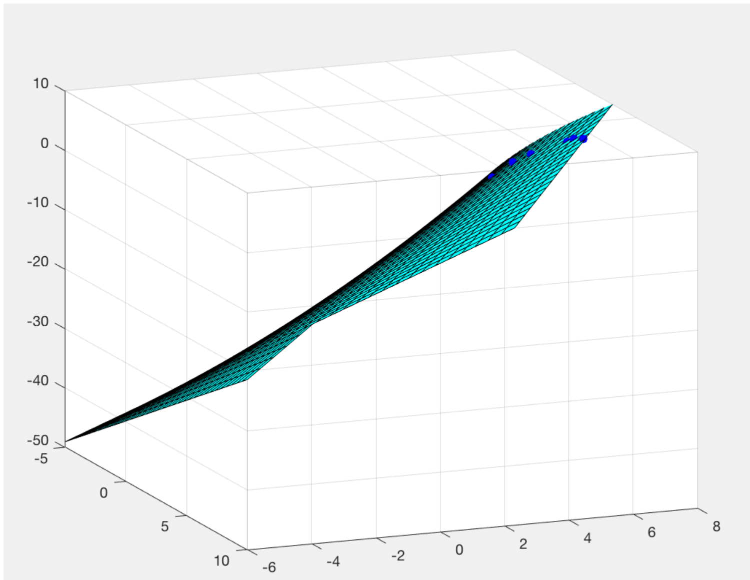
kita dapat menggunakan persamaan pada 16(a) untuk mencari nilai Ω

$$= \begin{pmatrix} -23.3572 \\ 2.7489 \\ 2.6568 \\ 0.1350 \end{pmatrix}$$

$$\Omega = \left(\begin{pmatrix} 1 & 3.2 & 5.8 & 18.56 \\ 1 & 4.1 & 5.1 & 20.91 \\ 1 & 5.1 & 4 & 20.4 \\ 1 & 6.9 & 2.2 & 15.18 \\ 1 & 7.3 & 1.7 & 12.41 \\ 1 & 7.9 & 0.9 & 7.11 \end{pmatrix}^T \begin{pmatrix} 1 & 3.2 & 5.8 & 18.56 \\ 1 & 4.1 & 5.1 & 20.91 \\ 1 & 5.1 & 4 & 20.4 \\ 1 & 6.9 & 2.2 & 15.18 \\ 1 & 7.3 & 1.7 & 12.41 \\ 1 & 7.9 & 0.9 & 7.11 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 3.2 & 5.8 & 18.56 \\ 1 & 4.1 & 5.1 & 20.91 \\ 1 & 5.1 & 4 & 20.4 \\ 1 & 6.9 & 2.2 & 15.18 \\ 1 & 7.3 & 1.7 & 12.41 \\ 1 & 7.9 & 0.9 & 7.11 \end{pmatrix}^T \begin{pmatrix} 3.2 \\ 4.5 \\ 4.1 \\ 3.2 \\ 3 \\ 1.8 \end{pmatrix} = \begin{pmatrix} -23.3572 \\ 2.7489 \\ 2.6568 \\ 0.1350 \end{pmatrix}$$

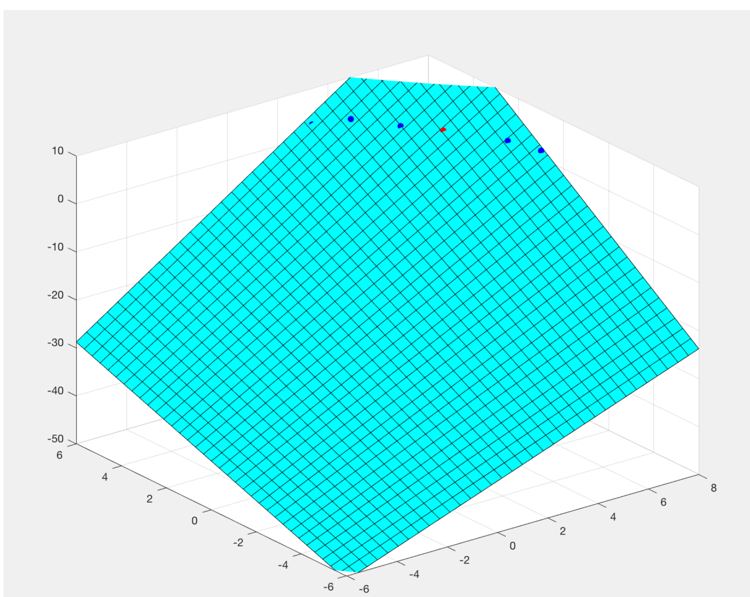
dimana $\hat{\omega_0} = -23.3572, \hat{\omega_1} = 2.7489, \hat{\omega_2} = 2.6568, \hat{\omega_3} = 0.1350.$

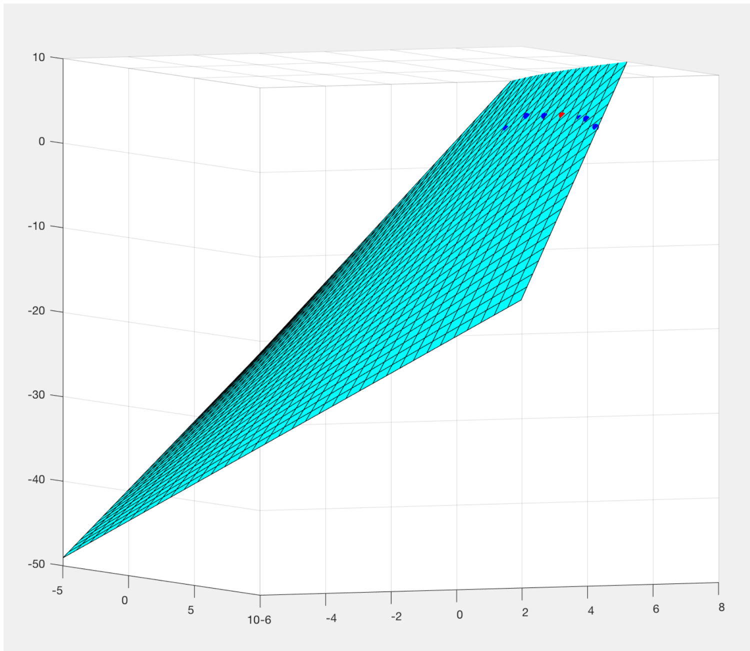




3. (5 points) Predict the y value of p4 using multivariate non-linear regression model in 16(c).

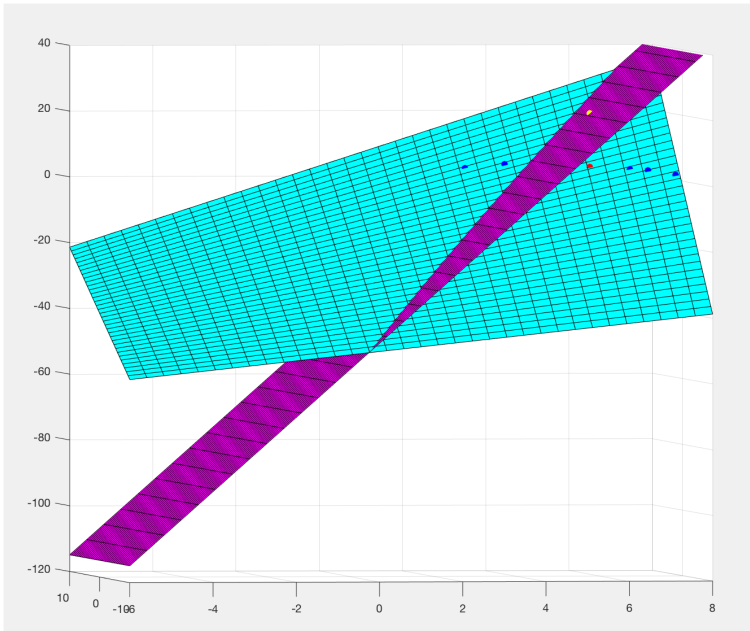
   Dengan menggunakan model di 16(c), nilai y_4 adalah 3.88328;

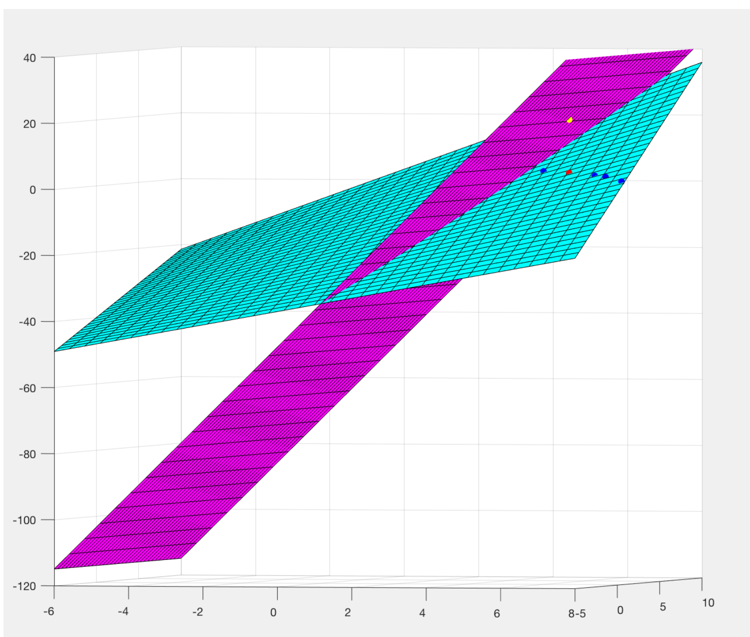




4. (5 points) Between two models resulted from 16(a) and 16(c), which model that gives better prediction to p4? Why? Give your explanation.

   Dari grafik, dapat dilihat bahwa model 16(c) mempunya nilai prediksi yang mendekati nilai-nilai dataset atau datatraining. Menurut saya, model 16(c) lebih baik daripada 16(a)





Referensi

[1] https://id.wikipedia.org/wiki/Regresi$_{L}inier$
[2] Introduction to Data Mining - Panning Tan, M. Steinbach

[3] https://en.wikipedia.org/wiki/Nonlinear$_{r}egression$

[4] Regression book

[5] Regression slide

[6] http://www.nickgillian.com/wiki/pmwiki.php/GRT/MLP

[7] Machine Learning - Tom Mitchell

[8] https://medium.com/towards-data-science/activation-functions-and-its-types-which-is-better-a9a5310cc8f

[9] Slide ANN-MLP Machine Learning