Методы оптимизации в машинном обучении

Андрей Данилов

February 2018

1. (a) $r_k = 0.99^k$
   По тесту корней линейная сходимость

   (b) $r_k = 0.99^{k^2}$
   По тесту корней сверхлинейная сходимость. Квадратичная сходимость отсутствует, т.к $\lim_{k\to\infty} 0.99^{k^2/2^k} = 1$

   (c) $r_k = 0.99^{2^k}$
   По тесту корней для квадратичной сходимости - квадратичная сходимость

   (d) $r_k = 1/k$
   По тесту корней сублинейная сходимость

   (e) $r_k = 1/\sqrt{k}$
   По тесту корней сублинейная сходимость

   (f) $r_k = 1/k^2$
   По тесту корней сублинейная сходимость

   (g) $r_k = 1/k!$
   По тесту отношений сверхлинейная сходимость. По тесту отношений для квадратичной сходимости квадратичная сходимость отсутствует

   (h) $r_k = 1/k^k$
   По тесту корней сверхлинейная сходимость. По тесту корней для квадратичной сходимости квадратичная сходимость отсутствует

   (i) $r_k = \begin{cases} 0.99^{2^k}, & k \text{ чётное} \\ r_{k-1}/k, & \text{иначе} \end{cases}$
   Для чётных k по тесту корней для квадратичной сходимости квадратичная сходимость. Для нечётных k также:

$$\lim_{k\to\infty} \left(\frac{0.99^{2^{k-1}}}{k} \right)^{1/2^k} = \lim_{k\to\infty} \frac{0.99^{\frac{2^{k-1}}{2^k}}}{k^{1/2^k}} = \sqrt{0.99}$$

   Поэтому в итоге квадратичная сходимость

(j) $r_k = \begin{cases} 0.5^{\frac{k+2}{2}}, & k \text{ чётное} \\ 0.5^{\frac{k-1}{2}}, & \text{иначе} \end{cases}$

В обоих случаях линейная сходимость по тесту корней, поэтому в итоге также линейная сходимость

2.

$$r_{k+1} \leq r_k - \frac{r_k^2}{c} \leq r_k, \text{ так как } r_k > 0$$

$$\frac{1}{r_{k+1}} - \frac{1}{r_k} = \frac{r_k - r_{k+1}}{r_{k+1} r_k} \geq \frac{r_k^2}{c} \frac{1}{r_k^2} = \frac{1}{c}$$

Далее по индукции. Для 0 утвеждение верно. Пусть верно для k, тогда для $k+1$:

$$\frac{1}{r_{k+1}} \geq \frac{1}{c} + \frac{1}{r_k} \geq \frac{1}{c} + \frac{k}{c} = \frac{k+1}{c}$$

3.

$$r_k \leq \left(1 - \frac{1}{Q}\right) r_{k-1} \leq \left(1 - \frac{1}{Q}\right)^k r_0 \leq \delta r_0$$

Последнее неравенство выполнено, если $\left(1 - \frac{1}{Q}\right)^k \leq \delta$.
Тогда:

$$k \ln\left(1 - \frac{1}{Q}\right) \leq \ln \delta$$

$$-\ln \delta \leq k \ln\left(1 + \frac{1}{Q-1}\right) \leq \frac{k}{Q-1}$$

$$k \geq Q \ln \frac{1}{\delta} + \ln \delta$$

Поскольку $\ln \delta \leq 0$, то из условия $k \geq Q \ln \frac{1}{\delta}$ следует последнее неравенство, а значит и $r_k \leq \delta r_0$

4. Обозначим $q = Mr_0$
Докажем по индукции, что $\forall k \; r_k = r_0 q^{2^k - 1}$
Для 0 верно, пусть верно для k, тогда

$$r_{k+1} = M r_k^2 = M r_0^2 q^{2^{k+1} - 2} = r_0 q^{2^{k+1} - 1}$$

Отсюда видно, что необходимо и достаточно, чтобы $Mr_0 < 1$, причём при выполнении этого условия скорость сходимости квадратичная.

5. (a)

$$Det(AXB(C^{-T} X^T C)^{-T}) = \frac{Det(A)Det(X)Det(B)Det(C)}{Det(X)Det(C)} = Det(AB)$$

(b)

$$\|uv^T - A\|_F^2 - \|A\|_F^2 = \sum_{i,j}(u_iv_j - a_{ij})^2 - \sum_{i,j} a_{ij}^2 =$$
$$= \sum_{i,j}(u_i^2v_j^2 - 2u_iv_ja_{ij}) = \|u\|\|v\| - 2u^TAv$$

(c) Обозначим $T = S^{-1}$

$$\sum_{i=1}^{n}\langle Ta_i,\ a_i\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}t_{jk}a_{ik} = \sum_{j=1}^{n}\sum_{k=1}^{n} t_{jk}\sum_{i=1}^{n} a_{ij}a_{ik} =$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{n} t_{jk}s_{kj} = \mathrm{tr}(TS) = \mathrm{tr}(I_n) = n$$

(d) По формуле Шермана-Моррисона $(2I_n + aa^T)^{-1} = \frac{I_n}{2} - \frac{aa^T}{4+2a^Ta}$

$$\mathrm{tr}\left(\left(2I_n + aa^T\right)^{-1}\left(uv^T + vu^T\right)\right) =$$
$$= \mathrm{tr}\left(\left(\frac{I_n}{2} - \frac{aa^T}{4+2a^Ta}\right)\left(uv^T + vu^T\right)\right) =$$
$$= \mathrm{tr}\left(uv^T\right) - \mathrm{tr}\left(\frac{aa^Tuv^T}{4+2a^Ta}\right) =$$
$$= (u,v) - \frac{(u,a)(v,a)}{2+(a,a)}$$

6. (a) $f(t) = det\,(A - tI_n)$

$$df(t) = det\,(A - tI_n)\,(A - tI_n, -I_ndt) = -det\,(A - tI_n)\,\mathrm{tr}(A - tI_n)dt$$
$$f'(t) = -det(A - tI_n)(\mathrm{tr}(A) - nt)$$
$$df'(t) = -d\left(det\,(A - tI_n)\right)(\mathrm{tr}(A) - nt) + det\,(A - tI_n)\cdot ndt$$
$$f''(t) = det(A - tI_n)\left((\mathrm{tr}(A) - nt)^2 + n\right)$$

(b) $f(t) = \|(A + tI_n)^{-1}b\|^2$

$$\begin{array}{ll} f_1(x) = \|x\|^2 & df_1(x) = 2x^Tdx \\ f_2(X) = X^{-1}b & df_2(X) = -X^{-1}dXX^{-1}b \\ f_3(t) = X + tI_n & df_3(t) = I_ndt \end{array}$$

$$f(t) = f_1(f_2(f_3(t)))$$
$$df(t) = -2((A + tI_n)^{-1}b)^T((A + tI_n)^{-1}I_n dt(A + tI_n)^{-1}b)$$
$$f'(t) = -2b^T(A + tI_n)^{-3}b$$
$$f''(t) = 6b^T(A + tI_n)^{-4}b$$

Продолжение в рукописном виде