Calcul numeric

Cele mai mici pătrate. Regresie liniară.

Paul Irofti
Andrei Pătrașcu
Cristian Rusu

Departmentul de Informatică
Facultatea de Matematică și Informatică
Universitatea din București
Email: `prenume.nume@fmi.unibuc.ro`

Recapitulare: Sisteme de ecuații liniare

Fie sistemul

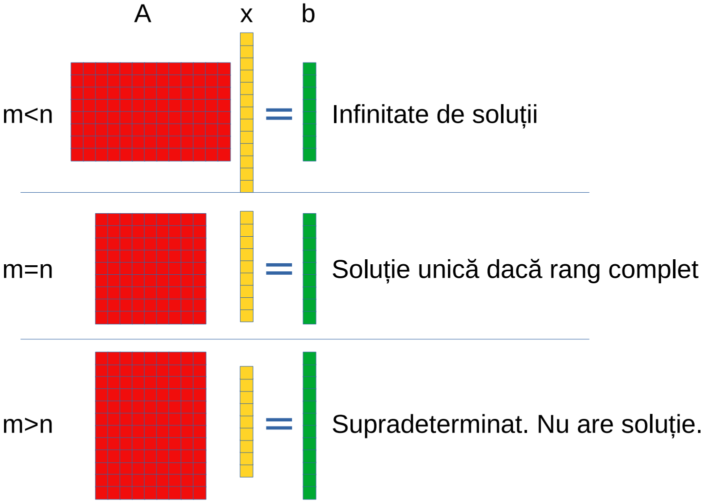
$$Ax = b$$

unde $A \in \mathbb{R}^{m \times n}$ și $b \in \mathbb{R}^n$ sunt date. Atunci sistemul este:

- Subdeterminat: $m < n$, posibil o infinitate de soluții
- Determinat: $m = n$, adesea soluție unică
- Supradeterminat: $m > n$, adesea nu are soluție

Teoremă. Sistemul $Ax = b$ are soluție dacă și numai dacă $b \in \mathrm{Im}(A)$.

Recapitulare: Sisteme de ecuații liniare



Recapitulare: Norme. Produs Scalar

Produs scalar euclidian: $\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$

- dacă x, y au norma unitate atunci $x^T y = \cos(\theta)$, unde θ este unghiul format de x si y
- în felul acesta, capătă sensul unei măsuri de similaritate (în opoziție cu distanța)

Norme $\|\cdot\|$: funcții care satisfac următoarele condiții

- pozitivitate: $\|x\| > 0, \forall x \in \mathbb{R}^n, x \neq 0$
- omogenitate: $\|\alpha x\| = |\alpha| \, \|x\|, \forall x \in \mathbb{R}^n, \alpha \in \mathbb{R}$
- inegalitatea triunghiului: $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathbb{R}^n$

Cazul supradeterminat

Sistemele supradeterminate ($m > n$) $\boldsymbol{Ax} = \boldsymbol{b}$ adesea nu au soluție pentru că $\boldsymbol{b} \notin \mathrm{Im}(A)$.

Scop: Minimizarea erorii de aproximare $\|\boldsymbol{Ax} - \boldsymbol{b}\|_p$.
Alegeri comune sunt $p = \{1, 2, \infty\}$ notate ℓ_1, ℓ_2, respectiv ℓ_∞.

Problema ℓ_2 este cea mai rapid de rezolvat:

$$\min_{x \in \mathbb{R}^n} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2 \tag{1}$$

denumită problema celor mai mici pătrate (sau *least squares*).

Soluții:

- minimizare $\phi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2$ cu ajutorul gradientului $\nabla\phi(\boldsymbol{x}) = 0$
- factorizare ortogonală $\boldsymbol{A} = \boldsymbol{QR}$ a.î. $\left\|(\boldsymbol{Q}^T\boldsymbol{A})\boldsymbol{x} - \boldsymbol{Q}^T\boldsymbol{b}\right\|_2$ se rezelovă ușor

Ecuațiile normale

Minimizare cu ajutorul gradientului:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \phi(\boldsymbol{x}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{2}$$

Rezolvăm ecuația $\nabla \phi(\boldsymbol{x}) = 0$:

$$\begin{aligned} \nabla \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 =& \nabla \frac{1}{2} \left(\sum_{i=1}^{m} \left[(\boldsymbol{a}^i \boldsymbol{x}_i)^2 - 2\boldsymbol{a}^i \boldsymbol{x}_i \boldsymbol{b}_i + \boldsymbol{b}_i^2 \right] \right) \\ =& \frac{1}{2} \sum_{i=1}^{m} \left[2\boldsymbol{a}^{i2} \boldsymbol{x}_i - 2\boldsymbol{a}^i \boldsymbol{b}_i \right] = \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^T \boldsymbol{b} \end{aligned}$$

Acestea sunt denumite ecuațiile normale:

$$\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T \boldsymbol{b} \tag{3}$$

unde

$$r_{LS} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{LS} \ ; \ \rho_{LS} = \|\boldsymbol{A}\boldsymbol{x}_{LS} - \boldsymbol{b}\|_2 \tag{4}$$

reprezintă reziduul minim, respectiv, mărimea reziduului minim.

Rezolvare folosind ecuațiile normale

Algorithm 1: Ecuații normale CMMP

Data: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ a.î. $\text{rang}(A) = n$ și $\boldsymbol{b} \in \mathbb{R}^m$
Result: soluția CMMP $\boldsymbol{x_{LS}} \in \mathbb{R}^n$

Calculează triunghiul inferior al matricei $\boldsymbol{C} = \boldsymbol{A}^T \boldsymbol{A}$
Actualizează vectorul țintă $\boldsymbol{d} = \boldsymbol{A}^T \boldsymbol{b}$
Calculează factorizarea Cholesky $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{G}^T$
Rezolvă folosind LTRIS $\boldsymbol{G}\boldsymbol{y} = \boldsymbol{d}$
Rezolvă folosind UTRIS $\boldsymbol{G^T x_{LS}} = \boldsymbol{y}$

Rezolvare folosind ecuațiile normale – complexitate

Algorithm 2: Ecuații normale CMMP

Data: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ a.î. rang$(A) = n$ și $\boldsymbol{b} \in \mathbb{R}^m$
Result: soluția CMMP $\boldsymbol{x}_{LS} \in \mathbb{R}^n$

Calculează triunghiul inferior al matricei $\boldsymbol{C} = \boldsymbol{A}^T \boldsymbol{A}$ $\quad O(\frac{mn^2}{2})$
Actualizează vectorul țintă $\boldsymbol{d} = \boldsymbol{A}^T \boldsymbol{b}$ $\quad O(mn)$
Calculează factorizarea Cholesky $\boldsymbol{C} = \boldsymbol{G}\boldsymbol{G}^T$ $\quad O(n^2)$
Rezolvă folosind LTRIS $\boldsymbol{G}\boldsymbol{y} = \boldsymbol{d}$ $\quad O(n^2)$
Rezolvă folosind UTRIS $\boldsymbol{G}^T \boldsymbol{x}_{LS} = \boldsymbol{y}$ $\quad O(n^2)$

Factorizarea QR

Factorizare ortogonală $\boldsymbol{A} = \boldsymbol{QR}$ a.î. $\left\|(\boldsymbol{Q}^T\boldsymbol{A})\boldsymbol{x} - \boldsymbol{Q}^T\boldsymbol{b}\right\|_2$ Pentru a obține matricea superior triughiulară $\boldsymbol{R} \in \mathbb{R}^{m\times n}$ se folosesc:

- rotații Givens
- reflector Householder
- eliminare Gaussiană (factorizare Doolittle)?

Aceste operații conduc la construcția matricei ortogonale $\boldsymbol{Q}$.

Remarcă

O matrice $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ este ortogonală dacă $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}_m$. Matricele ortogonale conservă norma ℓ_2!

$$\boldsymbol{Q}^T\boldsymbol{A} = \boldsymbol{R} = \begin{array}{c} \begin{bmatrix} \boldsymbol{R}_1 \\ 0 \end{bmatrix} \end{array} \begin{array}{c} n \\ m-n \end{array} \;;\; \boldsymbol{Q}^T\boldsymbol{b} = \begin{array}{c} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{d} \end{bmatrix} \end{array} \begin{array}{c} n \\ m-n \end{array} \qquad (5)$$

CMMP cu factorizare QR

Problema CMMP se reduce la

$$\min_{x\in\mathbb{R}^n} \|\boldsymbol{Ax}-\boldsymbol{b}\|_2^2 = \min_{x\in\mathbb{R}^n} \left\|\boldsymbol{Q}^T\boldsymbol{Ax}-\boldsymbol{Q}^T\boldsymbol{b}\right\|_2^2 = \min_{x\in\mathbb{R}^n} \|\boldsymbol{R_1x}-\boldsymbol{c}\|_2^2 + \|\boldsymbol{d}\|_2^2 \tag{6}$$

unde $\text{rang}(\boldsymbol{A}) = \text{rang}(\boldsymbol{R}_1) = n$
Problema se reduce la rezolvarea unui sistem superior triunghiular:

$$\boldsymbol{R}_1\boldsymbol{x}_{LS} = \boldsymbol{c}\ ;\ \rho_{LS} = \|\boldsymbol{d}\|_2 \tag{7}$$

Algorithm 3: CMMP cu factorizare QR

Data: $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ a.î. $\text{rang}(A) = n$ și $\boldsymbol{b} \in \mathbb{R}^m$
Result: soluția CMMP $\boldsymbol{x}_{LS} \in \mathbb{R}^n$

Calculează factorizarea $\boldsymbol{A} = \boldsymbol{QR}$
Aplică rotațiile/reflectorii primelor n poziții din $\boldsymbol{b}$
Rezolvă folosind UTRIS $\boldsymbol{R}(1:n, 1:n)\boldsymbol{x_{LS}} = \boldsymbol{b}(1:n)$

CMMP cu factorizare QR – complexitate

Problema CMMP se reduce la

$$\min_{x \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = \min_{x \in \mathbb{R}^n} \left\|\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{Q}^T\boldsymbol{b}\right\|_2^2 = \min_{x \in \mathbb{R}^n} \|\boldsymbol{R_1}\boldsymbol{x} - \boldsymbol{c}\|_2^2 + \|\boldsymbol{d}\|_2^2 \tag{6}$$

unde $\text{rang}(\boldsymbol{A}) = \text{rang}(\boldsymbol{R}_1) = n$
Problema se reduce la rezolvarea unui sistem superior triunghiular:

$$\boldsymbol{R}_1\boldsymbol{x}_{LS} = \boldsymbol{c} \ ; \ \rho_{LS} = \|\boldsymbol{d}\|_2 \tag{7}$$

Algorithm 4: CMMP cu factorizare QR

Data: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ a.î. $\text{rang}(A) = n$ și $\boldsymbol{b} \in \mathbb{R}^m$
Result: soluția CMMP $\boldsymbol{x}_{LS} \in \mathbb{R}^n$

Calculează factorizarea $\boldsymbol{A} = \boldsymbol{QR}$ $O(mn^2)$
Aplică rotațiile/reflectorii primelor n poziții din $\boldsymbol{b}$ $O(n^2)$
Rezolvă folosind UTRIS $\boldsymbol{R}(1:n, 1:n)\boldsymbol{x}_{\boldsymbol{LS}} = \boldsymbol{b}(1:n)$ $O(n^2)$

Regresia liniară

Recapitulare probabilități

Probabilitatea comună:

$$P(X = x_i, Y = y_j) = P(X = x_i \cap Y = y_j) = P(x_i, y_j) = \frac{n_{ij}}{N} \quad (8)$$

unde N reprezintă numărul total de evenimente, iar n_{ij} numărul de evenimente ce implică x_i și y_j.

Probabilitatea condițională:

$$P(X \mid Y) = \frac{|X \cap Y|}{|Y|} = \frac{\frac{|X \cap Y|}{N}}{\frac{|Y|}{N}} = \frac{P(X \cap Y)}{P(Y)} \quad (9)$$

Legea sumei și a produsului:

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y) \; ; \; P(x, y) = P(y \mid x) P(x) \quad (10)$$

Recapitulare probabilități

Expectanță, covarianță, varianță:

$$\mu = E[X] = \sum_{i=1}^{m} x_i p_i \;;\; \sum_{i=1}^{m} p_i = 1 \tag{11}$$

$$\Sigma = Cov(X, Y) = \frac{1}{m}\sum_{i=1}^{m}(x_i - E[X])(y_i - E[Y])^T \tag{12}$$

$$\sigma^2 = Cov(X, X) = Var(X) = E[(X - \mu)^2] \tag{13}$$

$$\tag{14}$$

Distribuție gaussiană:

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{15}$$

Vom nota cu $p(x) = \mathcal{N}(x \mid \mu, \sigma^2)$ sau $X \sim \mathcal{N}(\mu, \sigma^2)$.

Dacă x și y sunt variabile aleatoare gaussiene independente, atunci:

$$p(x + y) = \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2) \tag{16}$$

Estimarea verosimilității maxime

Lega lui Bayes:

$$\underbrace{P(x \mid y)}_{a\ \ posteriori} = \frac{\overbrace{P(y \mid x)}^{verosimilitate} \ \overbrace{P(x)}^{a\ \ priori}}{\underbrace{P(y)}_{dovezi}} \tag{17}$$

Fie setul de m probabilități $P(\boldsymbol{x}_i \mid \beta)$ condiționate de predictorii β. Atunci estimarea verosimilității maxime revine la:

$$\hat{\beta} = \arg\max_{\beta} P(\boldsymbol{X} \mid \beta) = \mathcal{L}_{\boldsymbol{X}}(\beta) = \prod_{i=1}^{m} P(\boldsymbol{x}_i \mid \beta) = \tag{18}$$

$$= \arg\max_{\beta} \log \mathcal{L}_{\boldsymbol{X}}(\beta) = \sum i = 1^m \log P(\boldsymbol{x}_i \mid \beta) = \tag{19}$$

$$= \arg\min_{\beta} - \sum_{i=1}^{m} \log P(\boldsymbol{x}_i \mid \beta) \tag{20}$$

unde $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m; \beta) = P(\boldsymbol{X} \mid \beta)$.
În literatură: *Maximum Likelihood Estimation (MLE)*.

Regresia liniară

Scop: stabilirea unei legături între variabila răspuns și una sau mai multe variabile de intrare (predictori).

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon \tag{21}$$

- regresie liniară simplă – un singur predictor

$$y = \beta_0 + \beta_1 x + \varepsilon = \begin{bmatrix} 1, x \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \varepsilon = \boldsymbol{x}^T \boldsymbol{\beta} + \varepsilon \tag{22}$$

- regresie liniară multiplă – mai mulți predictori

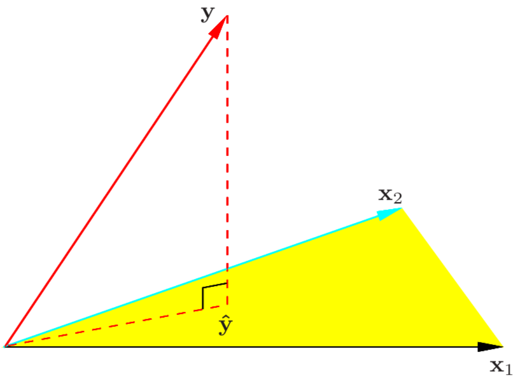
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_m x_m + \varepsilon = \boldsymbol{x}^T \boldsymbol{\beta} + \varepsilon \tag{23}$$

- regresie liniară univariată – răspuns scalar (precum sus)
- regresie liniară multivariată – răspuns vectorial

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots \beta_m x_{i,m} + \varepsilon_i \tag{24}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{e} \tag{25}$$

Proiecția



Perspectiva probabilistică pentru formularea CMMP

Regresia liniară privește (21) dintr-o perspectivă probabilistică:

$$p(y \mid \boldsymbol{x}) = \mathcal{N}(y \mid f(\boldsymbol{x}), \sigma^2) \tag{26}$$

unde $\boldsymbol{x}$ sunt intrările iar y sunt ieșirile perturbate cu un zgomot ε de medie $\mu = 0$ și varianță σ^2 (i.e. $\varepsilon \sim \mathcal{N}(0, \sigma^2)$).

Astfel (23), în care aproximăm $f(\boldsymbol{x})$ cu ajutorul predicotrilor β, devine:

$$p(y \mid \boldsymbol{x}, \beta) = \mathcal{N}(y \mid \boldsymbol{x}^T\beta, \sigma^2) \tag{27}$$

$$\iff \; y = \boldsymbol{x}^T\beta + \varepsilon, \; \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{28}$$

Remarcă

Verosimilitatea lui (28) este PDF-ul funcției y evaluată în punctul $\boldsymbol{x}^T\beta$.

Estimarea parametrilor β

Fie regresia liniară din (28), dat setul de antrenare $\{(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots, (\boldsymbol{x_m}, y_m)\}$ a.î. y_i și y_j sunt independente (între ele) date intrările x_i, respectiv, x_j cu $i \neq j$. Atunci:

$$
\begin{aligned}
p(\mathcal{Y} \mid \mathcal{X}) &= p(y_1, \ldots, y_m | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \beta) && (29) \\
&= \prod_{i=1}^{m} p(y_i \mid \boldsymbol{x}_i, \beta) = \prod_{i=1}^{m} \mathcal{N}(y_i \mid \boldsymbol{x}_i^T \beta, \sigma^2) && (30)
\end{aligned}
$$

Remarcă

Factorii și verosimilitatea lui (30) *reprezintă distribuții Gausiene.*

Remarcă

Fie estimatorii optimi $\beta^\star$ și intrarea de test $\boldsymbol{x}^\star$, atunci distribuția ieșirii $y^\star$ este $\mathcal{N}(y^\star \mid (\boldsymbol{x}^\star)^T \beta^\star)$

Soluția MLE

Estimarea parametriilor β din (30) cu ajutorul estimării verosimilității maxime se face rezolvând problema de optimizare:

$$\beta_{ML} = \arg\max_{\beta} p(\mathcal{Y} \mid \mathcal{X}, \beta) \tag{31}$$

ce poate fi rescrisă drept o problemă de minimizare a sumei:

$$-\log p(\mathcal{Y} \mid \mathcal{X}, \beta) = -\log \prod_{i=1}^{m} p(y_i \mid \boldsymbol{x}_n, \beta) = -\sum_{i=1}^{m} \log p(y_i \mid \boldsymbol{x}_n, \beta) \tag{32}$$

unde termenii reprezintă distribuții gaussiene:

$$\log p(y_i \mid \boldsymbol{x}_n, \beta) = -\frac{1}{2\sigma^2}\left(y_n - \boldsymbol{x}_n^T \beta\right)^2 + C \tag{33}$$

unde C reprezintă termenii liberi (i.e. ce nu implică β).

Soluția MLE prin metoda gradient

Problema de optimizare devine

$$\mathcal{L}(\beta) = \frac{1}{2\sigma^2} - \sum_{i=1}^{m} \left(y_n - \boldsymbol{x}_n^T \beta \right)^2 \tag{34}$$

$$= \frac{1}{2\sigma^2} \left(\boldsymbol{y} - \boldsymbol{X}\beta \right)^T \left(\boldsymbol{y} - \boldsymbol{X}\beta \right) = \frac{1}{2\sigma^2} \left\| \boldsymbol{y} - \boldsymbol{X}\beta \right\|_2^2 \tag{35}$$

Remarcă

Pentru a găsi punctul maximal în probleme de tipul (31) *se folosesc metode de gradient.*

$$\frac{\partial \mathcal{L}}{\partial \beta} = \frac{\partial}{\partial \beta} \left(\frac{1}{2\sigma^2} \left(\boldsymbol{y} - \boldsymbol{X}\beta \right)^T \left(\boldsymbol{y} - \boldsymbol{X}\beta \right) \right) \tag{36}$$

$$= \frac{1}{2\sigma^2} \frac{\partial}{\partial \beta} \left(\boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{y}^T \boldsymbol{X}\beta + \beta^T \boldsymbol{X}^T \boldsymbol{X}\beta \right) \tag{37}$$

$$= \frac{1}{\sigma^2} \left(-\boldsymbol{y}^T \boldsymbol{X} + \beta^T \boldsymbol{X}^T \boldsymbol{X} \right) \tag{38}$$

Soluția MLE este soluția CMMP

Estimarea β_{ML} prin maximizarea verosimilității implică rezolvarea (39) în punctul zero:

$$\frac{\partial \mathcal{L}}{\partial \beta} = \mathbf{0} \iff \beta_{ML}^T \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{y}^T \boldsymbol{X} \tag{40}$$

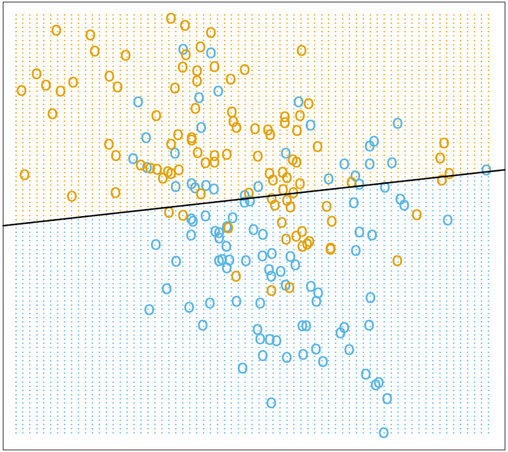
$$\iff \beta_{ML}^T = \boldsymbol{y}^T \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \tag{41}$$

$$\iff \beta_{ML} = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{y} \tag{42}$$

Remarcă

Ecuațiile (41) *reprezintă ecuațiile normale CMPP din* (3) *și sunt rezolvate eficient folosind Algoritmii 3 sau 5.*

Regresia liniară unidimensională



Regresia liniară bidimensională