

Identificarea sistemelor

Ingineria sistemelor, anul 3
Universitatea Tehnică din Cluj-Napoca

Lucian Buşoniu



Partea II

Baze matematice:
Regresie liniară. Teoria probabilităților și
statistică

Motivare

Majoritatea metodelor de identificare care urmează necesită **regresia liniară** și concepte de **teoria probabilităților și statistică**. Vom discuta aici aceste unelte matematice.

În această parte anumite notații (de ex. x , A) au o semnificație diferită de cea din restul materialului de curs.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Aproximare: Exemplu motivant

Studiem **venitul anual** (în EUR) al unei persoane bazat pe **nivelul de educație** și **experiența profesională** (ambele măsurate în ani).

Se dă un set de date (nivel de educație, experiență profesională, venit anual) de la un grup reprezentativ de persoane. Obiectivul este **predicția** venitului oricărei alte persoane, care nu face parte din setul de date, pornind de la nivelul său de educație și experiență.

Problema de regresie: Elemente

Elementele problemei:

- Un șir de eșantioane cunoscute $y(k) \in \mathbb{R}$, indexate de $k = 1, \dots, N$: y este *variabila dependentă* (măsurătoarea).
- Pentru fiecare k , un vector cunoscut $\varphi(k) \in \mathbb{R}^n$: conține *regresorii* $\varphi_i(k)$, $i = 1, \dots, n$, $\varphi(k) = [\varphi_1(k), \varphi_2(k), \dots, \varphi_n(k)]^\top$.
- Un *vector de parametri* $\theta \in \mathbb{R}^n$ **necunoscut**.

Model liniar al variabilei dependente:

$$\begin{aligned} y(k) &= \varphi^\top(k) \theta + e(k) \\ &= [\varphi_1(k), \varphi_2(k), \dots, \varphi_n(k)] \cdot [\theta_1, \theta_2, \dots, \theta_n]^\top + e(k) \\ &= \left[\sum_{i=1}^n \varphi_i(k) \theta_i \right] + e(k) \end{aligned}$$

unde $e(k)$ este eroarea modelului

O paranteză despre notația vectorilor

Toate variabilele vectoriale sunt implicit vectori coloană – standardul din automată. Adeseori le scriem sub forma unor linii transpuse, pentru a economisi spațiu vertical:

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = [q_1, q_2, q_3]^\top$$

De notat și: $Q^\top = [q_1, q_2, q_3]$.

De exemplu, în formula modelului liniar :

$$y(k) = [\varphi_1(k), \varphi_2(k), \dots, \varphi_n(k)] \cdot [\theta_1, \theta_2, \dots, \theta_n]^\top + e(k) = \varphi^\top(k)\theta + e(k)$$

$\varphi(k)$ este coloană, dar trebuie transformat în linie pentru ca produsul vectorial să fie corect. Deci îl transpunem: $\varphi^\top(k)$, obținând linia $[\varphi_1(k), \dots]$ (de notat că linia nu este transpusă). Pe de altă parte, θ trebuie să fie coloană în produsul vectorial, deci nu este transpus; dar pentru conveniență îl scriem sub forma unei linii transpuse, $[\theta_1, \dots]^\top$ (observați transpusa!).

Problema de regresie: Obiectiv

Obiectiv: identificarea comportamentului variabilei dependente din date.

Mai exact: Găsirea unor valori pentru parametrii θ astfel încât variabila dependentă aproximată $\hat{y}(k) = \varphi^\top(k)\theta$ să fie cât mai aproape posibil de $y(k)$ pentru orice k ; echivalent, astfel încât $e(k)$ să fie cât mai mici posibil.

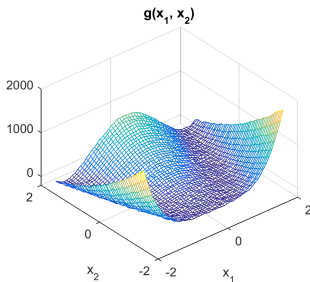
Vom clarifica ulterior ce înseamnă “cât mai aproape/mici”.

Regresia liniară este o metodă clasică și des folosită, de ex. Gauss a folosit-o pentru a calcula orbitele planetelor în 1809.

Problema de regresie: Două utilizări importante

- 1 k este o variabilă de timp, și modelăm seria temporală $y(k)$.
- 2 k este doar un index de eșantion, și $\varphi(k) = \phi(x(k))$ unde x este intrarea unei funcții necunoscute g . În acest caz $y(k)$ este ieșirea corespunzătoare, posibil afectată de zgomot, iar obiectivul este identificarea unui model al funcției g din date. Această problemă se numește *aproximarea unei funcții*, sau *învățarea supervizată*.

$g(x)$ necunoscut



$\hat{g}(x)$



Aproximare: Funcții de bază

Pentru aproximarea unei funcții, regresorii $\phi_i(k)$ din:

$$\phi(x(k)) = [\phi_1(x(k)), \phi_2(x(k)), \dots, \phi_n(x(k))]^\top$$

se numesc *funcții de bază*.

Aproximare: Formalism matematic pentru exemplul motivant

Studiem **venitul anual** y (în EUR) al unei persoane bazat pe **nivelul de educație** x_1 și **experiența profesională** x_2 (ambele măsurate în ani).

Se dă un set de date $(x_1(k), x_2(k), y(k))$ de la un grup reprezentativ de persoane. Obiectivul este **predicția** venitului oricărei alte persoane din nivelul său de educație (x_1) și experiență (x_2).

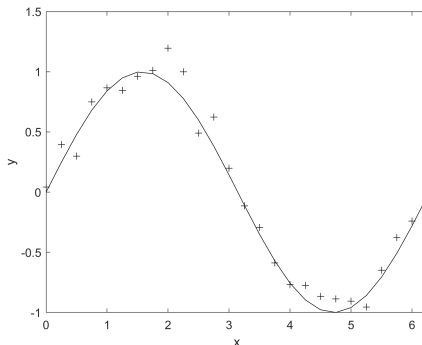
- Alegem funcțiile de bază $\phi(x) = [x_1, x_2, 1]^T$. Ne așteptăm ca venitul să evolueze conform cu $\theta_1 x_1 + \theta_2 x_2 + \theta_3 = \phi^T(x)\theta$, crescând liniar cu educația and experiența (de la un nivel minimal). Regresia implică găsirea parametrilor θ pentru care expresia este cel mai aproape de valorile reale ale venitului.
- În realitate, lucrurile sunt mai complicate... vom avea nevoie de variabile de intrare suplimentare, funcții de bază mai bune, etc.

Aproximare: Exemplu motivant 2

Studiem **timpul de reacție** y (în ms) al unui șofer în funcție de **vârsta** sa x_1 (în ani) și **oboseala** x_2 (de ex. pe o scară de la 0 la 1).

Se dă un set de date $(x_1(k), x_2(k), y(k))$ de la un grup reprezentativ de șoferi de diferite vârste și nivele de oboseală. Obiectivul este **predicția** timpului de reacție al oricărui alt șofer folosind vârsta (x_1) și oboseala (x_2) sa.

Aproximare: Exemplu recurent



Aproximăm $\sin(x)$ pe intervalul $[0, 2\pi]$. De notat: Metoda nu are access la funcție, ci doar la eșantioane stohastice! Vom folosi acest exemplu de-a lungul cursului, pentru a ilustra conceptele.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori**
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Exemplu regresori 1: Polinom în k

Util pentru modelarea seriilor temporale.

$$\begin{aligned} y(k) &= \theta_1 + \theta_2 k + \theta_3 k^2 + \dots + \theta_n k^{n-1} \\ &= \begin{bmatrix} 1 & k & k^2 & \dots & k^{n-1} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \dots \\ \theta_n \end{bmatrix} \\ &= \varphi^\top(k) \theta \end{aligned}$$

Exemplu regresori 2: Polinom în x

Util pentru aproximarea funcțiilor. De exemplu, polinomul de gradul 2 cu două variabile de intrare $x = [x_1, x_2]^T$ este:

$$y(k) = \theta_1 + \theta_2 x_1(k) + \theta_3 x_2(k) + \theta_4 x_1^2(k) + \theta_5 x_2^2(k) + \theta_6 x_1(k)x_2(k)$$

$$= \begin{bmatrix} 1 & x_1(k) & x_2(k) & x_1^2(k) & x_2^2(k) & x_1(k)x_2(k) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{bmatrix}$$

$$= \phi^T(x(k))\theta = \varphi^T(k)\theta$$



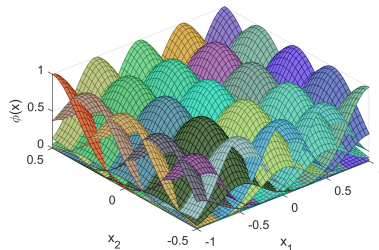
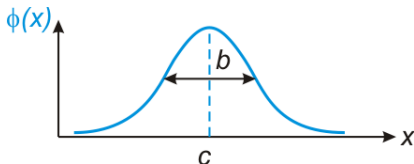
Conexiune: Proiect partea 1

Exemplu regresori 3: Funcții de bază Gaussiene

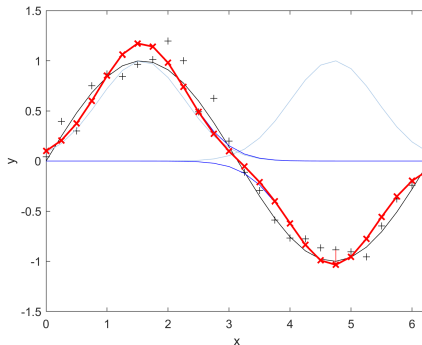
Utile pentru aproximarea funcțiilor:

$$\phi_i(x) = \exp \left[-\frac{(x - c_i)^2}{b_i^2} \right] \quad (1\text{-dim});$$

$$= \exp \left[-\sum_{j=1}^d \frac{(x_j - c_{ij})^2}{b_{ij}^2} \right] \quad (d\text{-dim})$$



Regresori: Exemplu recurent



Cum datele au un “munte” și o “vale”, **două RBFuri** ar trebui să fie suficiente (vezi soluția)

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei**
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Reamintim: Obiectivul problemei de regresie

Obiectiv: identificarea comportamentului variabilei dependente din date.

Mai exact: Găsirea unor valori pentru parametrii θ astfel încât variabila dependentă aproximată $\hat{y}(k) = \varphi^\top(k)\theta$ să fie cât mai aproape posibil de $y(k)$ pentru orice k ; echivalent, astfel încât $e(k)$ să fie cât mai mici posibil.

Sistem liniar

Scriind modelul pentru fiecare din cele N date, obținem un sistem de ecuații liniare:

$$y(1) = \varphi_1(1)\theta_1 + \varphi_2(1)\theta_2 + \dots \varphi_n(1)\theta_n$$

$$y(2) = \varphi_1(2)\theta_1 + \varphi_2(2)\theta_2 + \dots \varphi_n(2)\theta_n$$

...

$$y(N) = \varphi_1(N)\theta_1 + \varphi_2(N)\theta_2 + \dots \varphi_n(N)\theta_n$$

Reamintim că în aproximarea funcțiilor, $\varphi_i(k) = \phi_i(x(k))$

Sistemul se poate scrie în *formă matriceală*:

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \dots & \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) & \dots & \varphi_n(2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(N) & \varphi_2(N) & \dots & \varphi_n(N) \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \dots \\ \theta_n \end{bmatrix}$$

$$Y = \Phi\theta$$

cu noile variabile $Y \in \mathbb{R}^N$ și $\Phi \in \mathbb{R}^{N \times n}$.

Problema celor mai mici pătrate (CMMP)

Dacă $N = n$, sistemul se poate rezolva cu egalitate.

În practică, este mai bine să folosim $N > n$, de ex. datorită zgomotului. În acest caz, sistemul nu mai poate fi rezolvat cu egalitate, ci doar cu aproximare.

- *Eroarea* la k : $\varepsilon(k) = y(k) - \varphi^\top(k)\theta$,
vectorul de eroare $\varepsilon = [\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)]^\top$.
- **Funcția obiectiv** ce trebuie minimizată:

$$V(\theta) = \frac{1}{2} \sum_{k=1}^N \varepsilon(k)^2 = \frac{1}{2} \varepsilon^\top \varepsilon$$

Problema CMMP

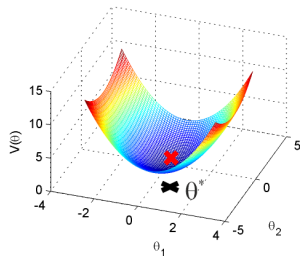
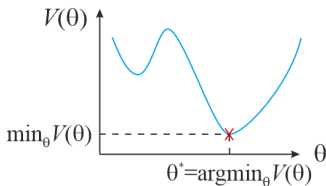
Găsește vectorul de parametri $\hat{\theta}$ care minimizează funcția obiectiv:

$$\hat{\theta} = \arg \min_{\theta} V(\theta)$$

Paranteză: Problema de optimizare

Data fiind o funcție V de variabilele θ , care poate fi de ex. obiectivul nostru CMMP, sau în general oricare altă funcție:

găsește *valoarea optimă a funcției* $\min_{\theta} V(\theta)$ și valorile $\theta^* = \arg \min_{\theta} V(\theta)$ ale variabilelor pentru care minimul este atins



De notat că în cazul regresiei liniare, folosim notația $\hat{\theta}$; vectorul $\hat{\theta}$ este soluția reală a problemei de optimizare dat fiind setul de date, dar rămâne totuși o estimare datorită zgomotului din date

Soluția formală a problemei de regresie

Pentru a deriva ușor soluția, pornim de la:

$$Y = \Phi\theta$$

și înmulțim la stânga cu $\Phi^\top$:

$$\Phi^\top Y = \Phi^\top \Phi \theta$$

O alta înmulțire la stânga, cu inversa matricii $\Phi^\top \Phi$, duce la:

$$\hat{\theta} = (\Phi^\top \Phi)^{-1} \Phi^\top Y$$

Observații:

- O metodă mai bună de a-l găsi pe $\hat{\theta}$ este să scriem problema de optimizare (minimizare MSE), și să o rezolvăm egalând derivata matriceală cu 0.
- Matricea $\Phi^\top \Phi$ trebuie să fie inversabilă, ceea ce necesită o alegere bună a modelului (ordin n , regresori φ), și folosirea unui set informativ de date.
- Valoarea optimă a funcției obiectiv este $V(\hat{\theta}) = \frac{1}{2}[Y^\top Y - Y^\top \Phi(\Phi^\top \Phi)^{-1} \Phi^\top Y]$.
- $(\Phi^\top \Phi)^{-1} \Phi^\top$ este pseudoinversa matricii Φ .

Expresie alternativă

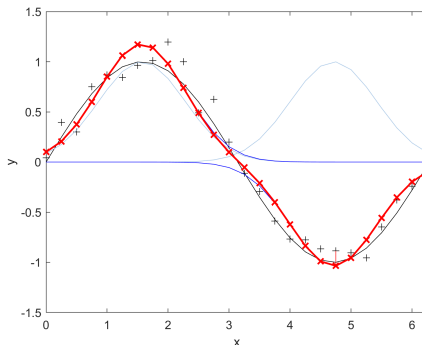
$$\Phi^T \Phi = \sum_{k=1}^N \varphi(k) \varphi^T(k), \Phi^T Y = \sum_{k=1}^N \varphi(k) y(k)$$

Soluția poate fi scrisă:

$$\hat{\theta} = \left[\sum_{k=1}^N \varphi(k) \varphi^T(k) \right]^{-1} \left[\sum_{k=1}^N \varphi(k) y(k) \right]$$

Avantaj: matricea Φ cu dimensiunile $N \times n$ nu mai trebuie calculată; este nevoie doar de matrici și vectori mai mici, de dimensiuni $n \times n$ respectiv n .

Soluția problemei de regresie: Exemplu recurent



Parametrii θ_1 , θ_2 **scalează & reflectă** RBFurile în așa fel încât **soluția aproximată** minimizează suma erorilor pătratice între datele reale și cele approximate.

Rezolvarea sistemului liniar în practică

În practică, ambele metode bazate pe inversarea de matrici se comportă prost din punct de vedere numeric. Există algoritmi mai buni, cum ar fi triangularizarea ortogonală.

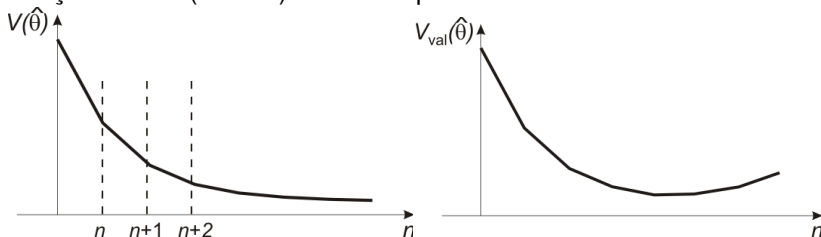
În majoritatea cazurilor, **MATLAB** alege automat un algoritm potrivit. Dacă Φ este stocată în variabila `PHI` și Y în `Y`, comanda care rezolvă sistemul de ecuații în sensul CMMP este împărțirea matriceală la stânga (backslash):

```
theta = PHI \ Y;
```

Dacă se dorește un control mai detaliat al algoritmului, se poate folosi funcția `linsolve` în loc de `\`.

Alegerea modelului

Considerăm că dată fiind o complexitate a modelului (număr de parametri) n , putem genera regresori $\varphi(k)$ care fac modelul mai expresiv (de ex., funcții de bază pe o grilă mai fină). Ne așteptăm ca funcția obiectiv (CMMP) să se comporte în următorul fel:



Observație: Dacă datele sunt afectate de zgomot, creșterea exagerată a lui n va duce la **supraantrenare**: performanțe bune pe datele de identificare, dar proaste pe date diferite. **Validarea pe un set separat de date** este esențială în practică!

Putem așadar crește treptat valoarea lui n până când eroarea V_{val} pe datele de validare începe să crească.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice**
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Exemplu analitic: Estimarea unui scalar

Model:

$$y(k) = b = 1 \cdot b = \varphi(k)\theta$$

unde $\varphi(k) = 1 \forall k$, $\theta = b$.

Pentru N date:

$$y(1) = \varphi(1)\theta = 1 \cdot b$$

...

$$y(N) = \varphi(N)\theta = 1 \cdot b$$

În formă matriceală:

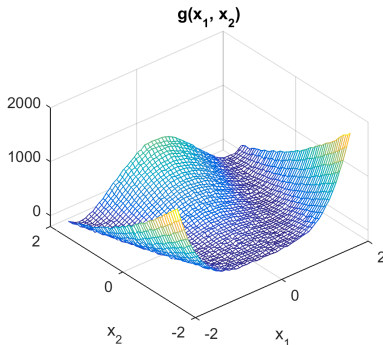
$$\begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \theta$$
$$Y = \Phi\theta$$

Exemplu analitic: Estimarea unui scalar (continuare)

$$\begin{aligned}\hat{\theta} &= (\Phi^T \Phi)^{-1} \Phi^T Y \\ &= \left(\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \\ &= N^{-1} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \\ &= \frac{1}{N} (y(1) + \dots + y(N))\end{aligned}$$

Intuiție: Estimarea este media tuturor măsurătorilor, filtrând zgomotul.

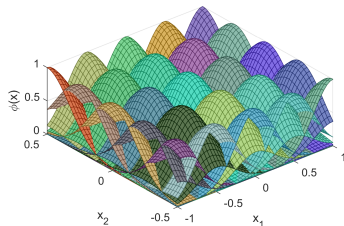
Exemplu: Aproximarea funcției lui Rosenbrock



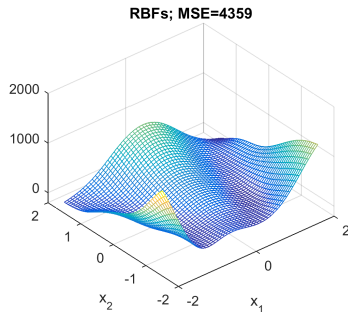
- Funcția Rosenbrock: $g(x_1, x_2) = (1 - x_1)^2 + 100[(x_2 + 1.5) - x_1^2]^2$ (necunoscută de algoritm).
- **Date de identificare:** 200 puncte de intrare (x_1, x_2) , distribuite aleator în spațiul $[-2, 2] \times [-2, 2]$; și ieșirile corespunzătoare $y = g(x_1, x_2)$, **afectate de zgomot**.
- **Date de validare:** grilă uniformă cu 31×31 puncte în $[-2, 2] \times [-2, 2]$ cu ieșirile corespunzătoare (afectate de zgomot).

Funcția Rosenbrock: Funcții de bază radiale

Reamintim funcțiile de bază radiale:



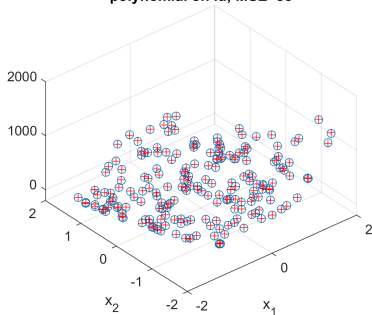
Rezultate cu 6×6 RBF-uri, cu centre pe o grilă echidistantă și lățimea egală cu distanța între centre:



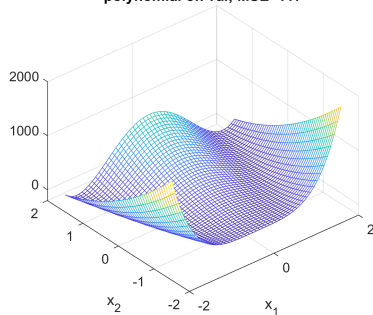
Funcția Rosenbrock: Aproximare polinomială

Polinom de gradul 4 în cele două intrări (15 parametri):

polynomial on id; MSE=88



polynomial on val; MSE=117

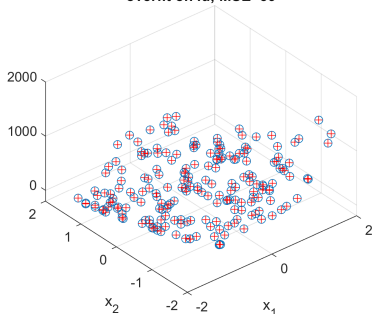


Funcția Rosenbrock: Exemplu de supraantrenare

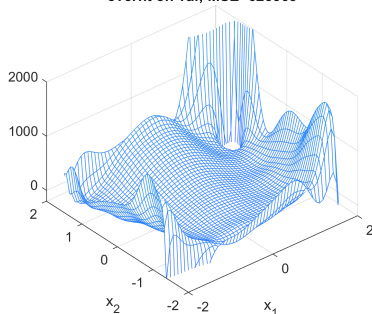
Din nou polinom, dar de data aceasta de gradul 13.

MSE pe datele de identificare este mai mic decât pentru gradul 4 mai sus. Pe datele de validare însă:

overfit on id; MSE=39



overfit on val; MSE=625363



Rezultat foarte prost! Gradul prea mare duce la supraantrenare.

Sumar regresie liniară

- Modele liniare pentru aproximarea funcțiilor și a seriilor temporale; regresori și parametri.
- Funcții de bază polinomiale și radiale.
- Scrierea problemei ca un sistem de ecuații liniare, obiectiv matematic, și soluție (teoretică și practică).
- Alegerea modelului și supraantrenarea.
- Exemple: estimarea unui scalar și funcția “banană”.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare

Variabilă aleatoare discretă

Fie un set $\mathcal{X}$ conținând valori posibile x . O variabilă aleatoare X corespunzând acestui set este descrisă de funcția de frecvență:

Definiție

Funcția de frecvență a variabilei X este lista probabilităților tuturor valorilor individuale $p(x_0), p(x_1), \dots$. Probabilitățile trebuie să fie pozitive: $p(x) \geq 0 \forall x$, și suma lor trebuie să fie 1:

$$P(X \in \mathcal{X}) = \sum_{x \in \mathcal{X}} p(x) = 1.$$

Exemplu: Numărul obținut ca urmare a aruncării unui zar este o variabilă aleatoare discretă, cu șase valori posibile: $\mathcal{X} = \{1, 2, \dots, 6\}$. Pentru un zar corect, funcția de frecvență este $p(x) = 1/6$ pentru toate valorile x , de la 1 la 6.

Observație: Setul $\mathcal{X}$ poate conține un număr finit sau infinit de elemente, dar în al doilea caz setul trebuie să fie numărabil (Intuitiv: valorile se pot enumera, sau pot fi asociate numerelor naturale $0, 1, 2, \dots$).

Motivație pentru variabile aleatoare continue

Dorim să caracterizăm unghiul la care se va opri roata unei rulete, ca o fracțiune dintr-o rotație completă: o variabilă continuă $x \in [0, 1]$. În acest caz, $\mathcal{X} = [0, 1]$. Pentru o ruletă corectă, fiecare valoare x are aceeași probabilitate.

Vom încerca întâi să definim o funcție de frecvență; aceasta trebuie să aibă aceeași valoare peste tot. Notăm această valoare cu c ; cum există o infinitate de valori ale lui x în orice subinterval de lungime finită din $[0, 1]$, dacă c este nenulă, probabilitatea oricărui astfel de interval este infinită! Deci singura valoare posibilă pentru $p(x) = c$ este 0, care nu are nicio utilitate practică; nu putem așadar folosi funcția de frecvență pentru a descrie acest caz.

Putem defini probabilități doar pentru intervale.

Variabilă aleatoare continuă

Fie un interval $\mathcal{X} \subseteq \mathbb{R}$ de numere reale, cu x o valoare individuală. O variabilă aleatoare X corespunzând acestui set este descrisă de densitatea de probabilitate:

Definiție

Densitatea de probabilitate a unei variabile aleatoare continue X este o funcție $f : \mathcal{X} \rightarrow [0, \infty)$, astfel încât probabilitatea de a obține o valoare într-un interval $[a, b] \subseteq \mathcal{X}$ este:

$$P(X \in [a, b]) = \int_a^b f(x) dx$$

Densitatea trebuie să satisfacă $P(X \in \mathcal{X}) = \int_{\mathcal{X}} f(x) dx = 1$.

Observație: Intervalul $\mathcal{X}$ poate fi și setul complet al numerelor reale.

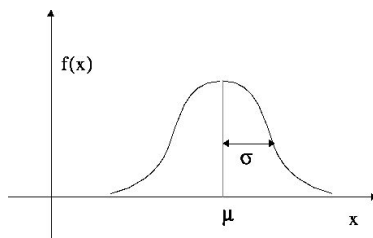
Exemplu 1: Densitate uniformă pentru ruletă

Să ne întoarcem la exemplul cu ruleta. Frațiune dintr-o rotație completă: $x \in [0, 1]$ și $\mathcal{X} = [0, 1]$.

Pentru a obține probabilități uniforme ca să reprezentăm o ruletă corectă, dorim ca $P(X \in [a, b]) = b - a$, pentru care trebuie ca $f(x) = 1$.

Exemplu 2: Densitate Gaussiană

Are formă similară cu funcțiile de bază Gaussiene, dar semnificație diferită.



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Parametri: media μ și varianța σ^2 (vor fi explicați mai târziu)

Distribuția Gaussiană intervine adeseori în natură: de ex., distribuția IQ-urilor într-o populație umană. Este numită de aceea și distribuția normală, și se notează $\mathcal{N}(\mu, \sigma^2)$.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare**
- 7 Vectori și secvențe aleatoare

Valoarea medie

Definiție

$$E\{X\} = \mu_X = \begin{cases} \sum_{x \in \mathcal{X}} p(x)x & \text{pentru variabile aleatoare discrete} \\ \int_{\mathcal{X}} f(x)x & \text{pentru variabile aleatoare continue} \end{cases}$$

Intuiție: media tuturor valorilor, ponderate de probabilitatea lor; valoarea “așteptată” în avans, dată fiind distribuția de probabilitate.

Valoarea medie se mai numește și *valoare așteptată* sau *speranță*.

Exemple:

- Pentru un zar unde X este numărul feței obținute,
 $E\{X\} = \frac{1}{6}1 + \frac{1}{6}2 + \dots + \frac{1}{6}6 = 7/2$.
- Dacă X are densitate Gaussiană $\mathcal{N}(\mu, \sigma^2)$, atunci $E\{X\} = \mu$. De notat constanta μ versus valoarea medie μ_X a variabilei X .

Valoarea medie a unei funcții

Considerăm o funcție $g : \mathcal{X} \rightarrow \mathbb{R}$ care depinde de variabila aleatoare X . Atunci, $g(X)$ este o și ea o variabilă aleatoare, cu valoarea medie:

$$E \{g(X)\} = \begin{cases} \sum_{x \in \mathcal{X}} p(x)g(x) & \text{discret} \\ \int_{\mathcal{X}} f(x)g(x) & \text{continuu} \end{cases}$$

Exemplu: Considerăm un joc cu un zar în care fața cu 6 puncte câștigă 10 lei, și celelalte fețe nu câștigă nimic. Avem $g(6) = 10$ și $g(x) = 0$ pentru toate celelalte valori ale lui x . Valoarea medie a acestui joc este $\frac{1}{6}0 + \dots + \frac{1}{6}0 + \frac{1}{6}10 = 10/6$ lei.

Varianța

Definiție

$$\text{Var} \{X\} = \sigma_X^2 = \mathbb{E} \{(X - \mu_X)^2\} = \mathbb{E} \{X^2\} - \mu_X^2$$

Intuiție: cât de “răspândite” sunt valorile aleatoare în jurul valorii medii.

$$\begin{aligned} \text{Var} \{X\} &= \begin{cases} \sum_{x \in \mathcal{X}} p(x)(x - \mu_X)^2 & \text{discret} \\ \int_{\mathcal{X}} f(x)(x - \mu_X^2) & \text{continuu} \end{cases} \\ &= \begin{cases} \sum_{x \in \mathcal{X}} p(x)(x^2 - \mu_X^2) & \text{discret} \\ \int_{\mathcal{X}} f(x)(x^2 - \mu_X^2) & \text{continuu} \end{cases} \end{aligned}$$

Cantitatea $\sigma_X = \sqrt{\text{Var} \{X\}}$ se numește *abaterea standard*.

Exemple:

- Pentru un zar, $\text{Var} \{X\} = \frac{1}{6}1^2 + \frac{1}{6}2^2 + \dots + \frac{1}{6}6^2 - (7/2)^2 = 35/12$.
- Dacă X are densitate Gaussiană $\mathcal{N}(\mu, \sigma^2)$, atunci $\text{Var} \{X\} = \sigma^2$.

Probabilitate: Independență

Definiție

Două variabile aleatoare X și Y sunt **independente** dacă:

- în cazul continuu, $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$.
- în cazul discret, $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$.

unde $f_{X,Y}$ este densitatea comună a vectorului (X, Y) , f_X și f_Y sunt densitățile pentru X și Y , și la fel pentru funcțiile de frecvență p .

Exemple:

- Evenimentul de a arunca 6 cu un zar este independent de evenimentul 6 la aruncarea anterioară (de fapt, de oricare altă valoare la orice aruncare anterioară).
- Evenimentul de a arunca două valori 6 *consecutive* nu este independent de aruncarea anterioară!

(Primul fapt este contra-intuitiv și multă lume nu îl înțelege, ducând la așa-numita *gambler's fallacy*. O secvență mai lungă de jocuri norocoase sau proaste nu are nici absolut nici o influență asupra jocului următor!)

Covarianță

$$\text{Var} \{X\} = E \{(X - \mu_X)^2\} = E \{(X - \mu_X)(X - \mu_X)\}$$

Înlocuim X cu Y în a doua paranteză.

Definiție

$$\text{Cov} \{X, Y\} = E \{(X - \mu_X)(Y - \mu_Y)\} = E \{(X - E \{X\})(Y - E \{Y\})\}$$

Intuiție: cât de “aliniată” sunt schimbările celor două variabile (covarianță pozitivă dacă variabilele se schimbă în direcții similare, negativă dacă se schimbă în direcții opuse).

Observație: $\text{Var} \{X\} = \text{Cov} \{X, X\}$.

Variabile necorelate

Definiție

Variabilele aleatoare X și Y sunt **necorelate** dacă $\text{Cov} \{X, Y\} = 0$. Altfel, ele se numesc **corelate**.

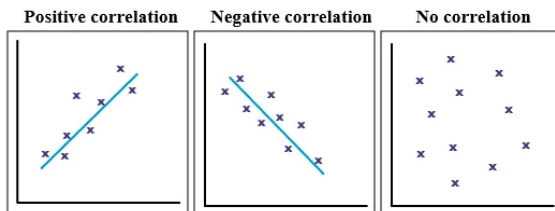
Exemple:

- Nivelul de educație al unei persoane este corelat cu venitul său.
- Culoarea părului este necorelată cu venitul (sau ar trebui să fie, în cazul ideal).

Observații:

- Dacă X și Y sunt independente, atunci sunt și necorelate.
- Dar nu și invers! Putem avea variabile necorelate care sunt totuși dependente.

Intuiție despre covarianță



Axele X și Y ale fiecărui grafic sunt cele două variabile aleatoare studiate.

- Corelate pozitiv (covarianță > 0): când X crește, Y crește.
- Corelate negativ (covarianță < 0): când X crește, Y scade.
- Necorelate (covarianță $= 0$): nici o relație.

Conținut

- 1 Problema de regresie și exemple motivante
- 2 Exemple de regresori
- 3 Problema CMMP și soluția ei
- 4 Exemple analitice și numerice
- 5 Variabile aleatoare discrete și continue
- 6 Proprietăți ale variabilelor aleatoare
- 7 Vectori și secvențe aleatoare**

Vectori de variabile aleatoare

Considerăm vectorul $\mathbf{X} = [X_1, \dots, X_N]^\top$ unde fiecare X_i este o variabilă aleatoare cu valori reale continue. Acest vector are o *funcție de densitate comună* $f(\mathbf{x})$, cu $\mathbf{x} \in \mathbb{R}^N$.

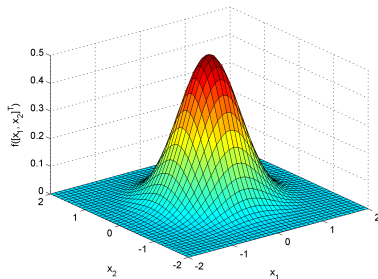
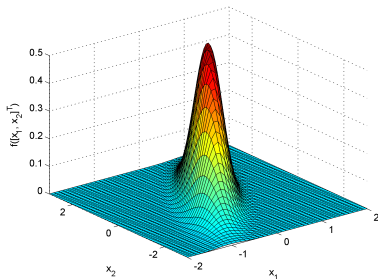
Definiții

Valoarea medie și matricea de covarianță a lui $\mathbf{X}$:

$$\begin{aligned} E\{\mathbf{X}\} &= \boldsymbol{\mu}_X := [E\{X_1\}, \dots, E\{X_N\}]^\top = [\mu_1, \dots, \mu_N]^\top \in \mathbb{R}^N \\ \text{Cov}\{\mathbf{X}\} &= \Sigma_X := \begin{bmatrix} \text{Cov}\{X_1, X_1\} & \text{Cov}\{X_1, X_2\} & \cdots & \text{Cov}\{X_1, X_N\} \\ \text{Cov}\{X_2, X_1\} & \text{Cov}\{X_2, X_2\} & \cdots & \text{Cov}\{X_2, X_N\} \\ \cdots & \cdots & \cdots & \cdots \\ \text{Cov}\{X_N, X_1\} & \text{Cov}\{X_N, X_2\} & \cdots & \text{Cov}\{X_N, X_N\} \end{bmatrix} \\ &= E\left\{(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^\top\right\} \in \mathbb{R}^{N,N} \end{aligned}$$

Observații: $\text{Cov}\{X_i, X_i\} = \text{Var}\{X_i\}$. De asemenea, $\text{Cov}\{X_i, X_j\} = \text{Cov}\{X_j, X_i\}$, deci matricea Σ_X este simetrică.

Exemplu: vector Gaussian



Densitatea comună Gaussiană a unui vector $\mathbf{X}$ se poate scrie:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^N \sqrt{\det(\Sigma)}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})^\top\right)$$

unde $\boldsymbol{\mu}$ este vectorul de valori medii și Σ matricea de covarianță (care trebuie să fie pozitiv definită, pentru ca $\det(\Sigma) > 0$ și Σ^{-1} să existe).

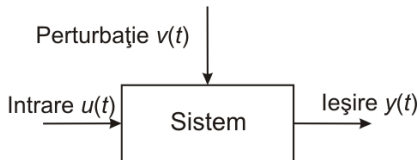
Proces stohastic

Definiție

Un **proces stohastic** $\mathbf{X}$ este o secvență de variabile aleatoare $\mathbf{X} = (X_1, \dots, X_k, \dots, X_N)$.

Avem așadar de-a face tot cu un vector de variabile aleatoare, cu o structură specifică: fiecare index din vector este asociat unui pas discret de timp k .

În identificarea sistemelor, semnalele (intrări, ieșiri, perturbații etc.) vor fi adesea procese stohastice.



Zgomot alb de medie zero

Definiție

Un proces stohastic $\mathbf{X}$ este **zgomot alb de medie zero** dacă:

$\forall k, E\{X_k\} = 0$ (medie zero), și $\forall k, k' \neq k, \text{Cov}\{X_k, X_{k'}\} = 0$ (valorile la pași diferiți de timp sunt necorelate). În plus, varianța $\text{Var}\{X_k\}$ trebuie să fie finită $\forall k$.

Cu notație vectorială, aceste proprietăți se pot scrie compact: media $\mu_{\mathbf{X}} = E\{\mathbf{X}\} = \mathbf{0} \in \mathbb{R}^N$ și matricea de covarianță $\Sigma_{\mathbf{X}} = \text{Cov}\{\mathbf{X}\}$ este diagonală (cu diagonala formată din numere finite și pozitive).

În identificarea sistemelor, măsurătorile sunt adesea afectate de zgomote, și vom presupune câteodată că aceste zgomote sunt albe și de medie zero.

Proces staționar

Valorile unui semnal la diferite momente de timp pot fi corelate (de ex. când semnalul depinde de ieșirea unui sistem dinamic). Vom presupune însă câteodată că semnalele sunt staționare:

Definiție

Procesul stohastic $\mathbf{X}$ este **staționar** dacă $\forall k, E\{X_k\} = \mu$, și $\forall k, k', \tau$, $\text{Cov}\{X_k, X_{k+\tau}\} = \text{Cov}\{X_{k'}, X_{k'+\tau}\}$.

Media este aceeași la fiecare pas, iar covarianța depinde doar de pozițiile relative ale pașilor de timp (și nu de pozițiile lor absolute).

Rezumat probabilități și statistică

- Variabile aleatoare (VA) discrete și funcția de frecvență.
- VA continue și densitatea de probabilitate.
- Valoarea medie și varianța unei VA.
- Independența, covarianța, și corelația a două VA.
- Vector de VA, cu valoare medie și matrice de covarianță.
- Proces stohastic, zgomot alb de medie zero, și proces staționar.