МЕТОД ЕЛІПСОЇДІВ ДЛЯ ЛІНІЙНОЇ РЕГРЕСІЇ

Стецюк П.І., Стовба В.О.
stovbaviktor@yandex.ua

Інститут кібернетики імені В.М. Глушкова, Київ

IV МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ
"ОБЧИСЛЮВАЛЬНИЙ ІНТЕЛЕКТ
(РЕЗУЛЬТАТИ, ПРОБЛЕМИ, ПЕРСПЕКТИВИ) - 2017"
16–18 травня 2017 року, м. Київ

Зміст

Зміст

Метод еліпсоїдів запропонували

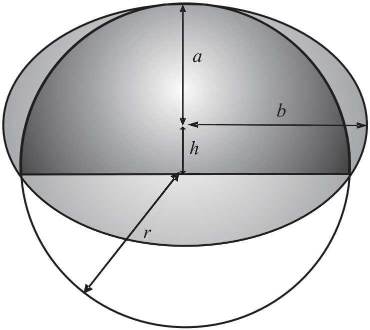
1976 **Юдін Д.Б. та Неміровський А.С.** як метод послідовних відсікань [1].

1977 **Шор Н.З.** як варіант методу з розтягом простору в напрямку субградієнта [2].

1. Юдин Д.Б., Немировский А.С. *Информационная сложность и эффективные методы решения выпуклых экстремальных задач* // Экономика и математические методы. – 1976. – Вып. 2. – С. 357–369.

2. Шор Н.З. *Метод отсечения с растяжением пространства для решения задач выпуклого программирования* // Кибернетика. – 1977. – № 1. – С. 94–95.

Ідея: мінімальний за об'ємом еліпсоїд $\mathcal{E}_n$



Еліпсоїд $\mathcal{E}_n$, що містить напівкулю в E^n, має наступні параметри:

$$a = \frac{n}{n+1}r,\ b = \frac{n}{\sqrt{n^2-1}}r,\ h = \frac{1}{n+1}r$$

де $\alpha = \frac{b}{a}$ та r – радіус кулі S_n.

Якщо простір "розтягнути"в напрямку a з коефіцієнтом $\alpha = \sqrt{\frac{n+1}{n-1}}$, то $\mathcal{E}_n$ трансформується в кулю у перетвореному просторі.

Відношення об'єму еліпсоїда $\mathcal{E}_n$ до об'єму кулі S_n рівне

$$q(n) = \frac{\mathrm{vol}(\mathcal{E}_n)}{\mathrm{vol}(S_n)} = \frac{1}{\alpha}\left(\frac{b}{r}\right)^n = \sqrt{\frac{n-1}{n+1}}\left(\frac{n}{\sqrt{n^2-1}}\right)^n < e^{-1/2n} < 1.$$

Зміст

В чому полягає задача?

Розглядається задача опуклого програмування: знайти

$$f_p^* = f_p(\theta_p^*) = \min_{\theta \in R^n} \left\{ f_p(\theta) = \|A\theta - b\|_p \right\}, \qquad (1)$$

за обмежень

$$\left\| \theta - \overline{\theta} \right\| \leq r, \qquad (2)$$

де A – $m \times n$-матриця; $b \in R^m$ – m-вимірний вектор; $\theta \in R^n$– n-вимірний вектор невідомих параметрів, $\overline{\theta}$ – центр кулі радіусу r, в якій локалізовано θ_p^*.

Тут $p \in R$ – скалярний параметр $(p \geq 1)$, який для вектора $y = A\theta - b = (y_1, \ldots, y_m)^T$ визначає L_p-норму

$$\|y\|_p = \left(\sum_{i=1}^{m} |y_i|^p \right)^{1/p}.$$

Випадок $p = \infty$ визначається як $\|y\|_\infty = \max_{i=1,\ldots,m} |y_i|$.

До задачі (1)-(2) зводиться задача знаходження параметрів лінійної регресії. Ця задача при

- $\boldsymbol{p = 1}$ – метод найменших модулів;
- $\boldsymbol{p = 2}$ – метод найменших квадратів;
- $\boldsymbol{p = \infty}$ – мінімаксний (чебишевський) метод.

Зміст

Вхідні параметри та ініціалізація алгоритму

Вхідні параметри:

- $\boldsymbol{p \geq 1}$ – визначає L_p-норму в (1);
- $\boldsymbol{\varepsilon_f > 0}$ – точність, з якою треба знайти $f_p^* = f_p(\theta_p^*)$.

Ініціалізація

Покладемо

- стартову точку $\theta_0 = \overline{\theta}$;
- початковий радіус $r_0 = r$;
- $B_0 := I_n$, де I_n – одинична $n \times n$-матриця.

Перейдемо до першої ітерації зі значеннями θ_0, r_0 та B_0.

Ітерація методу

Нехай на k-й ітерації знайдені θ_k, r_k, B_k. Перехід до ітерації $(k+1)$ полягає у виконанні такої послідовності дій.

Крок 1. Обчислимо $f_p(\theta_k)$. Якщо $f_p(\theta_k) = 0$, то "Зупинка: $k^* = k$ та $\theta_p^* = \theta_k$". Інакше обчислимо $g_p(\theta_k)$ – субградієнт функції $f_p(\theta)$ в точці θ_k за формулою

$$g_p(\theta_k) = (\|A\theta_k - b\|_p)^{1-p} \sum_{j=1}^{m} \left(sgn(a_j\theta_k - b_j) \cdot |a_j\theta_k - b_j|^{p-1} a_j^T \right),$$

де a_j – вектор-рядок матриці A з номером $j = \overline{1,m}$.
Якщо $r_k \left\| B_k^T g_p(\theta_k) \right\| \leq \varepsilon_f$, то "Зупинка: $k^* = k$ і $\theta_p^* = \theta_k$".
Інакше переходимо до кроку 2.

Крок 2. Покладемо $\xi_k := \dfrac{B_k^T g_p(\theta_k)}{\|B_k^T g_p(\theta_k)\|}$.

Крок 3. Обчислимо $\theta_{k+1} := \theta_k - h_k B_k \xi_k$, де $h_k = \dfrac{1}{n+1} r_k$.

Крок 4. Обчислимо нову матрицю та радіус

$$B_{k+1} := B_k + \left(\sqrt{\frac{n-1}{n+1}} - 1\right) (B_k \xi_k)\, \xi_k^T, \quad r_{k+1} := r_k \frac{n}{\sqrt{n^2-1}}.$$

Крок 5. Переходимо до ітерації $(k+1)$ зі значеннями $\theta_{k+1}, r_{k+1}, B_{k+1}$.

Збіжність методу

Теорема

Послідовність точок $\{\theta_k\}_{k=0}^{k^*}$ задовільняє нерівностям

$$\left\| B_k^{-1}(\theta_k - \theta_p^*) \right\| \le r_k, \quad k = 0, 1, 2, \dots, k^*$$

На кожній ітерації $k{>}0$ величина зменшення об'єму еліпсоїда $E_k = \{\theta{\in}R^n : \left\| B_k^{-1}(\theta_k{-}\theta) \right\| {\le} r_k\}$, який локалізує точку θ_p^*, є величиною сталою і рівною

$$q = \frac{vol(E_k)}{vol(E_{k-1})} = \sqrt{\frac{n-1}{n+1}} \left(\frac{n}{\sqrt{n^2-1}} \right)^n < \exp\left\{ -\frac{1}{2n} \right\} < 1.$$

Висновки

Алгоритм знаходження θ_p^* можна успішно застосовувати для випадку, коли кількість змінних $n = \overline{2, 10}$. Для зменшення в 10 разів об'єму еліпсоїда, в якому локалізовано точку θ_p^*, потрібно K ітерацій, де

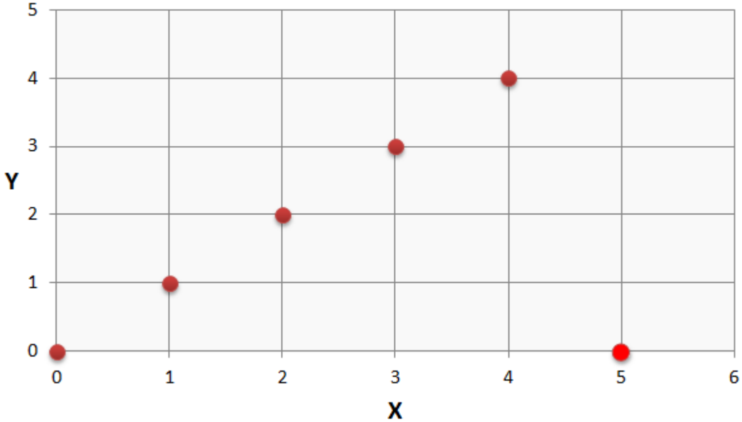
$$K = -\frac{\ln 10}{\ln q} \approx (2 \ln 10) n \approx 4.6 n,$$

тобто щоб на порядок покращити відхилення знайденого рекордного значення функції $f_p(\theta_p)$ від її оптимального значення f_p^*, необхідно зробити $4.6 n^2$ ітерацій.

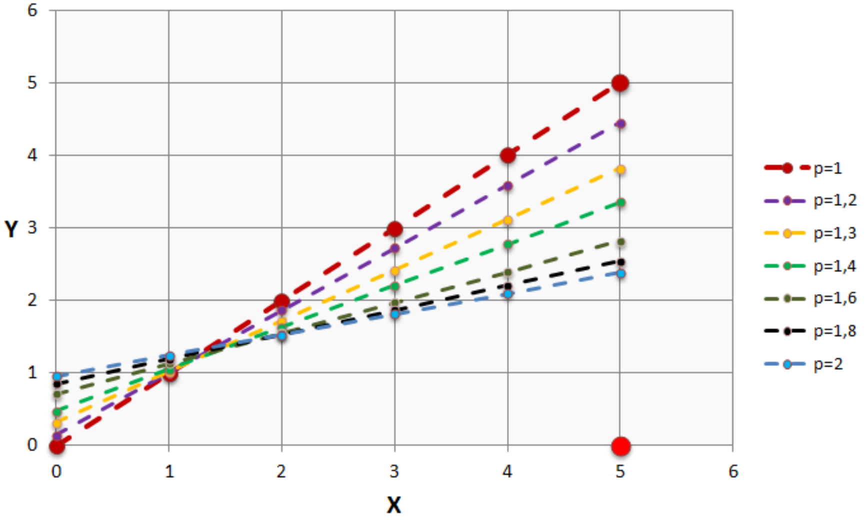
Зміст

Приклад застосування методу

Апроксимувати лінійною функцією $y = ax + b$ результати спостережень, серед яких наявне аномальне.



Приклад застосування методу



Результати роботи алгоритму

$$\boldsymbol{\varepsilon_f} = 10^{-12}, \ \boldsymbol{r} = 3$$

p	itn	a	b	f_r
1	200	1.0000	3.4408e-012	5.0000
1.2	119	0.86343	1.3768e-001	4.9047
1.3	111	0.70080	3.1609e-001	4.6989
1.4	107	0.57606	4.7512e-001	4.4615
1.6	108	0.42249	7.0521e-001	4.0324
1.8	107	0.33784	8.5195e-001	3.7011
2	104	0.28571	9.5238e-001	3.4503

Запитання?

ДЯКУЮ ЗА УВАГУ!