Методы оптимизации в машинном обучении, ВМК+Физтех, осень 2017

Домашняя работа 1: Скорости сходимости и матрично-векторное дифференцирование

Срок сдачи: 20 сентября 2017 (среда), 23:59 для ВМК
22 сентября 2017 (пятница), 23:59 для Физтеха

1 Пусть $(r_k)_{k=1}^{\infty}$ — одна из следующих последовательностей:

(a) $r_k := (0.99)^k$
(b) $r_k := (0.99)^{k^2}$
(c) $r_k := (0.99)^{2^k}$
(d) $r_k := 1/k$
(e) $r_k := 1/\sqrt{k}$
(f) $r_k := 1/k^2$
(g) $r_k := 1/k!$
(h) $r_k := 1/k^k$
(i) $r_k := \begin{cases} (0.99)^{2^k}, & \text{если } k \text{ четное,} \\ r_{k-1}/k, & \text{иначе} \end{cases}$
(j) $(r_k)_{k=1}^{\infty} := \left(1, \frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{4}, \frac{1}{16}, \frac{1}{8}, \dots\right)$

Для каждого из указанных вариантов классифицируйте $(r_k)_{k=1}^{\infty}$ по скорости сходимости (линейная, сублинейная, сверхлинейная). В случае сверхлинейной сходимости дополнительно выясните, имеет ли место квадратичная сходимость.

2 Пусть $C > 0$, и пусть $(r_k)_{k=1}^{\infty}$ — одна из следующих трех последовательностей:

(a) (Сублинейная) $r_k := C/k^{\gamma}$, где $\gamma > 0$.
(b) (Линейная) $r_k := Cq^k$, где $0 < q < 1$.
(c) (Квадратичная) $r_k := C(C^{-1}R)^{2^k}$, где $R > 0$ и $0 < C^{-1}R < 1$.

Пусть $0 < \varepsilon < 1$, и пусть $T(\varepsilon) := \min\{k \geq 1 : r_k \leq \varepsilon C\}$ — первый момент времени достижения относительной точности ε.

(a) Для каждого из трех вариантов последовательности выпишите явную формулу для T.
(b) Проанализируйте, насколько сильно T зависит от точности ε и параметра последовательности (γ, q, R соответственно).
(c) Заполните следующие таблицы, вписав в пустые ячейки соответствующие числовые значения T:

Сублинейная			
ε \ γ	1	2	0.5
10^{-1}			
10^{-3}			
10^{-5}			
10^{-7}			
10^{-12}			

Линейная			
ε \ q	0.9	0.999	0.99999
10^{-1}			
10^{-3}			
10^{-5}			
10^{-7}			
10^{-12}			

Квадратичная			
ε \ R	$0.9\,C$	$0.999\,C$	$0.99999\,C$
10^{-1}			
10^{-3}			
10^{-5}			
10^{-7}			
10^{-12}			

(*Рекомендация.* Напишите программу, которая заполнит эти таблицы автоматически. Достаточно выписать одну значимую цифру и степень (например: 3×10^8).)

3 Упростите каждое из следующих выражений:

(a) $\mathrm{Det}(AXB(C^{-T}X^TC)^{-T})$, где $A, B, C, X \in \mathbb{R}^{n\times n}$, $\mathrm{Det}(C) \neq 0$, $\mathrm{Det}(C^{-T}X^TC) \neq 0$.
(b) $\|uv^T - A\|_F^2 - \|A\|_F^2$, где $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$.
(c) $\mathrm{Tr}((2I_n + aa^T)^{-1}(uv^T + vu^T))$, где $a, u, v \in \mathbb{R}^n$.
(d) $\sum_{i=1}^n \langle S^{-1}a_i, a_i\rangle$, где $a_1, \dots, a_n \in \mathbb{R}^n$, $S := \sum_{i=1}^n a_i a_i^T$, $\mathrm{Det}(S) \neq 0$.

4 Пусть f — одна из следующих функций:

(a) $f : E \to \mathbb{R}$ — функция $f(t) := \mathrm{Det}(A - tI_n)$, где $A \in \mathbb{R}^{n \times n}$, $E := \{t \in \mathbb{R} : \mathrm{Det}(A - tI_n) \neq 0\}$.

(b) $f : \mathbb{R}_{++} \to \mathbb{R}$ — функция $f(t) := |(A + tI_n)^{-1} b|^2$, где $A \in \mathbb{S}^n_+$, $b \in \mathbb{R}^n$.

Для каждого из указанных вариантов вычислите первую и вторую производные f' и f''.

5 Пусть f — одна из следующих функций:

(a) $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := \dfrac{1}{2}\|xx^T - A\|_F^2$, где $A \in \mathbb{S}^n$.

(b) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ — функция $f(x) := \dfrac{\langle Ax, x\rangle}{|x|^2}$, где $A \in \mathbb{S}^n$.

(c) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ — функция $f(x) := \langle x, x\rangle^{\langle x, x\rangle}$.

Для каждого из указанных вариантов вычислите вектор градиент ∇f и матрицу гессиан $\nabla^2 f$ (относительно стандартного скалярного произведения в пространстве $\mathbb{R}^n$).

6 Пусть $f : \mathbb{S}^n_{++} \to \mathbb{R}$ — одна из следующих функций:

(a) $f(X) := \mathrm{Tr}(X^{-1})$.

(b) $f(X) := \langle X^{-1} v, v\rangle$, где $v \in \mathbb{R}^n$.

(c) $f(X) := (\mathrm{Det}(X))^{1/n}$.

Для каждого из указанных вариантов покажите, что вторая производная $D^2 f(X)[H, H]$ имеет постоянный знак для всех $X \in \mathbb{S}^n_{++}$ и всех $H \in \mathbb{S}^n$.

(*Подсказка.* В последнем пункте воспользуйтесь неравенством Коши–Буняковского.)

7 Пусть f — одна из следующих функций:

(a) $f : \mathbb{R}^2 \to \mathbb{R}$ — функция $f(x) := 2x_1^2 + x_2^2(x_2^2 - 2)$.

(b) $f : \mathbb{R}^2 \to \mathbb{R}$ — функция $f(x) := (1 - x_1)^2 + 100(x_2 - x_1^2)^2$.

(c) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ — функция $f(x) := \dfrac{\langle Ax, x\rangle}{|x|^2}$, где $A \in \mathbb{S}^n$.

Для каждого из указанных вариантов найдите все точки стационарности f и определите их тип (локальный минимум/максимум, седловая точка).

8 Пусть c — ненулевой вектор в $\mathbb{R}^n$, $\sigma > 0$, и пусть $f : \mathbb{R}^n \to \mathbb{R}$ — функция

$$f(x) := \langle c, x\rangle + \frac{\sigma}{3}|x|^3.$$

Найдите единственную точку стационарности функции f.