Laborator 8 - Regresia liniară şi Regresia polinomială

1 Regresia liniară

Observaţie.

→Dreapta de regresie liniară este de forma $y = ax + b$

→Determinăm coeficienţii a şi b din relaţiile

$$\begin{aligned} nb + a\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i, \end{aligned}$$

unde n reprezintă numărul punctelor tabelate, iar $x_i, y_i, i = \overline{1,n}$ reprezintă punctele date.

→Scris sub formă matriceală, avem

$$\begin{bmatrix} n & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{bmatrix}.$$

→Dreapta (linia) de regresie se poate determina şi cu ajutorul comenzii `p=polyfit(x,y,n)`urmată de apelul `polyval(p,x)`, care evaluează un polinom în valorile precizate ale variabilei

Problema 1. Se consideră setul de date

$$\begin{aligned} x &= [0.1, 0.4, 0.5, 0.7, 0.7, 0.9] \\ y &= [0.61, 0.92, 0.99, 1.52, 1.47, 2.03]\,. \end{aligned}$$

Să se determine dreapta de regresie liniară şi să se reprezinte grafic rezultatul aproximării în sensul celor mai mici pătrate.

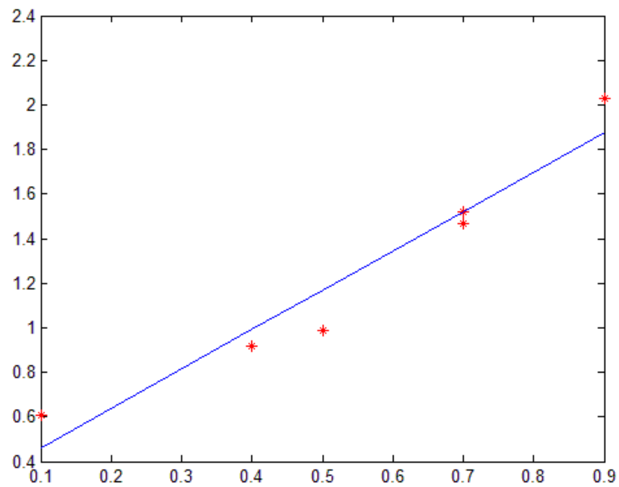
Soluţie.

```
x=[0.1,0.4,0.5,0.7,0.7,0.9];
```

```
y=[0.61,0.92,0.99,1.52,1.47,2.03];
c=polyfit(x,y,1)
%nu punem ";" la final pentru a vedea coeficientii dreptei de regresie
d=polyval(c,x);
axis([-2,10,-50,150])
plot(x,y,'*r');
hold on
plot(x,d)
% reprezentarea grafica se poate face si cu o singura comanda:  plot(x,d,x,y,'*r')
```

- Obţinem:

```
c =
 1.7646 0.2862
```



2 Regresia polinomială

Observaţie.

→Ecuaţia generală a estimatorului pătratic este de forma $y = a_0 + a_1 x + a_2 x^2$.

→Putem determina coeficienţii a_0, a_1, a_2 din:

$$\begin{aligned} na_0 + a_1\sum_{i=1}^{n}x_i + a_2\sum_{i=1}^{n}x_i^2 &= \sum_{i=1}^{n}y_i \\ a_0\sum_{i=1}^{n}x_i + a_1\sum_{i=1}^{n}x_i^2 + a_2\sum_{i=1}^{n}x_i^3 &= \sum_{i=1}^{n}x_iy_i \\ a_0\sum_{i=1}^{n}x_i^2 + a_1\sum_{i=1}^{n}x_i^3 + a_2\sum_{i=1}^{n}x_i^4 &= \sum_{i=1}^{n}x_i^2y_i. \end{aligned}$$

Scris sub formă matriceală, avem:

$$\begin{bmatrix} n & \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 \\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i^3 \\ \sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i^3 & \sum_{i=1}^{n}x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_iy_i \\ \sum_{i=1}^{n}x_i^2y_i \end{bmatrix}.$$

- Procedeu de rezolvare în Matlab:

1) folosind relaţiile de mai sus
sau:
2) Datele stocate în vectorul y se pot modela, de exemplu, printr-un polinom de gradul II, ceea ce înseamnă că ar putea fi reprezentat prin funcţia

$$y = a_0 + a_1x + a_2x^2,$$

coeficienţii determinându-se dintr-un sistem de 6 ecuaţii (dacă vectorul y are 6 elemente) cu 3 necunoscute de forma:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \\ 1 & x_6 & x_6^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}.$$

-Presupunând că sunt date elementele vectorului x şi ale vectorului corespondent y, se formează matricea A, a coeficienţilor sistemului, în funcţie de dimensiunea vectorului x, în raport cu puterile întâi şi a doua ale elementelor acestui vector.

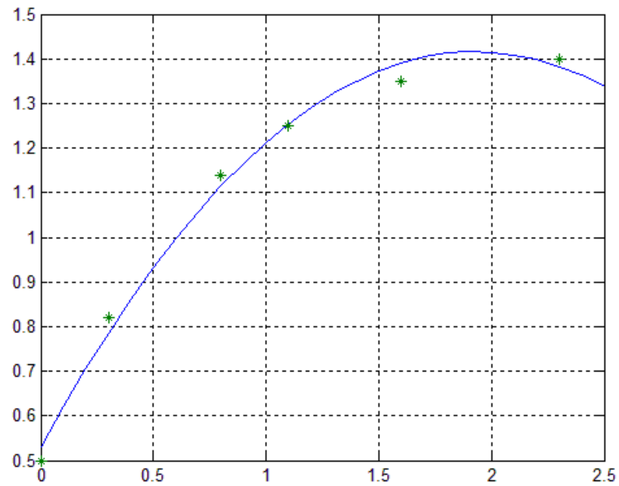
Exemplu.

```
x=[0,0.3,0.8,1.1,1.6,2.3]';
```

```
y=[0.5,0.82,1.14,1.25,1.35,1.40]';
A=[ones(size(x)) x x.^2];
%solutia acestui sistem este data de urmatoarea comanda
X=A\y
%realizam in continuare si reprezentarea grafica
T=(0:0.1:2.5)';
Y=[ones(size(T)) T T.^2]*X;
plot(T,Y,'-',x,y,'*')
grid on
```

- Obţinem:

```
X =
 0.5318
 0.9191
 -0.2387
```



Problema 2. Să se aproximeze un polinom de grad 3 cu datele generate din $y = 2 + 6x^2 - x^3$, cu erori aleatoare date. Erorile aleatoare au o distribuţie normală cu valoare medie 0. Să se folosească funcţiile `polyfit()` şi `polyval()` pentru determinarea coeficienţilor polinomului de grad 3 şi să se reprezinte grafic.

Soluţie.

```
x=0:0.25:6;
y=2+6*x.^2-x.^3;
y=y+randn(size(x));
```

```
xx=0:.02:6;
p=polyfit(x,y,3)
yy=polyval(p,xx);
plot(x,y,'*',xx,yy)
axis([0 6 0 40])
xlabel('x');ylabel('y');
```

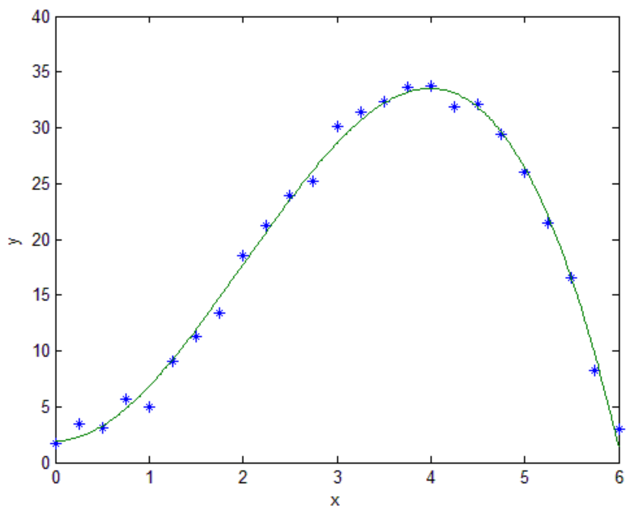
- După rulare, sunt afişaţi coeficienţii polinomului în ordinea descrescătoare a puterilor lui x :

```
p =
 -1.0007 6.0078 -0.1291 1.9407
```

% la fiecare rulare, se obţin alţi coeficienţi!!! s-a folosit funcţia `randn()`

- În acest caz, avem: $y = 1.9407 - 0.1291x + 6.0078x^2 - 1.0007x^3$

- reprezentarea grafică:



Problema 3. Să se aproximeze un polinom de grad 2 şi un polinom de grad 3 cu datele generate de $y = e^x$. Să se folosească funcţiile `polyfit()` şi `polyval()` pentru determinarea coeficienţilor polinomului şi să se reprezinte grafic.

Soluţie.

```
x=0:.25:4;
y=exp(x);
xx=0:0.02:4;
p2=polyfit(x,y,2)
```

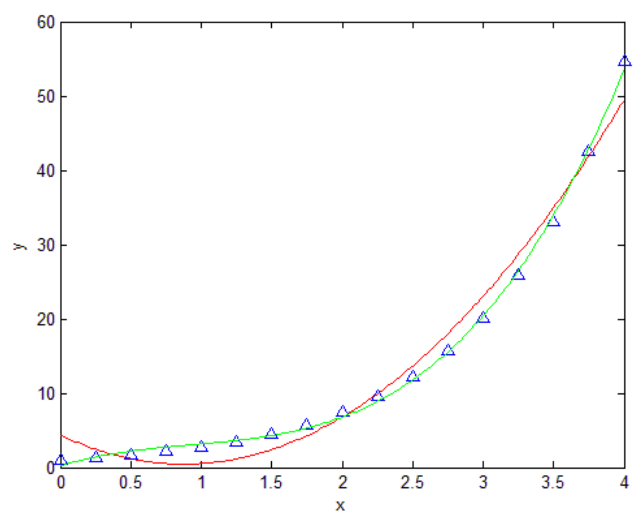
```
yy=polyval(p2,xx);
plot(x,y,'^',xx,yy,'r')
axis([0 4 0 60])
hold on
p3=polyfit(x,y,3)
yy=polyval(p3,xx);
plot(x,y,'^',xx,yy,'g')
xlabel('x');ylabel('y')
hold off
```

- După rulare, similar cu problema precedentă, sunt afişaţi coeficienţii polinomului în ordinea descrescătoare a puterilor lui x :

```
p2 =
 5.0227 -8.8196 4.3264
p3 =
 1.5648 -4.3659 5.7523 0.2189
```

- De asemenea, în figura de mai jos sunt reprezentate grafic aproximările polinoamelor de grad 2 şi 3 cu datele generate de o funcţie exponenţială.

Ce se observă? Răspuns: Polinomul de grad 3 aproximează datele foarte bine.



Problema 4. Aproximaţi un polinom de gradul 3 şi unul de gradul 5 cu ajutorul datelor generate de funcţia

$$y = \sin\left(\frac{1}{x+0.2}\right) + 0.2x,$$

funcţie peste care se suprapune un semnal de zgomot, distribuită normal cu deviaţia standard de 0.06.

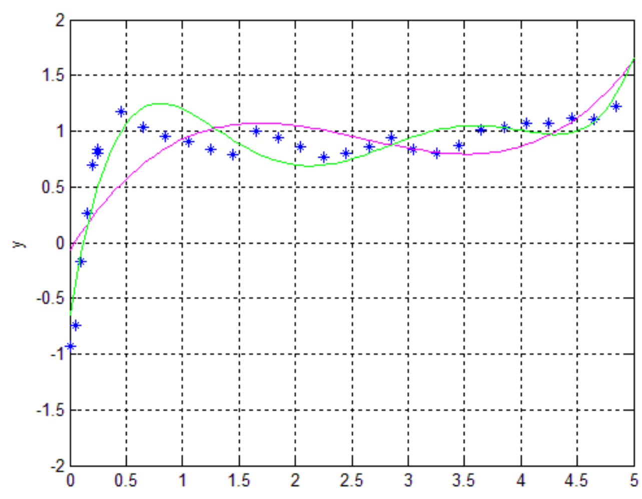
Soluţie.

```
xs=[0:0.05:.25 .25:0.2:4.85]
us=sin(ones(size(xs))./(xs+0.2))+0.2*xs+0.06*randn(size(xs))
save testdata xs us
```

- Cele 30 de valori de date sunt memorate în fişierul *testdata.mat*
- Programul următor încarcă datele, aproximează şi reprezintă grafic polinoamele generate

```
load testdata
xx=0:0.05:5;
p=polyfit(xs,us,3);
yy=polyval(p,xx);
plot(xs,us,'*',xx,yy,'m')
grid on
hold on
axis([0 5 -2 2])
p=polyfit(xs,us,5);
yy=polyval(p,xx);
plot(xx,yy,'g')
xlabel('x');ylabel('y')
hold off
```

- Figura următoare reprezintă rezultatele aproximărilor polinoamelor de gradul 3 şi 5. Sunt evidenţiate nepotrivirile acestor aproximaţii polinomiale:



3 Probleme propuse

1. Se consideră setul de date

$$\begin{aligned} x &= [0.1, 0.3, 0.5, 0.7, 0.9, 0.9] \\ y &= [0.71, 0.98, 0.99, 1.42, 1.75, 2.51]\,. \end{aligned}$$

Să se determine dreapta de regresie liniară şi să se reprezinte grafic rezultatul aproximării în sensul celor mai mici pătrate.

2. Să se construiască un polinom de gradul 2 care să aproximeze următoarele puncte obţinute prin măsurători experimentale:

$$\begin{aligned} x &= [0.2, 0.5, 0.6, 0.6, 0.7, 0.8] \\ y &= [0.89, 0.97, 1.09, 1.68, 1.97, 2.35]. \end{aligned}$$

Să se reprezinte grafic rezultatul aproximării.

3. Se consideră setul de date:

$$\begin{aligned} x &= [0, 0.4, 0.9, 1.2, 1.7, 2.5] \\ y &= [0.68, 0.99, 1.09, 1.59, 1.79, 2.49]\,. \end{aligned}$$

(a) Să se determine coeficienţii estimatorului pătratic (rezolvarea sistemului de ecuaţii - regresie polinomială)
(b) Să se reprezinte grafic rezultatul aproximării din regresia polinomială.

4. Să se aproximeze un polinom de grad 2 şi un polinom de grad 3 cu datele generate de funcţia $y = xe^{2x}$. Să se folosească funcţiile `polyfit()` şi `polyval()` pentru determinarea coeficienţilor polinomului şi să se reprezinte grafic.

5. Aproximaţi un polinom de gradul 4 şi unul de gradul 7 cu ajutorul datelor generate de funcţia

$$y = \cos\left(\frac{x}{x + 0.4}\right) + 0.4 \sin(x^2 + 0.4),$$

funcţie peste care se suprapune un semnal de zgomot, distribuită normal cu deviaţia standard de 0.05.