Numerikus Matematika

Baran Ágnes

Gyakorlat
Legkisebb négyzetes közelítések 2.

Legkisebb négyzetes közelítések

Példa

Határozzuk meg az alábbi adatokat legkisebb négyzetes értelemben legjobban közelítő

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

alakú modell paramétereit!

t_i	0.1	0.5	1.2	1.5	2	2.1	2.4	3	3.2
f_i	3.9	2.6	-0.8	0.3	3.2	3.8	3.2	-0.7	-0.9

Megoldás. A paramétereket az

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

Gauss-féle normálegyenlet megoldása szolgáltatja.

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

ahol

$$A = \begin{bmatrix} 1 & \cos(\pi t_1) & \sin(\pi t_1) \\ 1 & \cos(\pi t_2) & \sin(\pi t_2) \\ \vdots & & \\ 1 & \cos(\pi t_9) & \sin(\pi t_9) \end{bmatrix}, \quad f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_9 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Állítsuk elő a megadott adatokból az A mátrixot:

```
>> t=[0.1 0.5 1.2 1.5 2 2.1 2.4 3 3.2]';
>> f=[3.9 2.6 -0.8 0.3 3.2 3.8 3.2 -0.7 -0.9]';
>> A=[ones(9,1), cos(pi*t), sin(pi*t)];
```

Figyeljünk az A mátrix előállításánál: oszlopokat kell egymás mellé tennünk!

- Oldjuk meg a normálegyenletet!

Ha az f is oszlopvektorként adott:

```
>> x=(A'*A)\(A'*f)
x =
    1.4372
    2.0310
    1.1711
```

A legjobban illeszkedő adott alakú modell tehát:

$$F(t) = \underbrace{1.4372}_{x(1)} + \underbrace{2.0310}_{x(2)} \cos(\pi t) + \underbrace{1.1711}_{x(3)} \sin(\pi t)$$

Ha f-et nem oszlopvektorként adtuk meg, akkor a normálegyenlet megoldásánál hibaüzenetet kapunk.

- Ábrázoljuk az adatokat és az illesztett modellt!
  - Vegyünk fel sok (pl 100) pontot egy olyan intervallumban, ami tartalmazza a mérési helyeinket. Most pl. a $[0, 3.3]$ intervallum megfelelő:

    ```
    >> xx=linspace(0,3.3);
    ```

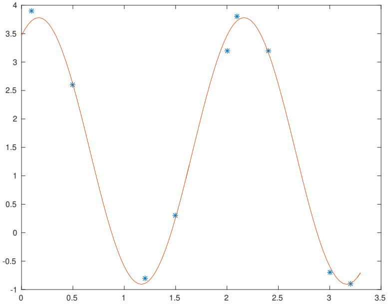
  - Az `xx` vektor elemeiben számoljuk ki a modell értékét. (A modell paraméterei az `x` vektorban vannak!)

    ```
    >> yy=x(1)+x(2)*cos(pi*xx)+x(3)*sin(pi*xx);
    ```

  - Ábrázoljuk az eredeti adatokat és az illesztett függvényt.

    ```
    >> plot(t,f,'*',xx,yy)
    ```

```
t=[0.1 0.5 1.2 1.5 2 2.1 2.4 3 3.2]';
f=[3.9 2.6 -0.8 0.3 3.2 3.8 3.2 -0.7 -0.9]';
A=[ones(9,1), cos(pi*t), sin(pi*t)];
x=(A'*A)\(A'*f);
xx=linspace(0,3.3);
yy=x(1)+x(2)*cos(pi*xx)+x(3)*sin(pi*xx);
figure; plot(t,f,'*',xx,yy)
```



Feladatok

(1) Határozza meg az alábbi adatokat legjobban közelítő

$$F(t) = a + \frac{b}{t}$$

alakú modell paramétereit!

t_i	1	1.2	1.4	1.4	1.5	1.7	1.9	2	2.1	2.2
f_i	4.2	3.8	3.4	3.3	3.3	3	2.8	2.8	2.75	2.7

(2) Határozza meg az alábbi adatokat négyzetesen legjobban közelítő

$$F(t) = x_1 \sin(t) + x_2 \sin(2t) + x_3 \sin(3t)$$

alakú modell paramétereit!

t_i	0.1	0.5	1.2	1.5	2	2.1	2.4	3	3.2	3.4	3.8	4	4.2	4.6	5
f_i	1	4.1	3	1	-1.5	-1.6	-1.7	-0.4	0.1	0.7	1.6	1.8	1.6	0.2	-2.5

3. feladat

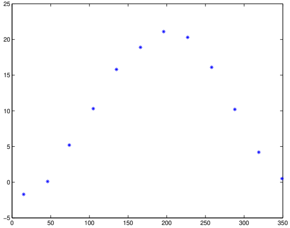
Havi középhőmérsékletek átlagai Budapesten (1901-1950)

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

Illesszünk az adatokra

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

alakú modellt.



4. feladat

Olvassa be a `trees.xlsx` állományt (Forrás: R programcsomag). Ez 31 fa (kései meggy) alábbi adatait tartalmazza: a törzs átmérője inch-ben (kissé félrevezetően: Girth oszlop), a fa magassága láb-ban (Height oszlop), a faanyag térfogata köbláb-ban (Volume oszlop). Ábrázolja a faanyag mennyiségét az átmérő függvényében, illetve a faanyag mennyiségét a magasság függvényében. A `plot3` függvény segítségével ábrázolja a faanyag mennyiségét az átmérő és a magasság függvényében. Illesszen egyenest a (Girth,Volume) adatokra, azaz becsülje meg a

$$\text{Volume} \sim x_1 + x_2 \cdot \text{Girth}$$

kapcsolatban az x_1, x_2 paramétereket! Végezze el a

$$\text{Volume} \sim x_1 + x_2 \cdot \text{Girth} + x_3 \cdot \text{Height}$$

közelítést is. Mindkét esetben számítsa ki az átlagos négyzetes hibát.

Példa

- Töltsük be a `carsmall` adathalmazt:

  ```
  >> load carsmall.mat
  ```

  Az adatállomány 100 autó adatait tartalmazza, most a `Horsepower`, `Weight`, `Acceleration`, `MPG` változókat fogjuk használni (a Wokspace-ben látjuk, hogy ezek 100 elemű oszlopvektorok). Az első három változóból próbáljuk megbecsülni a negyedik értéket.

- Ellenőrizzük, hogy az egyes vektorokban vannak-e hiányzó értékek!

  Használjuk az `isnan` függvényt, pl.

  ```
  >> sum(isnan(MPG))
     ans =
           6
  ```

  Csak azoknak az autóknak az adatait kellene megtartanunk, ahol mind a négy érték ismert. (Használhatjuk az `any` függvényt.)

- Ha töröljük azokat a sorokat, ahol van hiányzó érték akkor egy 93 elemű adathalmazunk maradt.
- Ábrázoljuk az MPG értékeket a súly függvényében!
- Keressük az MPG és Weight értékek közötti kapcsolatot lineáris függvény formájában:

$$\mathrm{MPG} \sim x_1 + x_2 \cdot \mathrm{Weight}$$

Ezt megoldhatjuk a `polyfit` függvénnyel, a Gauss-féle normálegyenlet megoldásával és a `fitlm` függvénnyel is.

- a polyfit függvénnyel: a (Weight,MPG) adatokra illesztünk egyenest.

  Figyeljünk, ne az eredeti Weight, MPG vektorokat használjuk, abban hiányzó értékek vannak!
  Emlékezzünk: a `polyfit` az illesztett egyenes együtthatóit a főegyütthatóval kezdve sorolja fel.

  Számítsuk ki az átlagos négyzetes hibát (a megfelelő képlet alkalmazásával, vagy használhatjuk a `immse` függvényt is)
- Oldjuk meg a feladatot a Gauss-féle normálegyenlettel is.

- a `fitlm` függvénnyel:

```
ans =
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate        SE          tStat 
                   __________    __________    _______

    (Intercept)        49.238        1.6504     29.834
    x1             -0.0086118    0.00053775    -16.014

Number of observations: 93, Error degrees of freedom: 91
Root Mean Squared Error: 4.16
R-squared: 0.738,  Adjusted R-Squared: 0.735
F-statistic vs. constant model: 256, p-value = 3.24e-28
```

Itt az együtthatókon kívül más információkat is kaptunk, ezekről ld. Statisztikából.

Ezután a Weight változón kívül használjunk még egy magyarázóváltozót, a Horsepower-t:

$$\text{MPG} \sim x_1 + x_2 \cdot \text{Horsepower} + x_3 \cdot \text{Weight}$$

Ebben az esetben is határozzuk meg az átlagos négyzetes hibát.