Методы оптимизации в машинном обучении, ШАД, весна 2018

Домашняя работа 1: Скорости сходимости последовательностей и матрично-векторное дифференцирование

Срок сдачи: 2 марта 2017 (пятница), 23:59

1 Классифицируйте каждую из следующих последовательностей $(r_k)_{k=1}^\infty$ по скорости сходимости (линейная, сублинейная, сверхлинейная). В случае сверхлинейной сходимости дополнительно выясните, имеет ли место квадратичная сходимость.

(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) $(r_k)_{k=1}^\infty := \left(1, \frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{4}, \frac{1}{16}, \frac{1}{8}, \dots\right)$

2 Пусть $(r_k)_{k=0}^\infty$ — последовательность положительных чисел, такая, что $r_{k+1} \le r_k - r_k^2/c$ для некоторого $c > 0$ и всех $k \ge 0$. Покажите что $r_k \le c/k$ для всех $k \ge 1$. (*Подсказка*: оцените разность $1/r_{k+1} - 1/r_k$ снизу.)

3 Пусть $Q > 1$, и пусть $(r_k)_{k=0}^\infty$ — неотрицательная последовательность, такая, что $r_{k+1} \le (1-Q^{-1})r_k$ для всех $k \ge 0$. Покажите, что для того, чтобы обеспечить $r_k \le \delta r_0$ для некоторого $0 < \delta < 1$, достаточно выбрать $k \ge Q \ln \frac{1}{\delta}$. (Таким образом, при такой скорости сходимости каждая новая правильная цифра после запятой гарантированно добавляется через каждые Q шагов.)

4 Пусть $(r_k)_{k=0}^\infty$ — рекуррентно заданная последовательность неотрицательных чисел $r_{k+1} := M r_k^2$, где $M > 0$ и $r_0 \ge 0$. Установите необходимое и достаточное условие на M и r_0, при котором последовательность $(r_k)_{k=0}^\infty$ будет сходиться к нулю. Какова при этом скорость сходимости?

5 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) $\sum_{i=1}^n \langle S^{-1}a_i, a_i\rangle$, где $a_1, \dots, a_n \in \mathbb{R}^n$, $S := \sum_{i=1}^n a_i a_i^T$, $\mathrm{Det}(S) \neq 0$.
(d) $\mathrm{Tr}((2I_n + aa^T)^{-1}(uv^T + vu^T))$, где $a, u, v \in \mathbb{R}^n$.

(*Подсказка*: Используйте формулу Шермана–Моррисона.)

6 Для каждой из следующих функций f вычислите первую и вторую производные f' и f'':

(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 $A \in \mathbb{S}_+^n$, $b \in \mathbb{R}^n$.

7 Для каждой из следующих функций f вычислите градиент ∇f и гессиан $\nabla^2 f$ (относительно стандартного скалярного произведения в пространстве $\mathbb{R}^n$):

(a) $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := \frac{1}{2}\|xx^T - A\|_F^2$, где $A \in \mathbb{S}^n$.
(b) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ — функция $f(x) := \frac{\langle Ax, x\rangle}{\|x\|^2}$, где $A \in \mathbb{S}^n$.
(c) $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ — функция $f(x) := \langle x, x\rangle^{\langle x, x\rangle}$.
(d) $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := \ln(\sum_{i=1}^m e^{\langle a_i, x\rangle})$, где $a_1, \dots, a_m \in \mathbb{R}^n$.

8 Покажите, что каждая из следующих функций $f : \mathbb{S}_{++}^n \to \mathbb{R}$ матричного аргумента является выпуклой:

(a) $f(X) := \langle X^{-1}, A\rangle$, где $A \in \mathbb{S}_+^n$.
(b) $f(X) := -(\mathrm{Det}(X))^{1/n}$.

(*Подсказка.* В последнем пункте используйте неравенство Коши–Буняковского.)

9 Покажите, что каждая из следующих функций f является вогнутой:

(a) $f : E \to \mathbb{R}$ — функция $f(x) := \ln(-Q(x))$, где $E := \{x \in \mathbb{R}^n : Q(x) < 0\}$, $Q : \mathbb{R}^n \to \mathbb{R}$ — функция $Q(x) := \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + c$, $A \in \mathbb{S}^n_+$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

(b) $f : \mathbb{R}^n_{++} \to \mathbb{R}$ — функция $f(x) := (\sum_{i=1}^n x_i^p)^{1/p}$, где $p < 1$, $p \neq 0$.

(c) $f : \mathbb{R}^n_{++} \to \mathbb{R}$ — функция $f(x) := \prod_{i=1}^n x_i^{\alpha_i}$, где $\alpha_1, \ldots, \alpha_n \geq 0$, $\sum_{i=1}^n \alpha_i = 1$.

(*Подсказка*: В некоторых пунктах могут оказаться полезными неравенства Коши–Буняковского и Йенсена.)

10 Пусть $g : \mathbb{R} \to \mathbb{R}$ — дифференцируемая функция, производная которой является липшицевой с параметром $L > 0$. Пусть $a \in \mathbb{R}^n$, $b \in \mathbb{R}$, и пусть $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := g(\langle a, x\rangle + b)$. Покажите, что градиент функции f является липшицевым с параметром $L\|a\|^2$. Установите отсюда, что функция потерь логистической регрессии $x \mapsto \sum_{i=1}^m \ln(1 + e^{\langle a_i, x\rangle})$, где $a_1, \ldots, a_m \in \mathbb{R}^n$, обладает липшицевым градиентом с параметром $\frac{1}{4}\sum_{i=1}^m \|a_i\|^2$.

11 Для каждой из следующих функций f найдите все точки стационарности и определите их тип (локальный минимум, локальный максимум, седловая точка):

(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 $f(x) := \frac{\langle Ax, x\rangle}{\|x\|^2}$, где $A \in \mathbb{S}^n$.

12 Для каждой из следующих функций найдите множество точек (глобального) минимума:

(a) $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := \langle c, x\rangle + \frac{\sigma}{3}\|x\|^3$, где $c \in \mathbb{R}^n$, $c \neq 0$, $\sigma > 0$.

(b) $f : E \to \mathbb{R}$ — функция $f(x) := \langle a, x\rangle - \ln(1 - \langle b, x\rangle)$, где $a, b \in \mathbb{R}^n$, $a, b \neq 0$, $E := \{x \in \mathbb{R}^n : \langle b, x\rangle < 1\rangle$.

(c) $f : \mathbb{R}^n \to \mathbb{R}$ — функция $f(x) := \langle c, x\rangle e^{-\langle Ax, x\rangle}$, где $c \in \mathbb{R}^n$, $c \neq 0$, $A \in \mathbb{S}^n_{++}$.

(d) $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ — функция $f(X) := \|AX - B\|_F^2$, где $A \in \mathbb{R}^{k\times m}$, $B \in \mathbb{R}^{k\times n}$, $\mathrm{Rank}(A) = m$.

(e) $f : \mathbb{S}^n_{++} \to \mathbb{R}$ — функция $f(X) := \langle X^{-1}, A\rangle + \ln \mathrm{Det}(X)$, где $A \in \mathbb{S}^n$.

(f) $f : \mathbb{S}^n_{++} \to \mathbb{R}$ — функция $f(X) := \langle X^{-1}, I_n\rangle - \langle A, X\rangle$, где $A \in \mathbb{S}^n$.

(Обратите внимание, что в некоторых задачах при определенных значениях параметров множество точек минимума может быть пустым.)