

Identificarea sistemelor

Ingineria sistemelor, anul 3
Universitatea Tehnică din Cluj-Napoca

Lucian Buşoniu



Partea III

Baze matematice: Regresie liniară și statistică

Motivare

Până acum, am discutat analiza în domeniul timp a răspunsurilor la treaptă, folosind concepte cunoscute din teoria sistemelor liniare.

Majoritatea metodelor de identificare care urmează necesită elemente noi: **regresia liniară** și concepte de **teoria probabilităților și statistică**. Le vom discuta aici.

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

Conținut

- 1 Regresia liniară
- 2 Concepte de teoria probabilităților și statistică

Conținut

- 1 Regresia liniară
 - Problema de regresie și exemple motivante
 - Exemple de regresori
 - Problema CMMP și soluția ei
 - Exemple analitice și numerice
- 2 Concepte de teoria probabilităților și statistică

Problema de regresie

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* $\varphi_i(k)$, $i = 1, \dots, n$, $\varphi(k) = [\varphi_1(k), \varphi_2(k) \dots, \varphi_n(k)]^T$.
- Un *vector de parametri* $\theta \in \mathbb{R}^n$ **necunoscut**.

Obiectiv: identificarea comportamentului variabilei dependente din date, folosind modelul liniar:

$$y(k) = \varphi^T(k)\theta$$

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

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: Exemplul 1

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: Exemplul 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.

Conținut

- 1 **Regresia liniară**
 - Problema de regresie și exemple motivante
 - **Exemple de regresori**
 - Problema CMMP și soluția ei
 - Exemple analitice și numerice
- 2 Concepte de teoria probabilităților și statistică

Exemplu regresori 1: Polinom în k

Util pentru modelarea seriilor temporale.

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

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:

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

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

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

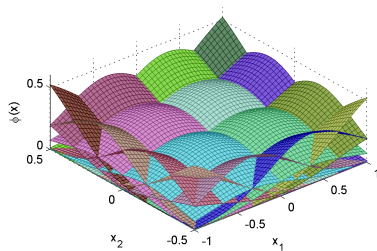
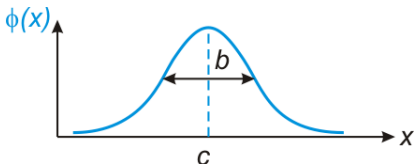
 **Conexiune:** Proiect partea 1

Exemplu regresori 3: Funcții de bază Gaussiene

Utile pentru aproximarea funcțiilor:

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

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



Conținut

- 1 Regresia liniară
 - Problema de regresie și exemple motivante
 - Exemple de regresori
 - Problema CMMP și soluția ei
 - Exemple analitice și numerice
- 2 Concepte de teoria probabilităților și statistică

Sistem liniar

Scrind 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

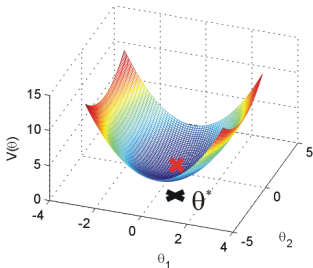
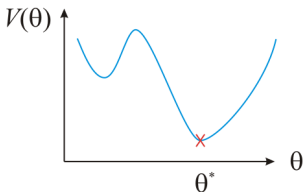
Găsește vectorul de parametri $\hat{\theta}$ care minimizează funcția obiectiv:

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

Paranteză: Problema de optimizare

Dată 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 Φ^\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:

- 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 .

Rezolvarea sistemului liniar

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.

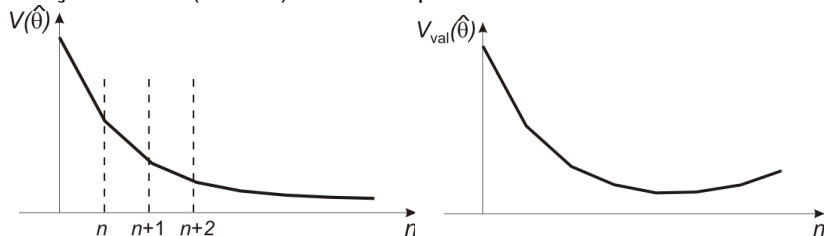
În majoritatea cazurilor, **MATLAB** alege automat un algoritm potrivit. Dacă Φ este stocată în variabila `PHI` și Y în `Y`, comanda care rezolvă sistemul de ecuații în sensul CMMP este împărțirea matriceală la stânga (backslash):

```
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 Regresia liniară
 - Problema de regresie și exemple motivante
 - Exemple de regresori
 - Problema CMMP și soluția ei
 - Exemple analitice și numerice
- 2 Concepte de teoria probabilităților și statistică

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:

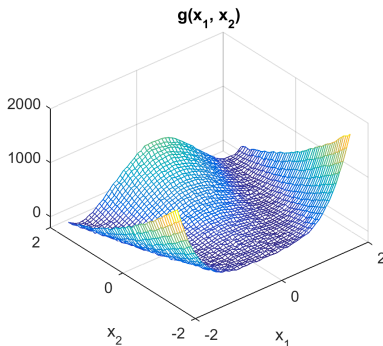
$$\begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \theta$$
$$Y = \Phi\theta$$

Exemplu analitic: Estimarea unui scalar (continuare)

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

Intuiție: Estimarea este media tuturor măsurătorilor, filtrând zgomotul.

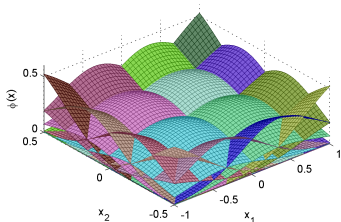
Exemplu: Aproximarea funcției lui Rosenbrock



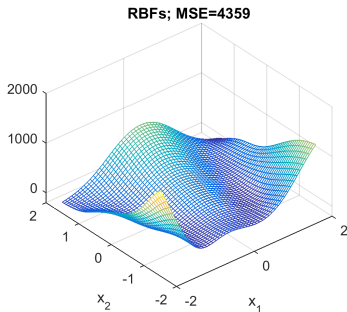
- Funcția Rosenbrock: $g(x_1, x_2) = (1 - x_1)^2 + 100[(x_2 + 1.5) - x_1^2]^2$ (necunoscută de algoritm).
- **Date de identificare:** 200 puncte de intrare (x_1, x_2) , distribuite aleator în spațiul $[-2, 2] \times [-2, 2]$; și ieșirile corespunzătoare $y = g(x_1, x_2)$, **afectate de zgomot**.
- **Date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



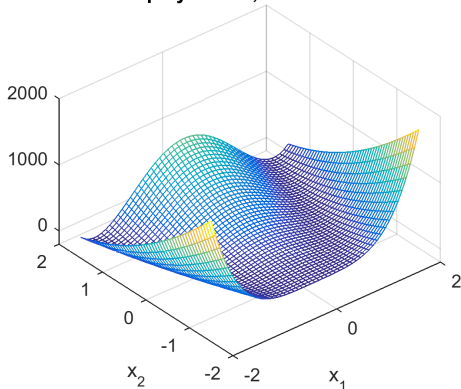
Rezultate cu 6×6 RBF-uri, cu centre pe o grilă echidistantă și lățimea egală cu distanța între centre:



Funcția Rosenbrock: Aproximare polinomială

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

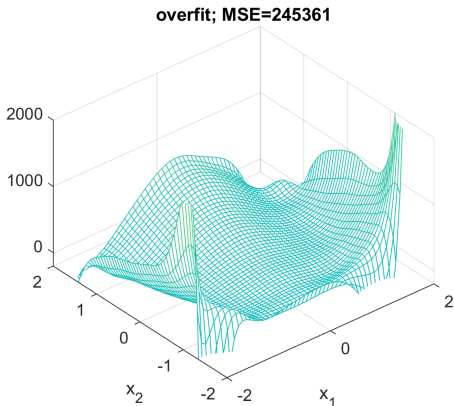
polynomial; MSE=110



Funcția Rosenbrock: Exemplet de supraantrenare

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

MSE pe datele de identificare este 55.32 (comparat cu 87.63 pentru gradul 4 mai sus). Pe datele de validare însă:



Rezultat foarte prost! Gradul prea mare duce la supraantrenare.

Conținut

- 1 Regresia liniară
- 2 **Concepte de teoria probabilităților și statistică**
 - Variabile aleatoare discrete și continue
 - Proprietăți ale variabilelor aleatoare
 - 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 $p(x_0), p(x_1), \dots$. Suma probabilităților 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 $0, 1, 2, \dots$).

Variabilă aleatoare continuă

Fie acum 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

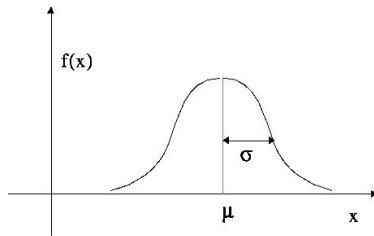
Dorim să caracterizăm unghiul la care se va opri roata unei rulete, ca o fracțiune dintr-o rotație completă: $x \in [0, 1]$. În acest caz, $\mathcal{X} = [0, 1]$. Pentru o ruletă corectă, fiecare valoare x are aceeași probabilitate.

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, și de aceea avem nevoie de densitatea f . Mai exact, pentru a obține probabilități uniforme, dorim ca $P(X \in [a, b]) = b - a$, ceea ce înseamnă că $f(x) = 1$.

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 Regresia liniară
- 2 **Concepte de teoria probabilităților și statistică**
 - Variabile aleatoare discrete și continue
 - **Proprietăți ale variabilelor aleatoare**
 - Vectori și secvențe aleatoare

Valoarea medie

Definiție

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

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 $E\{X\} = \mu.$

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:

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

Varianța

Definiție

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

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

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

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 $\text{Var} \{X\} = \sigma^2$.

Notăție

Vom nota generic $E\{X\} = \mu$ și $\text{Var}\{X\} = \sigma^2$.

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

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!)

Covarianța

Definiție

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

unde μ_X, μ_Y sunt valorile medii ale celor două variabile.

Intuiție: cât de “aliniate” sunt schimbările celor două variabile (covarianța 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.

Conținut

- 1 Regresia liniară
- 2 **Concepte de teoria probabilităților și statistică**
 - Variabile aleatoare discrete și continue
 - Proprietăți ale variabilelor aleatoare
 - **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} :

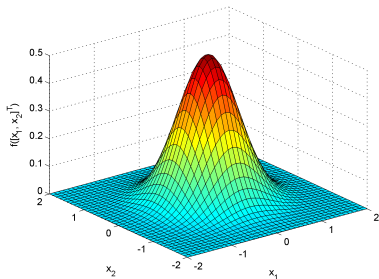
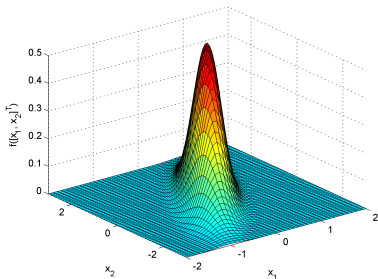
$$E \{ \mathbf{X} \} := [E \{ X_1 \}, \dots, E \{ X_N \}]^\top = [\mu_1, \dots, \mu_N]^\top, \text{ notată } \boldsymbol{\mu} \in \mathbb{R}^N$$

$$\text{Cov} \{ \mathbf{X} \} := \begin{bmatrix} \text{Cov} \{ X_1, X_1 \} & \text{Cov} \{ X_1, X_2 \} & \dots & \text{Cov} \{ X_1, X_N \} \\ \text{Cov} \{ X_2, X_1 \} & \text{Cov} \{ X_2, X_2 \} & \dots & \text{Cov} \{ X_2, X_N \} \\ \dots & \dots & \dots & \dots \\ \text{Cov} \{ X_N, X_1 \} & \text{Cov} \{ X_N, X_2 \} & \dots & \text{Cov} \{ X_N, X_N \} \end{bmatrix}$$

$$= E \left\{ (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top \right\}, \text{ notată } \Sigma \in \mathbb{R}^{N,N}$$

Observații: $\text{Cov} \{ X_i, X_j \} = \text{Var} \{ X_j \}$. De asemenea, $\text{Cov} \{ X_i, X_j \} = \text{Cov} \{ X_j, X_i \}$, deci matricea Σ 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).

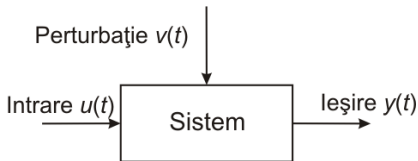
Proces stohastic

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 $\mu = E \{\mathbf{X}\} = \mathbf{0} \in \mathbb{R}^N$ și matricea de covarianță $\Sigma = \text{Cov} \{\mathbf{X}\}$ 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 X este **staționar** dacă $\forall k, E\{X_k\} = \mu$, și $\forall k, k', \tau$, $Cov\{X_k, X_{k+\tau}\} = Cov\{X_{k'}, X_{k'+\tau}\}$.

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