Numerikus matematika

Baran Ágnes

Előadás
Legkisebb négyzetek módszere

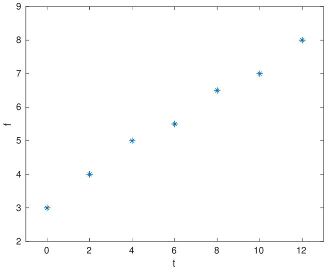
Legkisebb négyzetek módszere

Példa

Egy fél méter magas, téglatest alakú víztartályt egyenletes sebességgel töltenek fel vízzel. Amikor a tartályban 3 cm magasan áll a víz Péter elhatározza, hogy megméri a vízszint változását az idő függvényében. A következő méréseket végezte:

t_i (min)	0	2	4	6	8	10	12
f_i (cm)	3	4	5	5.5	6.5	7	8

Becsülje meg milyen magasan lesz a víz 20 perccel azután, hogy Péter elindította a mérést! Mikor indították el a tartály feltöltését? Kb mikor lesz tele a tartály?

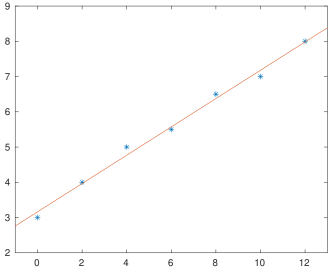


A feltöltés sebessége egyenletes $\implies$ a víz magassága az idő lineáris függvénye:

$$F(t) = x_1 + x_2 t,$$

ahol x_1 és x_2 értékét a mérések alapján határozzuk meg.

A méréseink esetlegesen hibával terheltek, így nem biztos, hogy a pontok egy egyenesre illeszkednek.

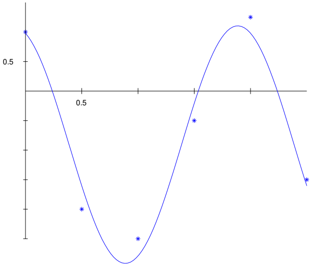


Példa

Megfigyelünk egy periodikus folyamatot, a méréseinkre egy

$$F(t) = x_1 + x_2 \cos \pi t + x_3 \sin \pi t$$

alakú modellt szeretnénk illeszteni, ahol x_1, x_2, x_3 értékét a mérések alapján határozzuk meg.



Mérési hibák miatt a modell nem biztos, hogy pontosan illeszkedik az adatokra.

Hogyan válasszuk meg a modell paramétereit, ha az adataink esetlegesen hibával terheltek?

t_i:	az i-edik megfigyelési hely
f_i:	az i-edik helyen megfigyelt érték
$F(t_i)$:	a modellünk értéke az i-edik helyen

Az i-edik helyen a modellünk értékének és a megfigyelt értéknek a négyzetes eltérése:

$$(F(t_i) - f_i)^2$$

Olyan paramétereket fogunk választani, melyre ezen négyzetes eltérések összege minimális:

$$\sum_i (F(t_i) - f_i)^2,$$

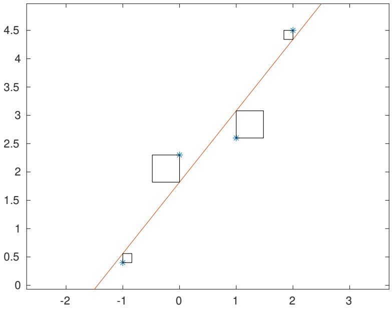
ahol az összegzést az összes megfigyelésre végezzük.

Olyan paramétereket fogunk választani, melyre ezen négyzetes eltérések összege minimális:

$$\sum_i (F(t_i) - f_i)^2,$$

ahol az összegzést az összes megfigyelésre végezzük.

Szemléletesen: négyzetek területösszegét minimalizáljuk



A modell

Olyan modellekkel foglalkozunk, melyek valamilyen adott $\varphi_1(t), \dots, \varphi_n(t)$ függvények lineáris kombinációi:

$$F(t) = x_1\varphi_1(t) + \cdots + x_n\varphi_n(t) = \sum_{j=1}^{n} x_j\varphi_j(t)$$

Példa

$$F(t) = x_1 \cdot \underbrace{1}_{\varphi_1(t)} + x_2 \underbrace{t}_{\varphi_2(t)}$$

$$F(t) = x_1 \cdot \underbrace{1}_{\varphi_1(t)} + x_2 \underbrace{\cos \pi t}_{\varphi_2(t)} + x_3 \underbrace{\sin \pi t}_{\varphi_3(t)}$$

$$F(t) = x_1 \underbrace{\sin t}_{\varphi_1(t)} + x_2 \underbrace{\sin 2t}_{\varphi_2(t)} + x_3 \underbrace{\sin 3t}_{\varphi_3(t)}$$

Legkisebb négyzetes közelítések

Adott m mérés:
a $t_1, t_2, \dots, t_m$ helyeken az
a $f_1, f_2, \dots, f_m$ megfigyelések.

A folyamatot leíró

$$F(t) = \sum_{j=1}^{n} x_j \varphi_j(t)$$

modell n darab ismeretlen paraméterét keressük úgy, hogy

$$J(x) = \sum_{i=1}^{m} (F(t_i) - f_i)^2$$

minimális legyen. Tipikusan $m \gg n$.

x_j: ismeretlen paraméterek ($j = 1, \dots, n$)
$\varphi_j(t)$: adott függvények ($j = 1, \dots, n$)

Legyen

$$A = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \dots & \varphi_n(t_1) \\ \varphi_1(t_2) & \varphi_2(t_2) & \dots & \varphi_n(t_2) \\ \vdots & & & \\ \varphi_1(t_m) & \varphi_2(t_m) & \dots & \varphi_n(t_m) \end{bmatrix} \in \mathbb{R}^{m \times n},$$

$$f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \in \mathbb{R}^m, \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n,$$

Ekkor

$$Ax = \begin{bmatrix} F(t_1) \\ F(t_2) \\ \vdots \\ F(t_m) \end{bmatrix}$$

A minimalizálandó függvény:

$$J(x) = \sum_{i=1}^{m} (F(t_i) - f_i)^2 = \|Ax - f\|_2^2$$

Minimum csak ott lehet, ahol

$$\frac{\partial J(x)}{\partial x_k} = 0, \quad k = 1, \ldots, n.$$

Ez az alábbi lineáris egyenletrendszerre vezet:

$$A^T Ax = A^T f$$

(Gauss-féle normálegyenlet)

Gauss-féle normálegyenlet

$$A^T A x = A^T f$$

- a Gauss-féle normálegyenlet mindig megoldható
- A megoldás a legkisebb négyzetes értelemben legjobban közelítő modell paramétereit adja.
- Ha az A mátrix oszlopvektorai lineárisan függetlenek, akkor a Gauss-féle normálegyenletnek egyetlen megoldása van.

Ha az A oszlopvektorai függőek (az $A^T A$ mátrix szinguláris), akkor végtelen sok megoldás van.

Szingularitás esetén javasolható:

- több adat felvétele
- a modell egyszerűsítése

Példa

Ha az illesztett függvény egy egyenes: $F(t) = x_1 + x_2 t$, akkor $\varphi_1(t) \equiv 1$ és $\varphi_2(t) = t$

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{bmatrix}$$

$$A^T A = \begin{bmatrix} m & \sum_{i=1}^{m} t_i \\ \sum_{i=1}^{m} t_i & \sum_{i=1}^{m} t_i^2 \end{bmatrix}, \qquad A^T f = \begin{bmatrix} \sum_{i=1}^{m} f_i \\ \sum_{i=1}^{m} t_i f_i \end{bmatrix}$$

szingularitás: az A oszlopvektorai lineárisan függőek, azaz

$$t_1 = t_2 = \cdots = t_m$$

1. Ha van legalább két különböző t_i érték, akkor a rendszer egyértelműen megoldható. A megoldás a J minimumhelye lesz.

2. Ha $t_1 = t_2 = \cdots = t_m =: t_0$, akkor

$$\begin{bmatrix} m & mt_0 \\ mt_0 & mt_0^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{m} f_i \\ t_0 \sum\limits_{i=1}^{m} f_i \end{bmatrix}$$

a 2. egyenlet az első t_0-szorosa $\quad \rightarrow \quad$ végtelen sok megoldás.

$$b = s \in \mathbb{R}, \quad a = \frac{1}{m} \sum_{i=1}^{m} f_i - st_0$$

Ha $b = 0$

$$a = \frac{1}{m} \sum_{i=1}^{m} f_i$$

Példa

Határozzuk meg az alábbi adatokat négyzetesen legjobban közelítő egyenes egyenletét!

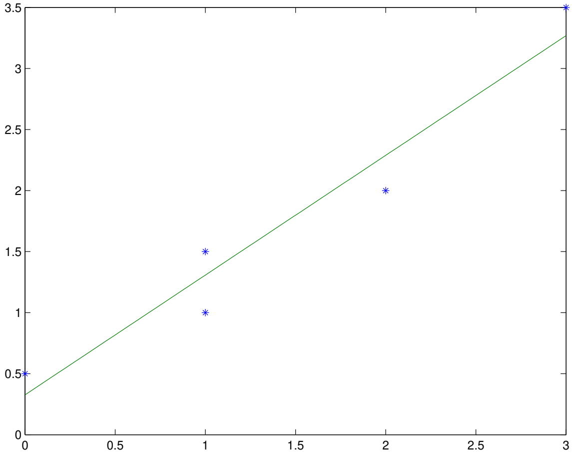
t_i	0	1	1	2	3
f_i	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{7}{2}$

A modell: $F(t) = a + b \cdot t$

$$\begin{bmatrix} 5 & 7 \\ 7 & 15 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{17}{2} \\ 17 \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 15 & -7 \\ -7 & 5 \end{bmatrix} \begin{bmatrix} \frac{17}{2} \\ 17 \end{bmatrix} = \begin{bmatrix} \frac{17}{52} \\ \frac{51}{52} \end{bmatrix}$$

Az illesztett modell: $F(t) = \frac{17}{52} + \frac{51}{52} t$


3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3

Példa

Határozzuk meg az alábbi adatokat négyzetesen legjobban közelítő egyenes egyenletét!

t_i	2	2	2	2	2
f_i	1	1	2	2	2

A modell: $F(t) = a + b \cdot t$

$$\begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \end{bmatrix}$$

$$5a + 10b = 8$$

$$b = s \in \mathbb{R}, \qquad a = \frac{8}{5} - 2s$$

Ha $s = 0$, akkor $F(t) \equiv \frac{8}{5}$

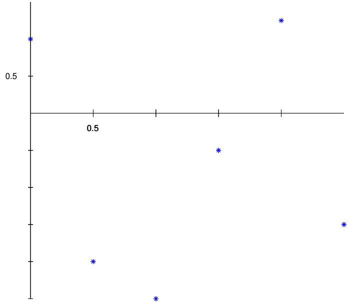
Példa

Határozzuk meg az alábbi adatokat négyzetesen legjobban közelítő

$$F(t) = x_1 + x_2 \cos(\pi t) + x_3 \sin(\pi t)$$

alakú modellt!

t_i	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$
f_i	1	-2	$-\frac{5}{2}$	$-\frac{1}{2}$	$\frac{5}{4}$	$-\frac{3}{2}$



$\varphi_1(t) \equiv 1$, $\varphi_2(t) = \cos(\pi t)$, $\varphi_3(t) = \sin(\pi t)$

$$A = \begin{bmatrix} 1 & \cos(\pi t_1) & \sin(\pi t_1) \\ 1 & \cos(\pi t_2) & \sin(\pi t_2) \\ \vdots & & \\ 1 & \cos(\pi t_6) & \sin(\pi t_6) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 6 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 3 \end{bmatrix}, \qquad A^T f = \begin{bmatrix} -\frac{17}{4} \\ \frac{19}{4} \\ -3 \end{bmatrix}$$

Az $A^T A x = A^T f$ Gauss-féle normálegyenlet megoldása:

$$x = \begin{bmatrix} -\frac{29}{32} \\ \frac{181}{96} \\ -\frac{67}{96} \end{bmatrix}$$

Példa

Határozzuk meg az alábbi adatokat négyzetesen legjobban közelítő

$$F(t) = x_1 + x_2 \cos(\pi t) + x_3 \sin(\pi t)$$

alakú modellt!

t_i	0	$\frac{1}{2}$	2	$\frac{5}{2}$
f_i	1	-2	$\frac{5}{4}$	$-\frac{3}{2}$

Az előző példából:

t_i	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$
f_i	1	-2	$-\frac{5}{2}$	$-\frac{1}{2}$	$\frac{5}{4}$	$-\frac{3}{2}$

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \to A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 \\ -2 \\ -\frac{5}{2} \\ -\frac{1}{2} \\ \frac{5}{4} \\ -\frac{3}{2} \end{bmatrix} \to f = \begin{bmatrix} 1 \\ -2 \\ \frac{5}{4} \\ -\frac{3}{2} \end{bmatrix}$$

Az A oszlopai lineárisan függőek $\to A^T A$ szinguláris

A szingularitás kezelése:

1. több adat felvétele (ld. előző példa)
2. a modell egyszerűsítése:

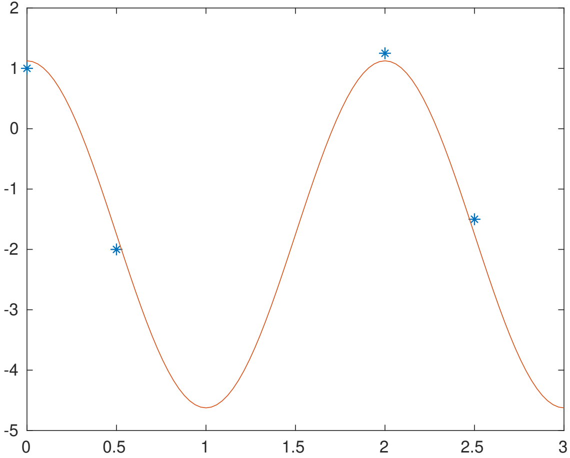
$$F(t) = x_1 + x_2 \cos(\pi t)$$

Ekkor

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}, \qquad f = \begin{bmatrix} 1 \\ -2 \\ \frac{5}{4} \\ -\frac{3}{2} \end{bmatrix}$$

Az $A^T A x = A^T f$ Gauss-féle normálegyenlet megoldása:

$$x = \begin{bmatrix} -1.7500 \\ 2.8750 \end{bmatrix}$$


2
1
0
-1
-2
-3
-4
-5
0
0.5
1
1.5
2
2.5
3

Példa

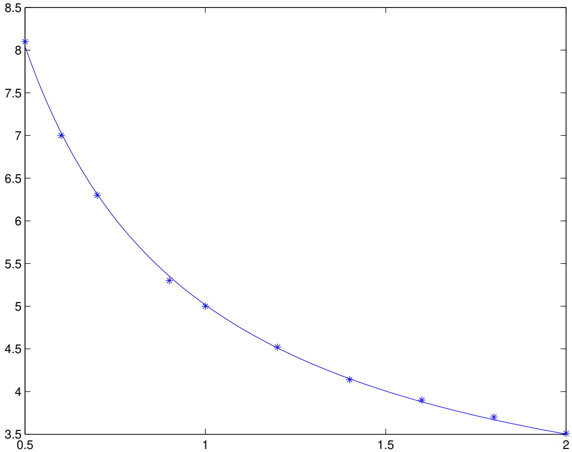
Határozzuk meg az alábbi adatokat négyzetesen legjobban közelítő $F(t) = a + \frac{b}{t}$ alakú modell paramétereit!

t_i	0.5	0.6	0.7	0.9	1	1.2	1.4	1.6	1.8	2
f_i	8.1	7	6.3	5.3	5	4.52	4.14	3.9	3.7	3.51

$$m = 10, \qquad A = \begin{bmatrix} 1 & \frac{1}{t_1} \\ 1 & \frac{1}{t_2} \\ \vdots & \\ 1 & \frac{1}{t_{10}} \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 10 & \sum_{i=1}^{10} \frac{1}{t_i} \\ \sum_{i=1}^{10} \frac{1}{t_i} & \sum_{i=1}^{10} \frac{1}{t_i^2} \end{bmatrix}, \qquad A^T f = \begin{bmatrix} \sum_{i=1}^{10} f_i \\ \sum_{i=1}^{10} \frac{f_i}{t_i} \end{bmatrix}$$

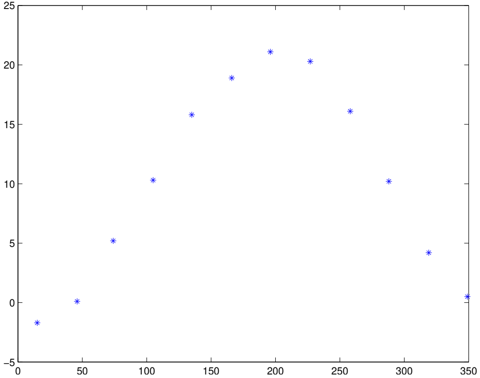
Megj.: Az $\left(\frac{1}{t_i}, f_i\right)$ adatokra illesztettünk egyenest.


8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
0.5
1
1.5
2

Példa

Havi középhőmérsékletek átlagai Budapesten (1901-1950)

t_i	15	46	74	105	135	166	196	227	258	288	319	349
f_i	-1.7	0.1	5.2	10.3	15.8	18.9	21.1	20.3	16.1	10.2	4.2	0.5



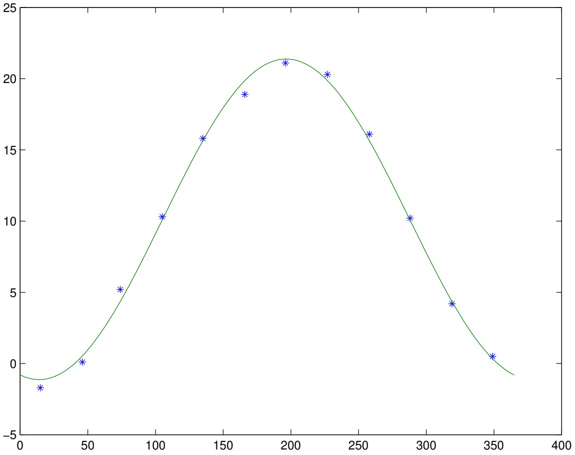
A modell:

$$F(t) = x_1 + x_2 \cos\left(2\pi \frac{t-14}{365}\right)$$

$$A = \begin{bmatrix} 1 & \cos\left(2\pi \frac{t_1-14}{365}\right) \\ 1 & \cos\left(2\pi \frac{t_2-14}{365}\right) \\ \vdots & \\ 1 & \cos\left(2\pi \frac{t_{12}-14}{365}\right) \end{bmatrix}$$

Az $A^T A x = A^T f$ Gauss-féle normálegyenlet megoldása (4 tizedesjegyre kerekítve):

$$x = \begin{bmatrix} 10.1248 \\ -11.2577 \end{bmatrix}$$


25
20
15
10
5
0
−5
0
50
100
150
200
250
300
350
400

Példa

(Matlab, carsmall adathalmaz) 93 autó esetén adott a lóerő, a súly és a gyorsulás értéke. Ezekből az adatokból szeretnénk megbecsülni, hogy az autó 1 gallon üzemanyaggal hány mérföldet tud megtenni (MPG). Feltételezzük, hogy az MPG érték a felsorolt jellemzők lineáris függvénye. Írjuk le ezt a kapcsolatot!

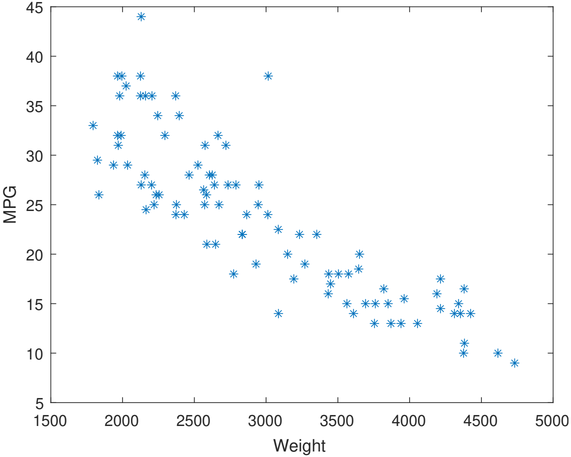
Legyen
$\varphi_1(t)$ a t autó esetén a lóerő
$\varphi_2(t)$ a t autó súlya
$\varphi_3(t)$ a t autó esetén a gyorsulás
$F(t)$ a t autó esetén a MPG érték

Kezdjük egy egyszerű modellel:

$$F(t) \approx x_1 + x_2\varphi_2(t),$$

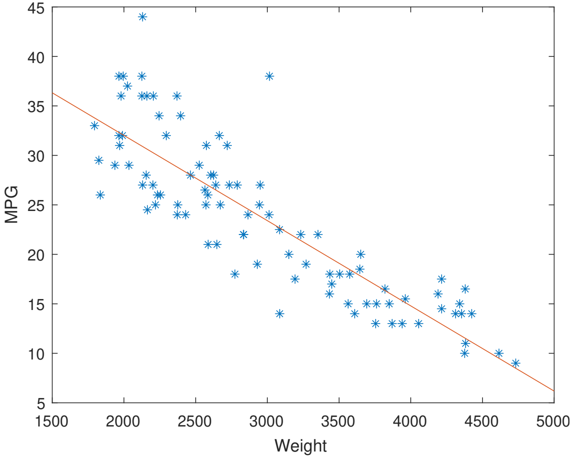
azaz az MPG értéket csak a súly függvényében vizsgáljuk.

Ábrázoljuk a (súly, MPG) párokat!



Látjuk, hogy a két érték között negatív kapcsolat van (minél nagyobb a súly, annál kevesebb mérföldet tud megtenni 1 gallon benzinnel).

Illesszünk egyenest a (súly, MPG) adatokra!



Az illesztett egyenes paraméterei: $x_1 = 49.2383$, $x_2 = -0.0086$.
A négyzetes eltérések összege: 1572.6

Próbálkozzunk egy bonyolultabb modellel:

$$F(t) \approx x_1 + x_2\varphi_1(t) + x_3\varphi_2(t),$$

azaz a lóerő és a súly segítségével becsüljük az MPG értéket.

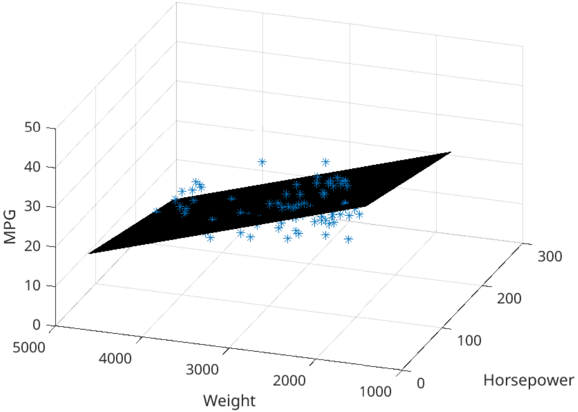
Az $A^TAx = A^Tf$ Gauss-féle normálegyenletet kell megoldanunk, ahol az A mátrixnak most 3 oszlopa van:

1. oszlop: az azonosan 1 vektor
2. oszlop: a lóerő értékek vektora
3. oszlop: a súly értékek vektora

Az f oszlopvektor az MPG értékek vektora

Az egyenlet megoldása után:

$$x_1 = 47.769, \quad x_2 = -0.042018, \quad x_3 = -0.0065651$$



A négyzetes eltérések összege: 1488.9

Ha mindhárom jellemzőt (lóerő, súly, gyorsulás) figyelembe vesszük a becslésnél:

$$F(t) \approx x_1 + x_2\varphi_1(t) + x_3\varphi_2(t) + x_4\varphi_3(t),$$

akkor az A mátrix 4 oszlopból áll:

1. oszlop: az azonosan 1 vektor
2. oszlop: a lóerő értékek vektora
3. oszlop: a súly értékek vektora
4. oszlop: a gyorsulás értékek vektora.

Az egyenlet megoldása után:

$$x_1 = 47.9768, \quad x_2 = -0.0429, \quad x_3 = -0.0065, \quad x_4 = -0.0116,$$

A négyzetes eltérések összege: 1488.8
(A javulás az előző modellhez képest minimális.)

Megjegyzés: a négyzetes hiba helyett gyakran az átlagos négyzetes hibát használjuk (így a hiba nem függ az adatok számától):

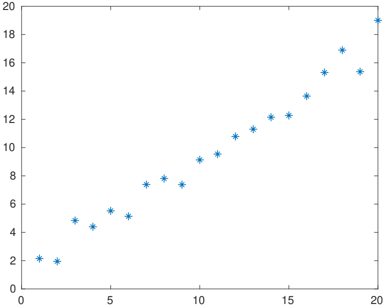
$$MSE = \frac{1}{m}\sum_{i=1}^{m}(F(t_i) - f_i)^2,$$

vagy ennek a négyzetgyökét (így a hibát és a megfigyeléseket ugyanazon a skálán mérjük):

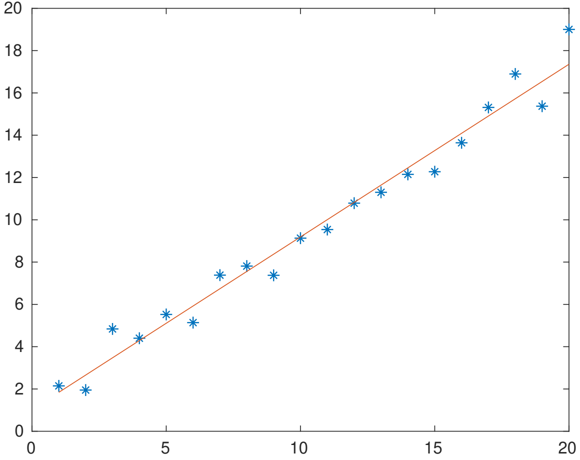
$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(F(t_i) - f_i)^2}.$$

Példa

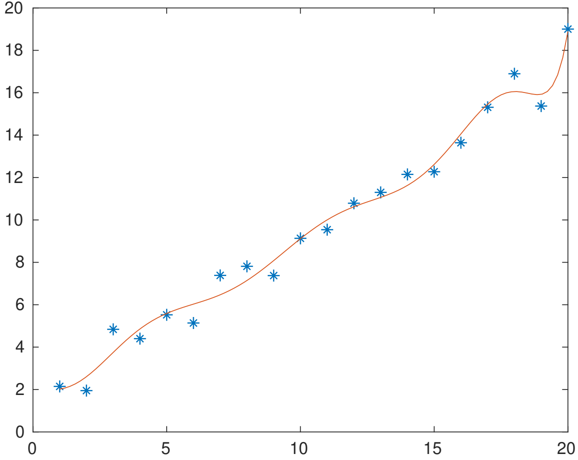
Tegyük fel, hogy megfigyelünk egy folyamatot, amely az $F(t) = 1.1 + 0.8t$ modellel írható le. ,,Felejtsük el'' a modellt, és végezzünk méréseket a $t = 1, \ldots, 20$ helyeken. A méréseink hibával terheltek, így az ábrán látható megfigyeléseket végeztük. Vizsgáljuk meg mi történik, ha modellt illesztünk a megfigyeléseinkre, de tegyük fel, hogy nincs elképzelésünk az illesztendő modellről, így a négyzetes hiba minimalizálása érdekében különböző fokszámú polinomokkal próbálkozunk. Minden esetben számoljuk ki az átlagos négyzetes hibát.



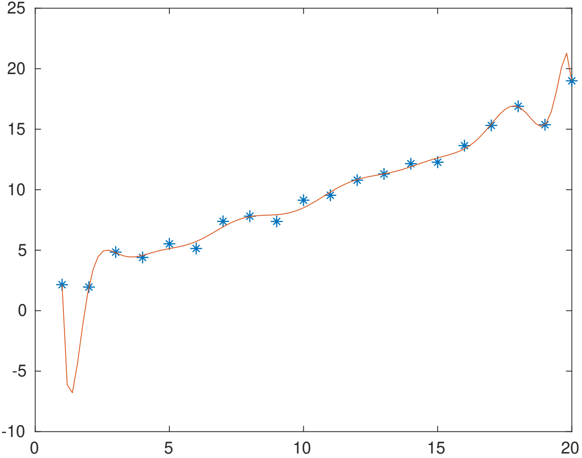
Ha egyenest illesztünk, akkor a megfigyelési helyeken az átlagos négyzetes hiba: 0.5992.



Ha egy kilencedfokú polinomot illesztünk, akkor a megfigyelési helyeken az átlagos négyzetes hiba: 0.3166.



Ha egy tizenötödfokú polinomot illesztünk, akkor a megfigyelési helyeken az átlagos négyzetes hiba: 0.0890.



Melyik a legjobb modell???

A megfigyelt értékekre a harmadik illeszkedik a legjobban, de a megfigyelt folyamatot mégsem jól írja le: rossz az általánosító képessége.

Vizsgáljuk meg az illesztett modellek értékét $t_a = 1.5$-ben és $t_b = 19.2$-ben, és hasonlítsuk össze az ,,elméleti értékkel" (amiből az adatokat generáltuk).

A polinom fokszáma	MSE (a megfigyelési helyeken)	az abszolút eltérés 1.5-ben	az abszolút eltérés 19.2-ben
1	0.5992	0.0509	0.2431
9	0.3166	0.0860	0.4311
15	0.0890	7.79	0.2469

Ha kellően sok adat áll rendelkezésre és kérdés, hogy milyen modellt válasszunk, akkor érdemes az adatainkat két részre bontani, tanuló- és tesztadatokra. A tanulóadatokra illesztjük a modellt, de az átlagos négyzetes hibát a tesztadatokon is mérjük, ez mutatja a modell általánosító képességét.