**Curs 2-3. Învăţare supervizată. Predicţie prin modele de regresie**

**liniară, polinomială şi multiliniară**

Să considerăm un proces (experimental)  care are ca variabilă de intrare (input) variabila  şi care are ca variabilă de ieşire (output) variabila 







Să considerăm acum că pentru , valoarea de output a procesului  este  Repetând acest proces experimental pentru diverse valori ale variabilei de input , colectăm valori ale variabilei de output , valori pe care le trecem într-un tabel de forma:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | x3 | … | xn-1 | xn |
| Y | y1 | y2 | y3 | … | yn-1 | yn |

**Target:** Modelarea dependenţei funcţionale dintre variabila de input  şi variabila de output  cu scopul de a realiza o **PREDICŢIE** (ideal optimă) asupra valorii posibile a variabilei output data fiind valoarea variabilei input. Aşadar, se doreşte determinarea unei funcţii  de o anumită formă pentru care

.

Funcţia  va mima comportamentul/dependenţa dintre cele două variabile.

***Observaţie:***

Dacă valorile variabilei  sunt diferite două câte două, deci pot fi ordonate crescător,

, putem aproxima dependenţa funcţională dintre variabila  şi variabila , printr-o tehnică de interpolare. Aşadar, se poate determina o funcţie  , de un anumit tip (polinomială, polinomială pe porţiuni) astfel încât,

.

Cu ajutorul acestei funcţii putem mima comportamentul procesului , adică putem aproxima , putem realiza predicţii**, însă doar pentru valori din intervalul **

Sunt momente în care tehnica de interpolare nu se poate folosi, sau nu este bine să o folosim.

1. **Tehnica de interpolare, se poate folosi exclusiv dacă valorile  sunt distincte.**

Se poate întampla, însă, ca experimentul procesului  să fie efectuat, de mai multe ori, pentru aceeaşi valoare a variabilei input şi acesta să genereze valori diferite a variabilei output 

***Exemplu:*** Cazul proceselor care implică măsurare, adică variabilase obţine printr-o măsurătoare asociată valorii variabilei . De exemplu, o persoană efectuează măsurători repetate ale temperaturii în sol la aceleaşi adâncimi. La fiecare măsurare a temperaturii la o anumită adâncime, poate obţine valori diferite.

1. **Procesul experimental  , dacă este o modelare matematică a unui proces fizico-ingineresc** sau dacă este un proces experimental, este supus erorilor. Aceste erori vor face ca valorile variabilei output că fie contaminate şi ele de erori. Prin urmare, dacă pentru  procesul  a returnat valoarea valoarea  poate să nu fie cea corectă (adevarată) ci valoarea coretă plus o eroare

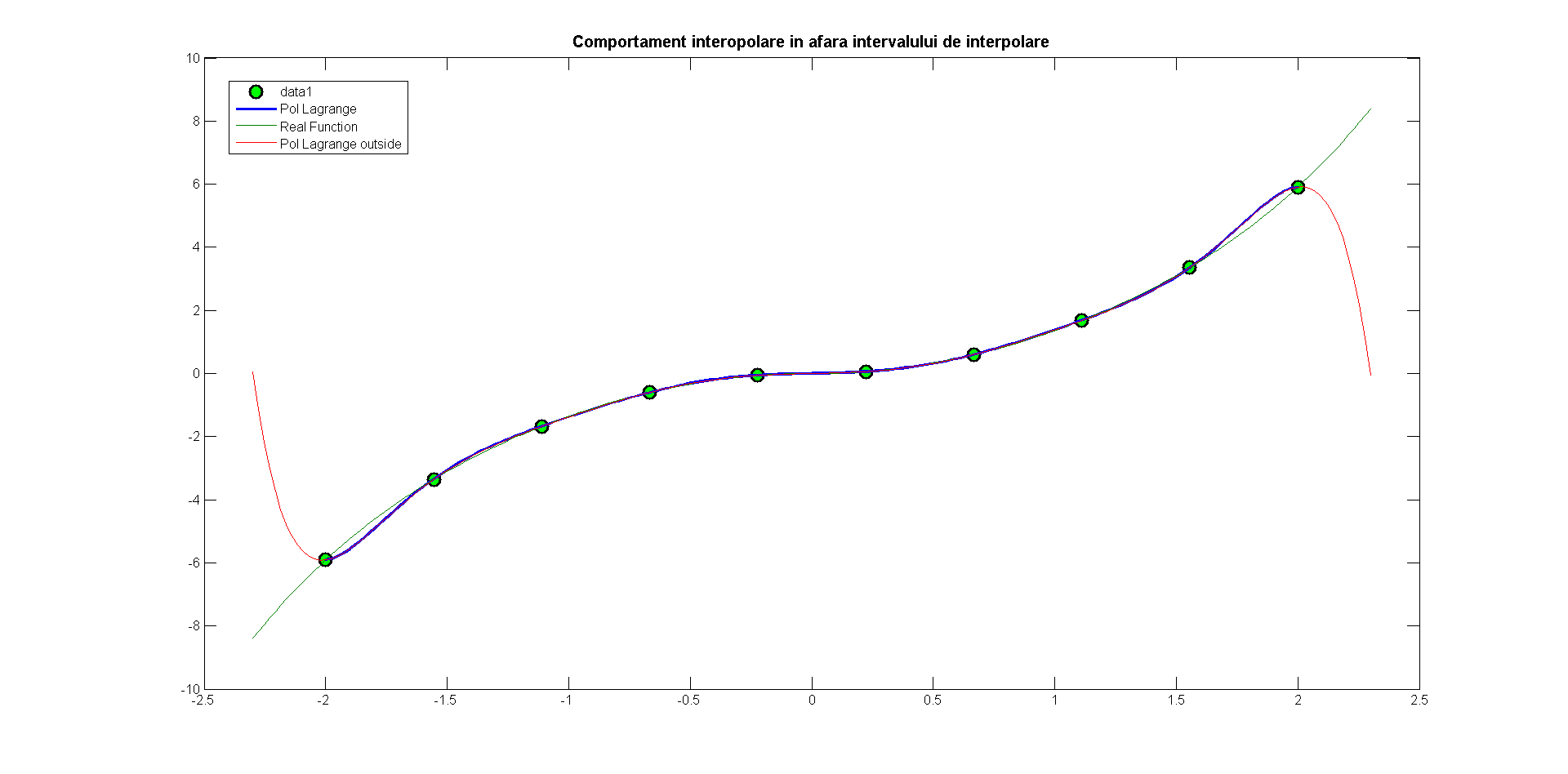


Din acest motiv condiţiile de interpolare, , se pot relaxa în sensul

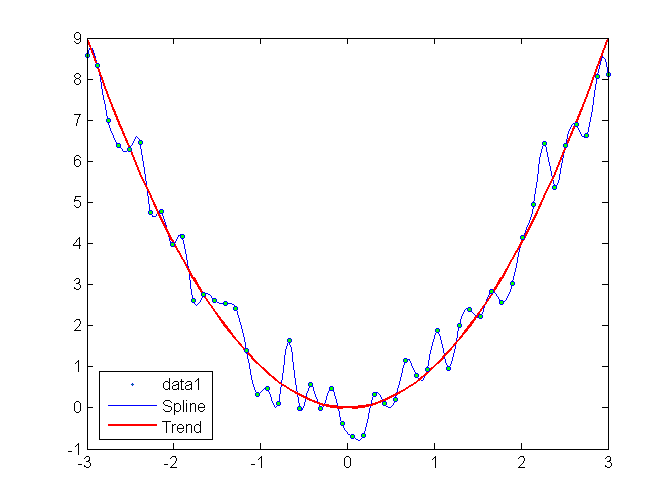
.

1. **Chiar dacă putem aplica o tehnică de interpolare, comportamentul procesului  se poate mima prin funcţia de interpolare doar în interiorul intervalului;** pentru valori din exteriorul acestui interval funcţia de interpolare  nefiind modelată**. Nu se pot realiza predicţii in afara domeniului datelor.**

**Exemplu:**

****

1. Dacă condiţiile  ,pentru a putea interpola, sunt îndeplinite însă numărul de date este foarte mare atunci interpolarea oferă o funcţie care mimează comportamentul procesului  însă într-o manieră de aşa numită **supra-condiţionare**. Comportamentul unei astfel de funcţii poate apărea de tipul următor



Deşi interpolarea cu funcţiile spline cubice este coretă, funcţia de interpolare nu surprinde **trendul datelor**, surprinde numai comportamentul local dat de condiţiile de interpolare . Trendul (comportamentul global) este surprins de funcţia al cărei grafic este roşu. Are loc aşa numitul fenomen de **OVERFITTING.**

**Modele de regresie ---> DATA DRIVEN MODELS**

Să considerăm că avem colectate (date statistice), pe care le avem trecute în tabelul de mai jos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | x3 | … | xn-1 | xn |
| Y | y1 | y2 | y3 | … | yn-1 | yn |

**Se pune problema determinării unei dependenţe funcţionale dintre variabila de input şi variabila de output , prin intermediul unei funcţii pe care de această dată o vom nota prin , care să surprindă cât mai bine trendul oferit de datele din tabel.**

Pentru determinarea trendului, pentru început se reprezintăpunctele , într-un sistem de coordonate . Distribuţia spaţială a norului de puncte  va defini prin forma sa tipul de funcţie al cărei grafic să îl reprezinte geometric cât mai fidel.

Datorită faptului că, se poate întâmpla ca să existe pentru care , funcţia  nu va mai putea să îndeplinească neapărat condiţiile de interpolare  (**NICI NU E DORIT ACEST LUCRU---OVERFITTING).** Funcţia  poartă denumirea de **funcţie de regresie** şi scriem

 sau 

unde prin variabila  exprimăm eroarea **modelului de regresie.**

Avem,

**** unde, este eroarea în nodul 

Determinarea parametrilor funcţiei de regresie se va face prin metoda numită **Metoda celor mai mici pătrate (MCMMP, Least Square Error---LSE sau mai exact Mean Least Square Error---MLSE)** atribuită lui Gauss.

***Observaţie:* Provenienţa datele din tabel poate fi foarte diversă: date colectate în urma unor chestionare, date ce provin din măsurători asociate unor fenomene, date ce provin din simulări numerice, date experimentale, date preluate din observaţii cu echipamente etc.**

Datele pe baza cărora se face analiza dependenţei pot avea sau nu incertitudine ataşată.

**Exemple:**

1. Variabila **X ->** **vârsta** iar variabila **Y-> salariu**. Ne propunem să detrminăm dependenţa/corelaţia dintre vârsta unui angajat (se un anumit profil) şi salariul său. Ambele variabile nu au asociată incertitudine
2. Variabila **X -> altitudine** iar variabila **Y -> concentraţie de agent poluant**. Să presupunem că efectuăm experimente în sensul măsurării concentraţiei unui agent poluant la o anumite altitudini într-o anumită zonă. Achizitionarea datelor se face cu o dronă. Ne propunem să determină dependenţa dintre altitudine şi concentarţie. Ambele variabile au asociate incertitudine deoarece la achizitionarea datelor se folosesc instrumente cu eroare de măsurare iar conditiile atmosferice pot introduce o supra incertitudine.

**Modelul de regresie polinomial**

Să presupunem că avem colectate datele următoare care vizează variabilele  şi 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | x3 | … | xn-1 | xn |
| Y | y1 | y2 | y3 | … | yn-1 | yn |

Dorim să modelăm dependenţa funcţională dintre variabila  şi variabila printr-o funcţie polinomială de grad *k* de forma



Atunci pentru a determina funcţia avem nevoie să determinăm cei *k+1* parametrii ai acesteia: 

În fiecare nod ,

unde avem valoarea variabilei  rezultată din procesul experimental  aceasta se va aproxima prin  de unde va rezulta eroarea de aproximare



Această eroare este diferenţa **dintre** **observaţie** (measurement data) şi **predicţia modelului** .

**Metoda celor mai mici pătrate (MCMMP) presupune determinarea parametrilor funcţiei  astfel încât suma pătratelor erorilor să fie minimă (de fapt, mai corect ar fi să spunem că media pătratelor erorilor este minimă).**

Se defineşte funcţia de cost



pe care trebuie să o minimizăm. Minimizarea acestei funcţii se face uşor folosind metoda analitică.

Pentru aceasta vom determina punctele critice egalând cu zero derivatele parţiale ale funcţiei .



Împărţim cu 2 fiecare ecuaţie şi obţinem



Adică sistemul:



care se numeşte sistemul normal lui Gauss, sistem liniar de *k+1* ecuaţii în necunoscutele 

Matricea acestui sistem este Hessiana funcţiei  , adică

.

Se demonstrează, folosind criteriul lui Sylvester, că această matrice este pozitiv definită deci soluţia unică a sistemului liniar este punct de minim global al funcţiei  Rezolvarea sistemului liniar ne oferă valorile parametrilor modelui de regresie polinomială. Dacă notăm prin  soluţia sistemului liniar, funcţia de regresie polinomială se scrie:



Pentru evaluarea erorii modelului se folosesc 2 indicatori şi anume eroarea medie (**Mean Square Error, MSE**)

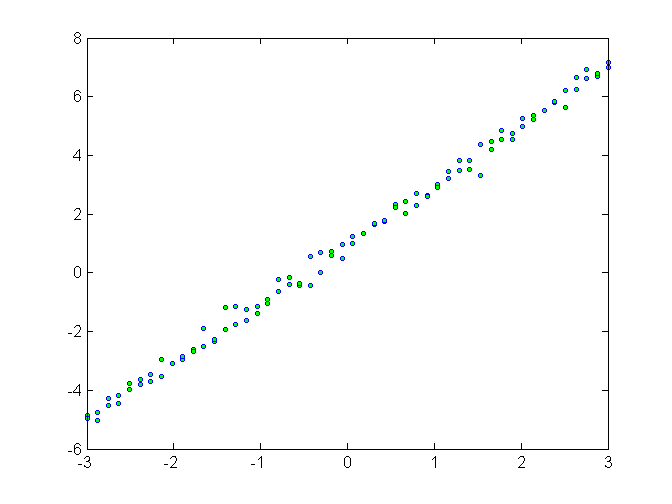
 sau radicalul său (**Root Mean Square Error, RMSE**),



Să particularizăm acum prin cazurile polinoamelor de gradul 1 şi 2.

**Regresia liniară**

In feneral, acest tip de regresie se foloseşte atunci când norul de puncte  se distribuie spaţial după o dreaptă (linie). Un exemplu este reprezentat de figura de mai jos:



Modelăm dependenţa funcţională dintre variabila şi variabila printr-o funcţie liniară de forma



În fiecare nod, unde avem valoarea variabilei  rezultată din procesul experimental  aceasta se va aproxima prin  de unde va rezulta eroarea de aproximare



Parametrii modelului liniar de regresie  se determină folosind MCMMP, minimizând suma pătratelor erorilor adică funcţia



Egalăm cu zero derivatele parţiale ale funcţiei  şi obţinem sistemul liniar



Matricea sistemului este Hessiana funcţiei  şi calculând minorii

 ultima fiind inegaliatea Cauchy-Buniakovski-Schwarz. Rezultă că soluţia sistemului este punct de minim global al funcţiei . Dacă notăm prin  soluţia sistemului, funcţia liniară de regresie se scrie

.

Eroarea modelului de regresie devine

.

**Aplicaţie:**

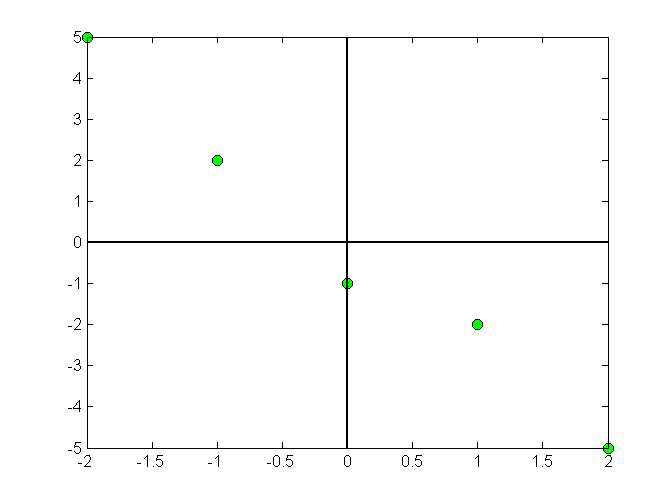
**Să se determine modelul de regresie care modelează cât mai bine dependenţa funcţională dintre variabilele X si Y, având la dispoziţie datele:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **X = *xi*** | **-2** | **-1** | **0** | **1** | **2** |
| **Y = *yi*** | **5** | **2** | **-1** | **-2** | **-5** |

**Să se reprezinte grafic, cât mai fidel, curba de regresie împreună cu norul de puncte.**

Soluţie:

Pentru a vedea ce model de regresie se potriveşte cel mai bine reprezentăm grafic, într-un sistem de coordonate carteziene , norul de puncte asociat tabelului de date.



Se observă o distribuţie spaţială liniară deci alegem să modelăm dependenţa funcţională dintre variabilele  şi  printr-un **model liniar de regresie:**



Parametrii  ai funcţiei se determină folosind MCMMP minimizând funcţia



Egalăm cu zero derivatele parţiale ale funcţiei  şi obţinem sistemul liniar

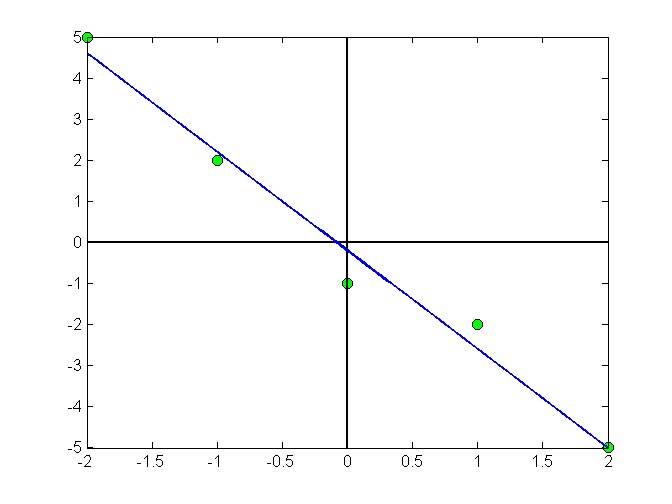


Avem  deci sistemul devine

. Rezultă funcţia de regresie

.

Reprezentarea grafică a funcţiei în, împreună cu norul de puncte asociat este:



*O*

*y*

*x*

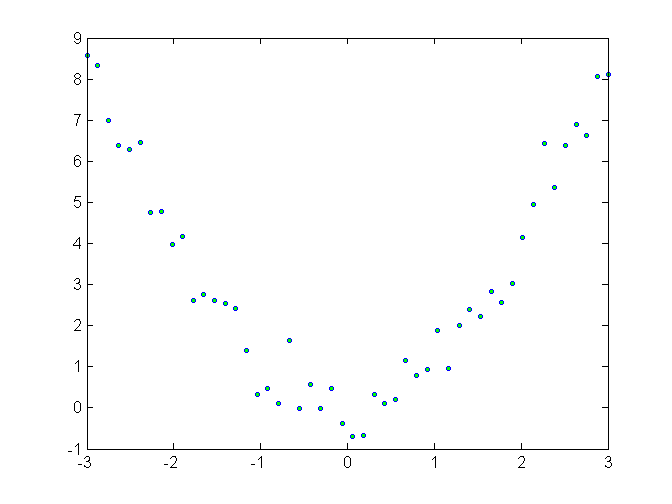
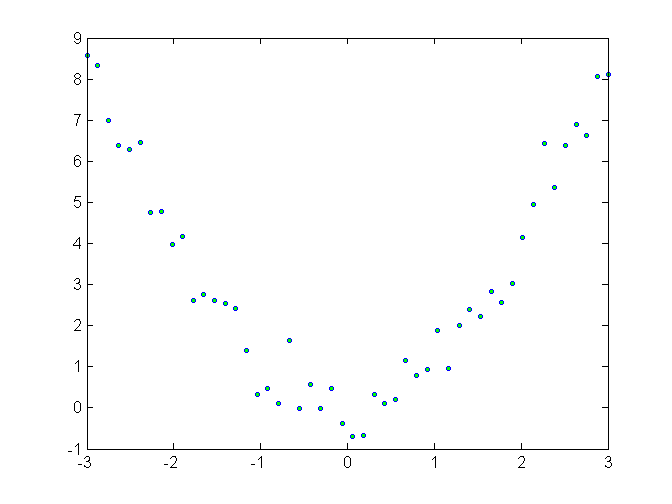
**Regresia pătratică (parabolică)**

Acest tip de regresie se foloseşte atunci când norul de puncte  se distribuie spaţial după o parabolă.

Să presupunem că avem datele următoare

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | x3 | … | xn-1 | xn |
| Y | y1 | y2 | y3 | … | yn-1 | yn |

obţinute din procesul experimental şi reprezentând grafic punctele  obţinem norul de puncte din figura de mai jos

****

Evaluând distribuţia spaţială a norului de puncte, putem observa că acesta are trendul unei parabole ceea ce face ca să căutam o **functie de regresie** de gradul al doilea.

Modelăm dependenţa funcţională dintre variabila şi variabila printr-o funcţie de gradul al doilea



Parametrii  ai funcţiei se determină folosind MCMMP minimizând funcţia



Egalăm cu zero derivatele parţiale ale funcţiei  şi obţinem sistemul liniar





Rezolvarea sistemului liniar ne oferă valorile parametrilor modelui de regresie polinomială. Dacă notăm prin  soluţia sistemului liniar, funcţia de regresie polinomială se scrie:



Eroarea (medie) modelului de regresie (Mean Square Error, MSE)devine

.

**Aplicaţie:**

**Să se determine modelul de regresie care modelează cât mai bine dependenţa funcţionala dintre variabilele X si Y, având la dispoziţie datele:**

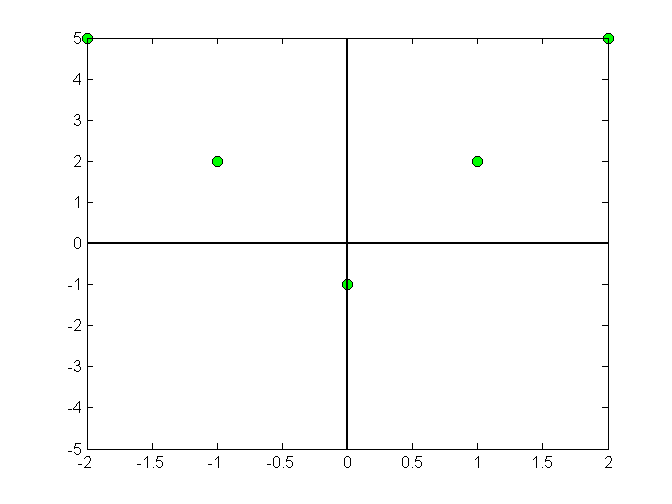
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **X = *xi*** | **-2** | **-1** | **0** | **1** | **2** |
| **Y = *yi*** | **5** | **2** | **-1** | **2** | **5** |

**Să se reprezinte grafic, cât mai fidel, curba de regresie împreună cu norul de puncte.**

*Soluţie:*

Pentru a vedea ce model de regresie se potriveşte cel mai bine reprezentăm grafic, într-un sistem de coordonate carteziene , norul de puncte asociat tabelului de date.

*y*



*O*

*x*

Se observă o distribuţie spaţială ce aduce cu o parabolă deci alegem să modelăm dependenţa funcţională dintre variabilele  şi  printr-un model parabolic de regresie:



Parametrii  ai funcţiei se determină folosind MCMMP minimizând funcţia



Egalăm cu zero derivatele parţiale ale funcţiei  şi obţinem sistemul liniar



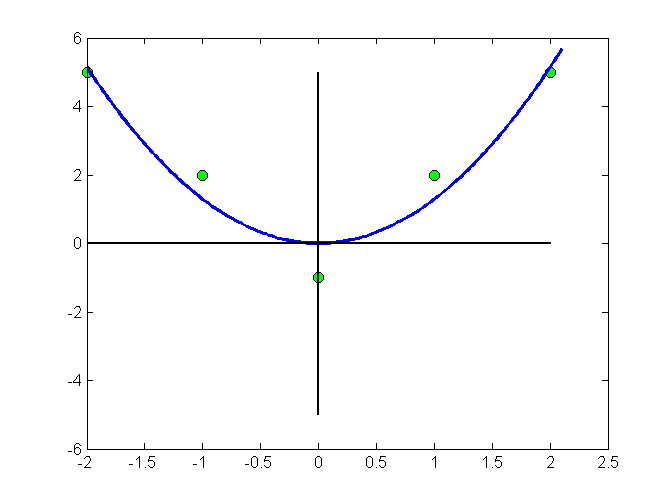


Avem  deci sistemul devine

 cu soluţia  Rezultă funcţia de regresie

.

Reprezentarea grafică a norului de puncte şi a parabolei de regresie este



*x*

*y*

*O*

**Componenta probabilistică/statistică a modelului liniar de regresie**

Plecăm de la sistemul de ecuaţii pe baza căruia se determină parametrii modelului liniar de regresie

.

Determinanatul sistemului este 

Atunci, folosind Cramer,



Aşadar, panta modelului e proporţională cu covarianţa dintre variabile. Dacă variabilele ar fi independente, covarianţa ar fi 0 deci panta 0 ceea ce face dreapta de regresie să fie orizontală.



Eroarea modelului se scrie





**Modelul de regresie multiliniară**

Cu ajutorul modelelor de regresie studiate anterior putem modela un proces (experimental)  care are ca variabilă de intrare (input) variabila  şi ca variabilă de ieşire (output) variabila 







Aşadar, dependenţa funcţională se scrie.

Cele mai multe procese aflate sub studiu au mai multe variabile de input (numite variabile independente) în proces  şi o variabilă de output  (numită variabilă dependentă)











În acest context pentru un set de valori  a variabilelor de intrare valoarea de output a procesului  este  Repetând acest proces experimental pentru diverse valori ale variabilelor de input , colectăm valori ale variabilei de output , valori pe care le trecem într-un tabel de forma:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X1 | x1(1) | x2(1) | x3(1) | … | xn-1(1) | xn(1) |
| X2 | x1(2) | x2(2) | x3(2) | … | xn-1(2) | xn(2) |
| .  .  . | .  .  . | .  .  . | .  .  . | .  .  . | .  .  . | .  .  . |
| Xk | x1(k) | x2(k) | x3(k) | … | xn-1(k) | xn(k) |
| Y | y1 | y2 | y3 | … | yn-1 | yn |

Modelul de regresie multiliniară descrie dependendenţa dintre variabilele  şi variabila printr-o funcţie de forma



Sau

,

unde variabila  defineşte eroarea modelului.

Parametrii  se determină folosind MCMMP minimizând suma pătratelor erorilor individuale



Funcţia de minimizat se scrie:



Considerăm notaţiile



Atunci, matriceal funcţia de minimizat se scrie

***Design Matrix***

.

Minimizarea acestei funcţii implică condiţia ca gradientul funcţiei  să fie zero adică



 de unde solutia

.

**Modele de regresie liniare ca modele de învaţare supervizată**

Să presupunem că avem la dispoziţie un set de date care conţine suprafaţa şi preţul unor apartamente dintr-un cartier din Bucureşti (de fapt Colentina ☺). Plotarea acestor date generează graficul,

****

Având la dispoziţie aceste date ne punem problema să facem o predicţie asupra altor apartamente din aceeaşi zona. Setul de date conţine un număr de n=75 de intrări în care variabila input *X* este **suprafaţa** locuinţei iar variabila output ***Y*** este **preţul** locuinţei.

*O pereche*  *se va numi “****training example****” iar întregul set de date* *se va numi “****training set****”*

***Descrierea problemei de învăţare supervizată:***

*Dat fiind setul de antrenare*  *se doreşte învăţarea unei funcţii*  astfel încât este ***un bun predictor*** *pentru valoarea*  *asociată inputului* .

Schematic, procesul se poate descrie astfel:

Set de antrenare

Algoritm de învăţare

Alg

*h*

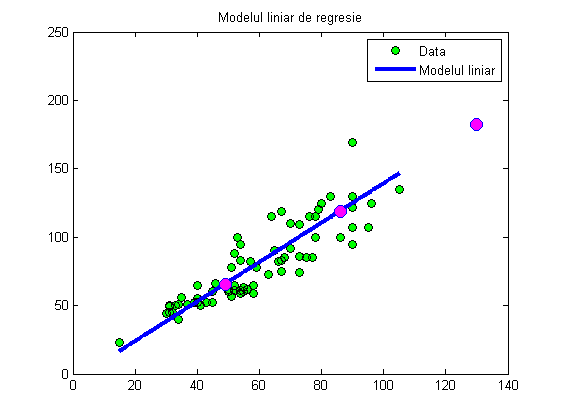
*h(x)*

*predictia pretului*

*x*

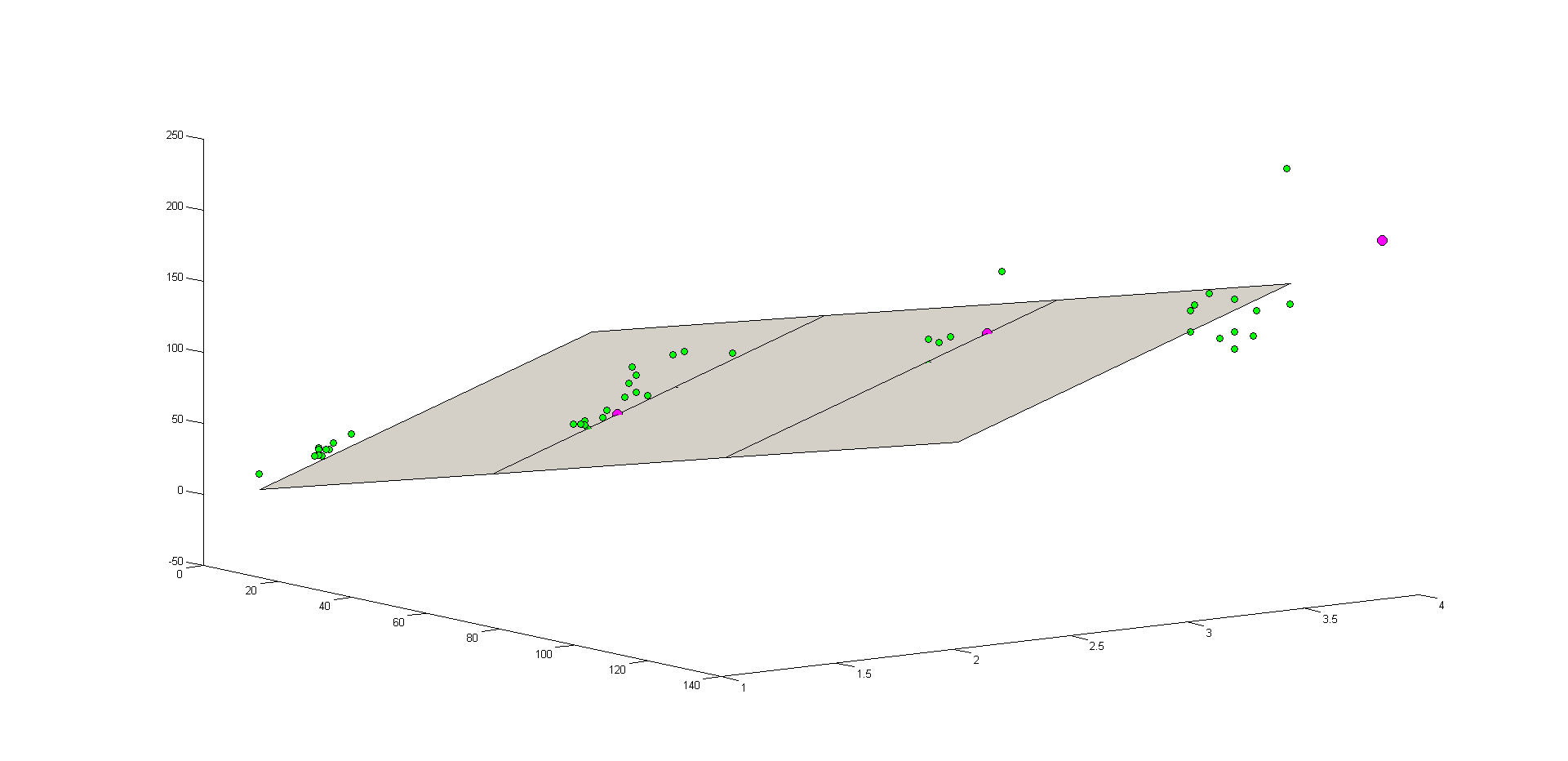
*suprafaţa*

Folosind modelul liniar de regresie găsim funcţia şi trei dintre predicţii (pentru x=49, 86 ,130) se găsesc în figura de mai jos.



Dacă adăugăm informaţie suplimentară, luând în calcul numărul de camere al apartamentelor  şi respectiv suprafaţa  avem o problemă de regresie multiliniară cu estimarea funcţiei de regresie.

Predicţiile pentru aceleaşi valori ale variabilei input se găsesc in figura următoare



**Observaţie:** Predicţiile cu cele două metode de regresie sunt uşor diferite,

|  |  |  |  |
| --- | --- | --- | --- |
| Metoda | X=49, nr\_cam=2 | X=86, nr\_cam=3 | X=130, nr\_cam=4 |
| Regresie liniară | 65.6885 | 119.0592 | 182.5271 |
| Regresie multiliniară | 64.8071 | 123.0543 | 193.8709 |

**Intepretarea probabilistică a modelelor liniare de regresie**

Estimarea parametrilor modelelor de regresie se bazează pe minimizarea unei funcţii cost care ia în considerare suma pătratelor erorilor (diferenţa între predicţia modelului teoretic şi datele efectiv observate).

X (input)

Y (output)





Observaţie:  reprezintă aria unui pătrat de latură , deci curba de regresie reprezintă acea curbă care generează cele mai mici pătrate (***geometrice***) cu punctele norului.

Totuşi, de ce este important ca această condiţie să fie satisfăcută. Să abordăm problema din punct de vedere probabilistic. Vom considera cazul general al regresiei multiliniare, în sensul în care variabila de output (“**target variable**”) se aproximează printr-o funcţie liniară multi-dimensională de tipul



Notaţii: .

În noile notaţii aproximarea se scrie,

.

Atunci în fiecare nod,  , unde . Eroarea  este termenul în care sunt incluse caracteristicile imposibil de modelat ale procesului sub observaţie, şi respectiv posibilele erori perturbatoare care ţin de achiziţia de date (“***random noise***”). Să considerăm adevărată ipoteza că aceste erroi au aceeaşi **distribuţie Gaussiană** şi sunt **independente.** Aşadar,

.

Rezultă, că densitatea de repartiţie a erorii este,



Intrucât,  rezultă că distribuţia lui  condiţionată de având ca parametrii  este,

.

Datorită independenţei errorilor, putem determina distribuţia vectorului aleator

 condiţionat de ca fiind,



Date fiind datele de selecţie , parametrii  ai modelului de determnă cu metoda verosimilităţii maxime considerând funcţia de verosimilitate,

.

Maximizarea acestei funcţii presupune minimizarea funcţiei



care este exact funcţia de cost din metoda celor mai mici pătrate.

Concluzie: Sub condiţiile probabilistice ale distribuţiilor erorilor (e.g. Gaussiane), estimarea parametrilor modelului liniar prin metoda celor mai mici pătrate este echivalentă cu estimarea parametrilor folosind metoda verosimilităţii maxime. Deasemenea, din estimare reiese că valorile parametrilor modelului nu depind de dispersia  a distribuţiei erorilor. Acest lucru nu mai este valabil dacă erorile nu sunt identic distribuite.

**UNDERFITTING vs OVERFITTING**

Să considerăm problema anterioară in care se doreşte realizarea unei predicţii a preţului unor locuinţe având la dispoziţie un set de antrenare. Ne restricţionăm la cazul apartamentelor cu 2 camere iar figura de mai jos surprinde modelarea cu o regresie liniară .

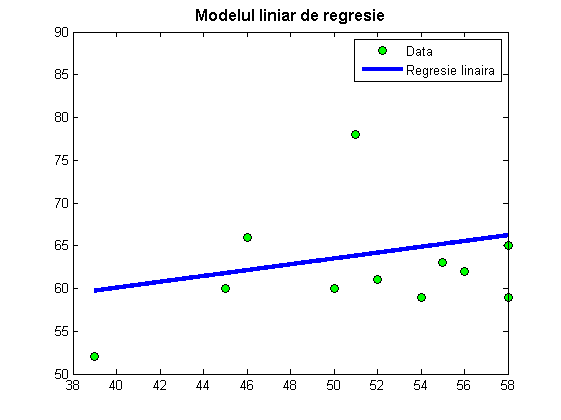
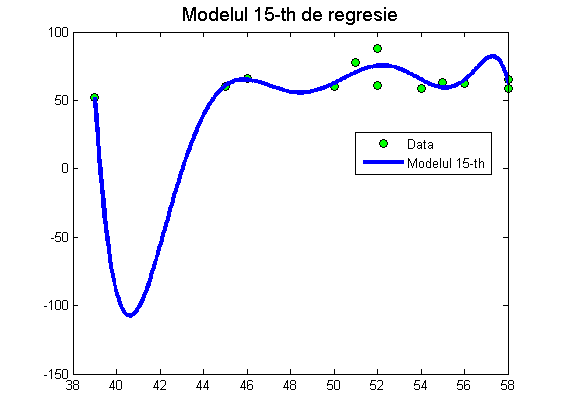
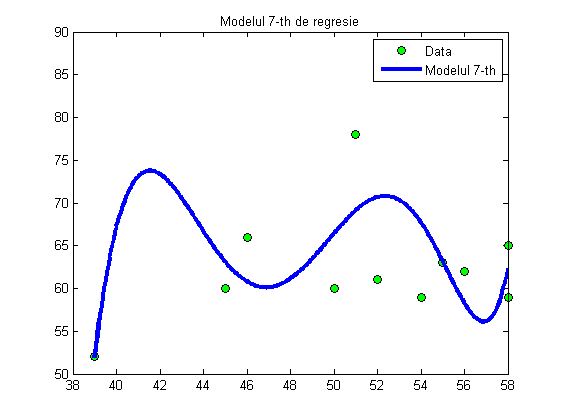
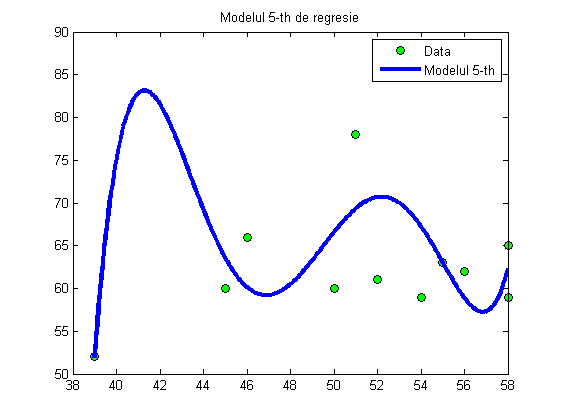
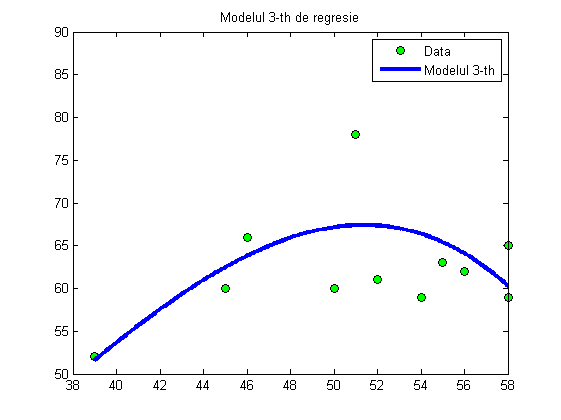
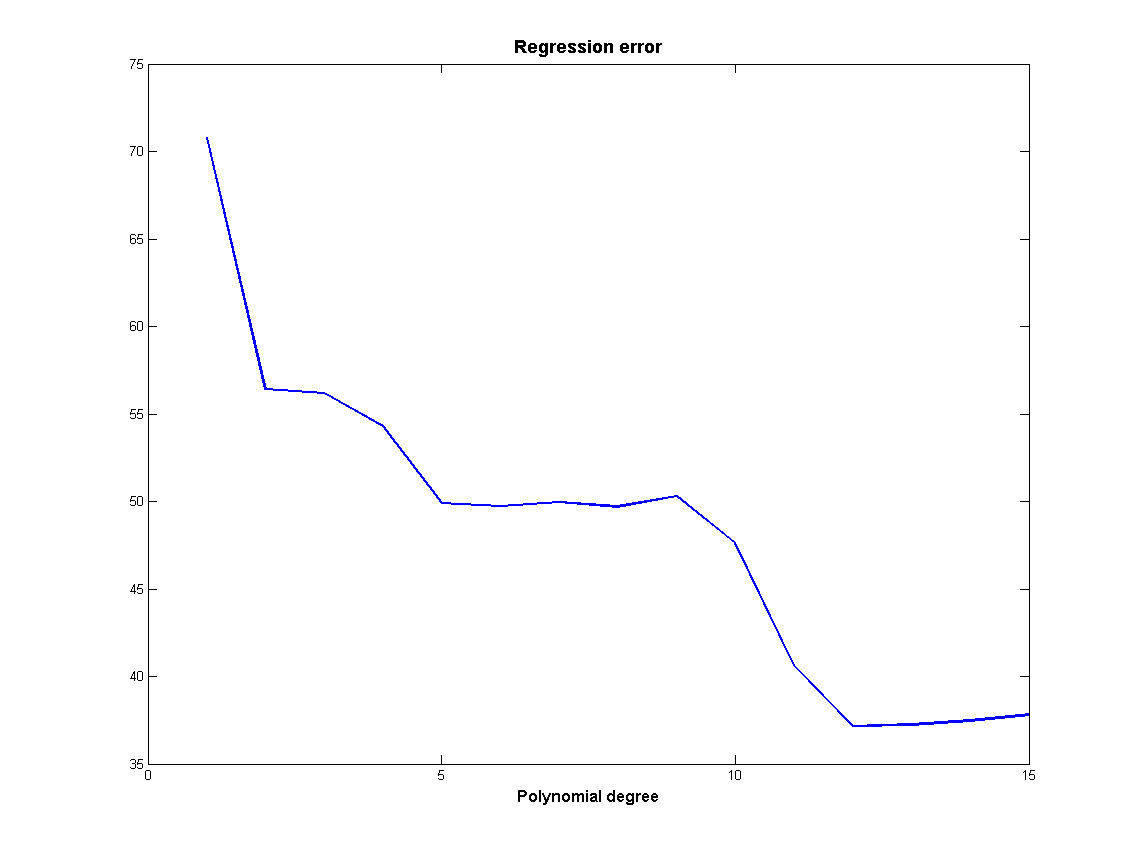


Figura de mai sus surprinde trendul însă, dreapta de regresie pare destul de departe de puncte, caz în care eroarea totală este destul de mare ceea ce face ca predicţia să nu fie cu o acurateţe acceptabilă. Ce putem face este să realizăm modelări cu regresii polinomiale cu grade din ce în ce mai mari. Figurile de mai jos prezintă modelele polinomiale de regresie de ordine 3,5,7,15.



In figura următoare este plotarea eroriilor modelelor.



Se observă că eroarea scade, până la un punct după care începe să staţioneze, sau chiar să scadă uşor. Au loc două efecte. In regresia liniară şi cele cu polinoame de grade mici are loc efectul de **UNDERFITTING, în care datele arată o structură pe care modelul nu o surprinde.** In cazul regresiilor polinomialeîn care gradul polinomului este mare are loc un fenomen de **OVERFITTING, caz în care modelul nu va realiza prediţii bune pentru anumite valori ale variabilei input. Ideea de overfitting vizează o potrivire foarte bună a datelor din setul de antrenare, dar care pentru anumite date noi de intrare oferă predicţii departe de realitate. Două soluţii remediază într-o oarecare măsură situaţia de overfitting:**

1. **Regresii liniare cu ponderare locală**
2. **Regularizare**

**Regresii liniare cu ponderare locală (Locally weighted linear regression)**

Regresia liniară are la bază determinarea unui model liniar care pe cazul general (regresie multiliniară) se scrie



Parametrii modelului se determină după rezolvarea problemei de optimizare,

.

Predicţia se va face, evident prin

.

Se observă că pentru fiecare input nou  predicţia se realizează folosind aceeaşi parametrii ai modelului de regresie. Regresia liniară cu ponderare locală îşi are originile într-o metodă numerică numită “**Element-free Galerkin**” în care se foloseşte o funcţie de pondere, cu care se ponderează pătratul erorilor din funcţia cost în funcţie de distanţa euclidiană de la input la dată. Noua funcţie cost devine,



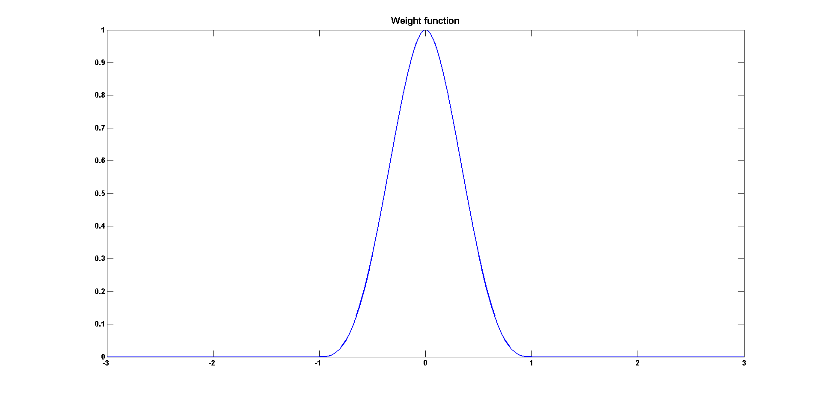
iar parametrii se determină prin,

.

Funcţiile de pondere, au proprietatea că descrescătoare, derivabile, sunt zero peste un anumit prag şi iau valoarea 1 în zero. Un exemplu de functie de pondere este funcţia spline quadratică

,

al cărei grafic este mai jos.

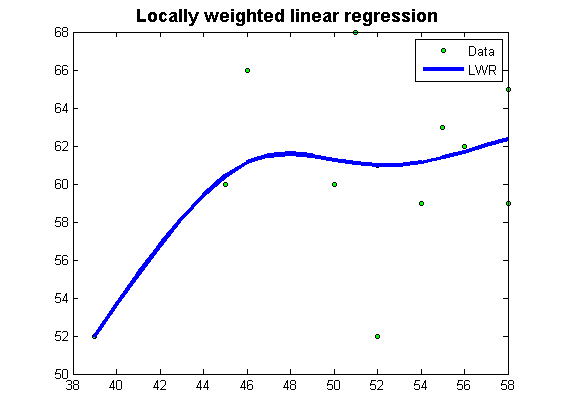


În cazul modelului liniar, pentru fiecare valoare de input , definim funcţia de cost

,

pe care o minimizăm egalând cu zero derivatele parţiale ale funcţiei  ,de unde soluţia.

Rezultatul implementării este in figura de mai jos, în care efectul local este evident (valoarea lui r=12);



**Modele de regresii cu regularizare**

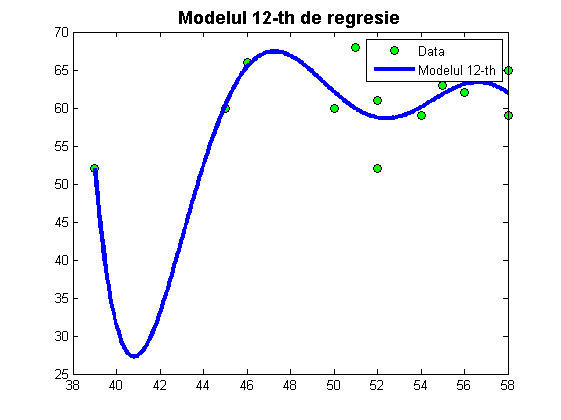
1. **Regularizare Tikhonov sau Ridge**

O soluţie pentru atenuarea overfittingului este folosirea unui termen de regularizare în funcţia de cost a modelului. Să considerăm modelul polinomial de grad 12 a cărui soluţie o avem in figura următoare. Se defineşte o nouă funcţie de cost prin adăugarea unui termen suplimentar. Tehnica se numeşte regularizare Tikhonov sau Ridge (denumire folosita in machine learning).

Noua funcţie de cost devine,

,

în care parametrul de regularizare are o valoare apriori aleasă pe baza unui proces de “trial-error”.



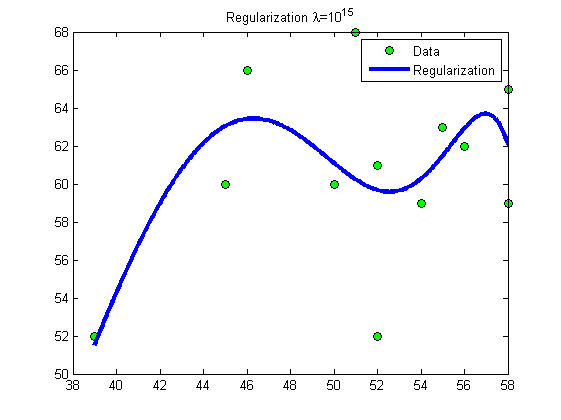
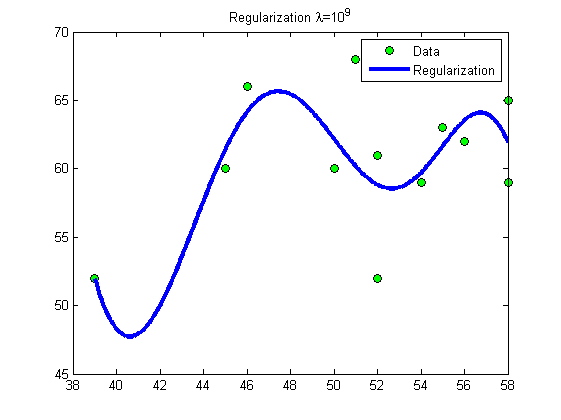
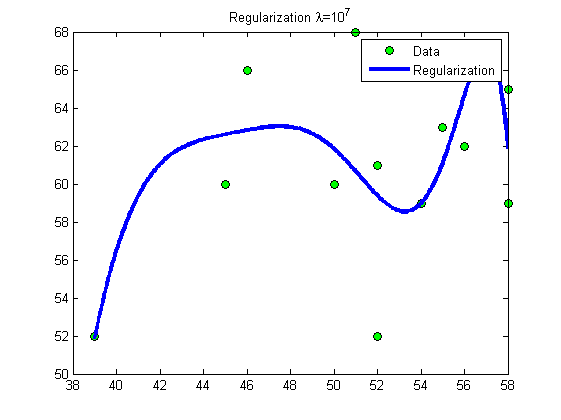
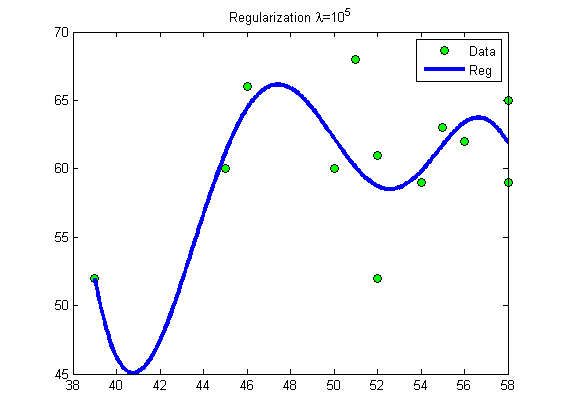
Setăm zero gradientul funcţiei cost şi obţinem

 sau



Solutia sistemului se scrie  unde 

Rezultatele modelelor de regresie cu regularizare Ridge pentru  sunt prezentate în figurile de mai jos.



1. **Regularizare Lasso**

In cazul regularizari Lasso adaugarea termenulu penalizator face ca noua funcţie de cost sa se scrie,

,

Ca și în cazul regularizării Ridge parametrul  controlează influența termenului penalizator, respective se realizează un balans între bias și variance, asă numitul “Bias-Variance Tradeoff” .i.e. creșterea valorii parametrului generează creșteea biasului și scaderea varianței și viceversa. Valoarea optimală se găsește în urma unui process de trial-error sau cross-validare.

Procesul de optimizare al funcției obiectiv de mai sus se realizează nu direct prin tehnica “gradient descent”, datorită lipsei derivabilității în origine, ci prin metoda subgradient descent, coordinate gradient sau proximal gradient decent.

COMENTARII:………………

**Linear regression as single layer neural network**

Activation

linear





Optimization:

Stochastic gradient descent



***Cost function: MSE***

















.