Cozma Laura-Elena

344

Tehnici de Optimizare

Tema 1

Exercițiul 1

a. Fie funcția $f(x) = \sum_{i=1}^{m} \log(\, a_{(i)}^T x + b_i)$. Aflăm gradientul funcției f în x. Acesta este dat de formula generală:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Calculăm $\frac{\partial f(x)}{\partial x_1}$.

$$f(x) = \log\left(a_{(1)}^T x + b_1\right) + \log\left(a_{(2)}^T x + b_2\right) + \cdots + \log\left(a_{(m)}^T x + b_m\right)$$

Pentru a calcula $\frac{\partial f(x)}{\partial x_1}$, calculăm mai întâi derivata după x1 a primului termen al sumei, pentru a deduce formula generală:

$\frac{\partial(\log(a_{(1)}^T x+b_1))}{\partial x_1} = \frac{a_{(1)1}}{a_{(1)}^T x+b_1}$, știind că:

$\frac{\partial(a_{(1)}^T x+b_1)}{\partial x_1} = \frac{\partial(a_{(1)1}x_1+a_{(1)2}x_2+\cdots+a_{(1)n}x_n+b_1)}{\partial x_1} = a_{(1)1}$, unde $a_{(1)1}$ reprezintă primul element al vectorului $a_{(1)}$.

Analog,

$$\frac{\partial(\log(a_{(2)}^T x + b_2))}{\partial x_1} = \frac{a_{(2)1}}{a_{(2)}^T x + b_2}$$

...

$$\frac{\partial(\log(a_{(m)}^T x + b_m))}{\partial x_1} = \frac{a_{(m)1}}{a_{(m)}^T x + b_m}$$

De unde rezultă că:

$$\frac{\partial f(x)}{\partial x_1} = \sum_{i=1}^{m} \frac{a_{(i)1}}{a_{(i)}^T x + b_i}$$

Deci, în cazul general:

$$\frac{\partial f(x)}{\partial x_j} = \sum_{i=1}^{m} \frac{a_{(i)j}}{a_{(i)}^T x + b_i}$$

Așadar, gradientul funcției f în x va fi:

$$\nabla f(x) = \begin{bmatrix} \displaystyle\sum_{i=1}^{m} \frac{a_{(i)1}}{a_{(i)}^T x + b_i} \\ \vdots \\ \displaystyle\sum_{i=1}^{m} \frac{a_{(i)n}}{a_{(i)}^T x + b_i} \end{bmatrix}$$

În cazul general, matricea Hessiana a functiei f în x este:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

Calculăm $\frac{\partial f(x)}{\partial x_1 \partial x_1}$:

$$\frac{\partial f(x)}{\partial x_1 \partial x_1} = \frac{\partial\left(\sum_{i=1}^{m} \frac{a_{(i)1}}{a_{(i)}^T x + b_i}\right)}{\partial x_1} = \sum_{i=1}^{m} \frac{-a_{(i)1} a_{(i)1}}{(a_{(i)}^T x + b_i)^2}$$, știind că:

$$\frac{\partial\left(\frac{a_{(1)1}}{a_{(1)}^T x + b_1}\right)}{\partial x_1} = \frac{-a_{(1)1} a_{(1)1}}{(a_{(1)}^T x + b_1)^2}$$

...

$$\frac{\partial\left(\frac{a_{(m)1}}{a_{(m)}^T x + b_m}\right)}{\partial x_1} = \frac{-a_{(m)1} a_{(m)1}}{(a_{(m)}^T x + b_m)^2}$$

Calculăm $\frac{\partial f(x)}{\partial x_1 \partial x_2}$:

$\frac{\partial f(x)}{\partial x_1 \partial x_2} = \frac{\partial(\sum_{i=1}^{m} \frac{a_{(i)1}}{a_{(i)}^T x + b_i})}{\partial x_2} = \sum_{i=1}^{m} \frac{-a_{(i)1} a_{(i)2}}{(a_{(i)}^T x + b_i)^2}$, știind că:

$$\frac{\partial(\frac{a_{(1)1}}{a_{(1)}^T x + b_1})}{\partial x_2} = \frac{-a_{(1)1} a_{(1)2}}{(a_{(1)}^T x + b_1)^2}$$

...

$$\frac{\partial(\frac{a_{(m)1}}{a_{(m)}^T x + b_m})}{\partial x_2} = \frac{-a_{(m)1} a_{(m)2}}{(a_{(m)}^T x + b_m)^2}$$

Deducem că pentru cazul general formula va fi:

$$\frac{\partial f(x)}{\partial x_k \partial x_p} = \sum_{i=1}^{m} \frac{-a_{(i)k} a_{(i)p}}{(a_{(i)}^T x + b_i)^2}.$$

Așadar, matricea Hessiana a lui f în x este:

$$\nabla^2 f(x) = \begin{bmatrix} \sum_{i=1}^{m} \frac{-a_{(i)1}^2}{(a_{(i)}^T x + b_i)^2} & \sum_{i=1}^{m} \frac{-a_{(i)1} a_{(i)2}}{(a_{(i)}^T x + b_i)^2} & \cdots & \sum_{i=1}^{m} \frac{-a_{(i)1} a_{(i)n}}{(a_{(i)}^T x + b_i)^2} \\ \sum_{i=1}^{m} \frac{-a_{(i)2} a_{(i)1}}{(a_{(i)}^T x + b_i)^2} & \sum_{i=1}^{m} \frac{-a_{(i)2}^2}{(a_{(i)}^T x + b_i)^2} & \cdots & \sum_{i=1}^{m} \frac{-a_{(i)2} a_{(i)n}}{(a_{(i)}^T x + b_i)^2} \\ \vdots & \cdots & \cdots & \vdots \\ \sum_{i=1}^{m} \frac{-a_{(i)n} a_{(i)1}}{(a_{(i)}^T x + b_i)^2} & \sum_{i=1}^{m} \frac{-a_{(i)n} a_{(i)2}}{(a_{(i)}^T x + b_i)^2} & \cdots & \sum_{i=1}^{m} \frac{-a_{(i)n}^2}{(a_{(i)}^T x + b_i)^2} \end{bmatrix}$$

Se observă că matricea Hessiana poate fi scrisă ca sumă de matrice:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{-a_{(1)1}^2}{\left(a_{(1)}^T x+b_1\right)^2} & \frac{-a_{(1)1}a_{(1)2}}{\left(a_{(1)}^T x+b_1\right)^2} & \cdots & \frac{-a_{(1)1}a_{(1)n}}{\left(a_{(1)}^T x+b_1\right)^2} \\ \frac{-a_{(1)2}a_{(1)1}}{\left(a_{(1)}^T x+b_1\right)^2} & \frac{-a_{(1)2}^2}{\left(a_{(1)}^T x+b_1\right)^2} & \cdots & \frac{-a_{(1)2}a_{(1)n}}{\left(a_{(1)}^T x+b_1\right)^2} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{-a_{(1)n}a_{(1)1}}{\left(a_{(1)}^T x+b_1\right)^2} & \frac{-a_{(1)n}a_{(1)2}}{\left(a_{(1)}^T x+b_1\right)^2} & \cdots & \frac{-a_{(1)n}^2}{\left(a_{(1)}^T x+b_1\right)^2} \end{bmatrix} + \begin{bmatrix} \frac{-a_{(2)1}^2}{\left(a_{(2)}^T x+b_2\right)^2} & \frac{-a_{(2)1}a_{(2)2}}{\left(a_{(2)}^T x+b_2\right)^2} & \cdots & \frac{-a_{(2)1}a_{(2)n}}{\left(a_{(2)}^T x+b_2\right)^2} \\ \frac{-a_{(2)2}a_{(2)1}}{\left(a_{(2)}^T x+b_2\right)^2} & \frac{-a_{(2)2}^2}{\left(a_{(2)}^T x+b_2\right)^2} & \cdots & \frac{-a_{(2)2}a_{(2)n}}{\left(a_{(2)}^T x+b_2\right)^2} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{-a_{(2)n}a_{(2)1}}{\left(a_{(2)}^T x+b_2\right)^2} & \frac{-a_{(2)n}a_{(2)2}}{\left(a_{(2)}^T x+b_2\right)^2} & \cdots & \frac{-a_{(2)n}^2}{\left(a_{(2)}^T x+b_2\right)^2} \end{bmatrix} +$$

$$\cdots + \begin{bmatrix} \frac{-a_{(m)1}^2}{\left(a_{(m)}^T x+b_m\right)^2} & \frac{-a_{(m)1}a_{(m)2}}{\left(a_{(m)}^T x+b_m\right)^2} & \cdots & \frac{-a_{(m)1}a_{(m)n}}{\left(a_{(m)}^T x+b_m\right)^2} \\ \frac{-a_{(m)2}a_{(m)1}}{\left(a_{(m)}^T x+b_m\right)^2} & \frac{-a_{(m)2}^2}{\left(a_{(m)}^T x+b_m\right)^2} & \cdots & \frac{-a_{(m)2}a_{(m)n}}{\left(a_{(m)}^T x+b_m\right)^2} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{-a_{(m)n}a_{(m)1}}{\left(a_{(m)}^T x+b_m\right)^2} & \frac{-a_{(m)n}a_{(m)2}}{\left(a_{(m)}^T x+b_m\right)^2} & \cdots & \frac{-a_{(m)n}^2}{\left(a_{(m)}^T x+b_m\right)^2} \end{bmatrix}$$, relație echivalentă cu

$$\nabla^2 f(x) = \frac{1}{\left(a_{(1)}^T x + b_1\right)^2} \begin{bmatrix} -a_{(1)1}^2 & -a_{(1)1}a_{(1)2} & \cdots & -a_{(1)1}a_{(1)n} \\ -a_{(1)2}a_{(1)1} & -a_{(1)2}^2 & \cdots & -a_{(1)2}a_{(1)n} \\ \vdots & \cdots & \cdots & \vdots \\ -a_{(1)n}a_{(1)1} & -a_{(1)n}a_{(1)2} & \cdots & -a_{(1)n}^2 \end{bmatrix} + \cdots$$

$$+ \frac{1}{\left(a_{(m)}^T x+b_m\right)^2} \begin{bmatrix} -a_{(m)1}^2 & -a_{(m)1}a_{(m)2} & \cdots & -a_{(m)1}a_{(m)n} \\ -a_{(m)2}a_{(m)1} & -a_{(m)2}^2 & \cdots & -a_{(m)2}a_{(m)n} \\ \vdots & \cdots & \cdots & \vdots \\ -a_{(m)n}a_{(m)1} & -a_{(m)n}a_{(m)2} & \cdots & -a_{(m)n}^2 \end{bmatrix}$$

Se observă că matricile de forma $\begin{bmatrix} -a_{(i)1}^2 & -a_{(i)1}a_{(i)2} & \cdots & -a_{(i)1}a_{(i)n} \\ -a_{(i)2}a_{(i)1} & -a_{(i)2}^2 & \cdots & -a_{(i)2}a_{(i)n} \\ \vdots & \cdots & \cdots & \vdots \\ -a_{(i)n}a_{(i)1} & -a_{(i)n}a_{(i)2} & \cdots & -a_{(i)n}^2 \end{bmatrix}$ pot fi scrise ca $-a_{(i)}a_{(i)}^T$.

Astfel, matricea Hessiana devine:

$$\nabla^2 f(x) = \frac{-a_{(1)}a_{(1)}^T}{\left(a_{(1)}^T x + b_1\right)^2} + \ldots + \frac{-a_{(m)}a_{(m)}^T}{\left(a_{(m)}^T x + b_m\right)^2} = -\frac{aa^T}{\left(a^T x + b\right)^2} \ (2)$$

Determinăm constanta Lipschitz cu ajutorul relației:

$$||\nabla^2 f(x)||_2 \leq L \ (2)$$

Cu ajutorul proprietății normei $\left||x + y|\right| \leq \left||x|\right| + ||y||$ și al relației 1, relația 2 devine:

$$||\nabla^2 f(x)|| = \left\| \frac{-a_{(1)}a_{(1)}^T}{\left(a_{(1)}^T x + b_1\right)^2} + \cdots + \frac{-a_{(m)}a_{(m)}^T}{\left(a_{(m)}^T x + b_m\right)^2} \right\| \leq \left\| \frac{-a_{(1)}a_{(1)}^T}{\left(a_{(1)}^T x + b_1\right)^2} \right\| + .. + \left\| \frac{-a_{(m)}a_{(m)}^T}{\left(a_{(m)}^T x + b_m\right)^2} \right\| \leq Cmax \sum_{i=1}^{m} \frac{1}{\left(a_{(i)}^T x + b_i\right)^2}$$

unde Cmax = $max_{i=\{1..m\}} ||-a_{(i)}a_{(i)}^T||$.

I. În cazul din enunț în care $a_{(i)}^T x + b_i > 1 <=> \frac{1}{(a_{(i)}^T x+b_i)^2} < 1$, ultima relație devine:

$$||\nabla^2 f(x)|| \leq \frac{Cmax}{m}$$

Așadar, putem alege $L = \frac{Cmax}{m}$.

II. În cazul în care $a_{(i)}^T x + b_i > 0$, acesta ar putea fi oricât de mic astfel încât $\frac{1}{(a_{(i)}^T x+b_i)^2}$ să aibă o valoare foarte mare tinzând la infinit, deci $||\nabla^2 f(x)||$ să nu poată fi marginită. Așadar, funcția nu va avea gradient continuu Lipschitz.

b. Alegem n = 1 și m = 2. Funcția devine:

$$f(x) = \log(a_1 x + b_1) + \log(a_2 x + b_2)$$

Derivata funcției este:

$$f'(x) = \frac{a_1}{a_1 x + b_1} + \frac{a_2}{a_2 x + b_2}$$

Din $f'(x) = 0$ rezultă că:

$$a_1 a_2 x + a_1 b_2 + a_1 a_2 x + a_2 b_1 = 0$$

Deci, punctul de extrem este:

$$x = \frac{-a_1 b_2 - a_2 b_1}{2 a_1 a_2}$$

c. Vrem să verificăm dacă Hessiana funcției f este pozitiv semidefinită sau nu, iar aceasta se reduce la a arăta că suma din membrul stâng este pozitiv semidefinită.

$$\nabla^2 f(x) = \frac{-a_{(1)}a_{(1)}^T}{\left(a_{(1)}^T x + b_1\right)^2} + \ldots + \frac{-a_{(m)}a_{(m)}^T}{\left(a_{(m)}^T x + b_m\right)^2}$$

$aa_{(i)}^T$ este o matrice pozitiv semidefinită, conform definiției, pentru orice i. (*)

Fiecare termen de tipul $\frac{-1}{\left(a_{(1)}^T x+b_1\right)^2}$ este negativ. (**)

Din relațiile (*) și (**) și din faptul că sumă de matrice negativ semidefinite este tot o matrice negativ semidefinită, rezultă că matricea Hessiană este negativ semidefinită, deci funcția f(x) este concavă. Așadar, orice punct de extrem x* este punct de maxim.