Домашнее задание №1. Выпуклые множества. Выпуклые и гладкие функции. Матрично-векторное дифференцирование. Субдифференциал.

Дедлайн: 27 марта 2022 года, 23:59 по Москве

Действует система[1] мягких дедлайнов!

Просьба присылать задания в виде **одного PDF-файла** (можно использовать LaTeX, а можно просто аккуратно сфотографировать рукописные решения и **собрать фотографии в один PDF-файл**) на почту **Горбунова Эдуарда**: ed-gorbunov@yandex.ru. Кроме того, просьба **указывать следующую тему письма:**

«Методы оптимизации в ML, весна 2022. Домашнее задание 1»

1. **Выпуклые множества (1 балл).** Какие из следующих операций не сохраняют выпуклость, если $X, Y \subseteq \mathbb{R}^n$ - выпуклые множества? Ответы обосновать.

   a) $X \cup Y$

   b) $X \times Y = \{(x, y) \mid x \in X, y \in Y\}$

   c) $\alpha X + \beta Y = \{\alpha x + \beta y \mid x \in X,\ y \in Y,\ \alpha, \beta \in \mathbb{R}\}$

   d) $\alpha X = \{\alpha x \mid x \in X,\ \alpha \in \mathbb{R}_-\}$

   e) $X^c = \{x \in \mathbb{R}^n \mid x \notin X\}$

2. **Матрично-векторное дифференцирование (3 балла).**

   3.1. Посчитайте $df(x)$ и $\nabla f(x)$ для функции $f(x) = \log(x^\top \mathrm{A} x)$ и выберите правильный вариант ответа.

   a) $\nabla f(x) = \frac{2\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A} x},\ df(x) = \frac{x^\top 2\left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A} x}$

   b) $\nabla f(x) = \frac{\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A} x},\ df(x) = \frac{x^\top \left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A} x}$

   c) $\nabla f(x) = \frac{x^\top \left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A} x},\ df(x) = \frac{\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A} x}$

[1] Указанная дата – мягкий дедлайн, т.е. работы можно присылать некоторое время после дедлайна. Но количество дней просрочки влияет на финальное количество набранных баллов: **если прислать задание на k-й день после дедлайна, то заработанные баллы будут домножены на коэффициент** $\max\{0, 1 - 0.1 \cdot k\}$. Таким образом, если прислать задание на первый день после дедлайна, то баллы будут домножены на 0.9, если на второй день – на 0.8 и так далее. Отсюда следует, что на 10-й день и позднее после дедлайна за задание можно набрать только 0 баллов. Задание можно присылать частями в разные дни. Тогда баллы за каждую часть, присланную после мягкого дедлайна, будут домножены на свой коэффициент.

d) $\nabla f(x) = \frac{x^\top 2\left(A + A^\top\right)dx}{x^\top A x}$, $df(x) = \frac{2\left(A + A^\top\right)x}{x^\top A x}$

3.2. Посчитайте $df(x)$, $d^2 f(x)$ и $\nabla f(x)$, $\nabla^2 f(x)$ для функции $f(x) = \frac{1}{p}\|x\|_2^p$, $p > 1$ и выберите все правильные результаты.

a) $\nabla f(x) = \|x\|_2^{p-1} x$

b) $df(x) = \|x\|_2^{p-2} x^\top dx$

c) $\nabla^2 f(x) = (p-2)\|x\|_2^{p-4} x^\top x + \|x\|_2^{p-2}$

d) $\nabla^2 f(x) = (p-1)\|x\|_2^{p-3} x x^\top + \|x\|_2^{p-1} I$

e) $d^2 f(x) = dx^\top \left((p-2)\|x\|_2^{p-4} x x^\top + \|x\|_2^{p-2} I\right) dx$

f) $d^2 f(x) = dx^\top \left((p-1)\|x\|_2^{p-3} x x^\top + \|x\|_2^{p-1} I\right) dx$

3.3. Верно ли, что у функции $f(x) = \frac{1}{n}\sum\limits_{i=1}^{n} \log\left(1 + \exp(a_i^\top x)\right) + \frac{\mu}{2}\|x\|_2^2$, $a_i \in \mathbb{R}^n$, $\mu > 0$, гессиан $\nabla^2 f(x) = \frac{1}{n}\sum\limits_{i=1}^{n}\left(-\frac{\exp(a_i^\top x)}{(1+\exp(a_i^\top x))^2} a_i a_i^\top + \frac{\exp(a_i^\top x)}{1+\exp(a_i^\top x)} a_i a_i^\top\right) + \mu \mathrm{I}$?

3. **Выпуклые функции (3 балла).** Докажите выпуклость следующих функций (можно пользоваться любыми свойствами/утверждениями/теоремами с лекций):

(a) $f(x) = \left(\sum\limits_{i=1}^{n} e^{x_i}\right)$;

(b) $f(x) = \frac{\|Ax - b\|^2}{1 - x^\top x}$, на множестве $\mathrm{X} = \{x \mid \|x\|^2 < 1\}$;

(c) $F(x) = \frac{f^2(x)}{g(x)}$ при условии, что $f(x)$ и $g(x)$ дважды непрерывно дифференцируемы, $f(x)$ является выпуклой и принимает только неотрицательные значения, $g(x)$ вогнута и принимает только положительные значения;

(d) $f(x,t) = -\log(t^2 - x^T x)$, на множестве $\mathrm{E} = \{(x,t)\} \in \mathbb{R}^{n \times 1} \mid \|x\|_2 < t\}$;

(e) $f(x) = \sum\limits_{i=1}^{n} w_i \ln\left(1 + \exp(a_i^\top x)\right) + \frac{\mu}{2}\|x\|_2^2$, $\mu, w_1, \ldots, w_n > 0$, $a_i \in \mathbb{R}^n$;

(f) $f(x) = \ln \sum\limits_{i=1}^{n} \exp\left(\max\{0, x_i\}^2\right)$. *Подсказка:* для начала докажите, что функция $g(x) = \ln\left(\sum\limits_{i=1}^{n} \exp x_i\right)$ является выпуклой (воспользуйтесь определением выпуклой функции и неравенством Гёльдера[2])

4. **(1 балл)** Посчитайте субдифференциал функции $f(x) = |c^\top x|$, $x \in \mathbb{R}^n$.

5. **(1 балл)** Посчитайте субдифференциал функции $f(x) = \|x\|_1$, $x \in \mathbb{R}^n$.

6. **(1 балл)** Определите константы μ и L для функции $f(x) = \|x\|_2^2$.

[2]**Неравенство Гёльдера**: пусть $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, тогда

$$\sum_{i=1}^{n} |x_i y_i| \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}}.$$

Дополнительные задачи[3]

Дедлайн: 27 марта 2022 года, 23:59 по Москве

Действует система[4] мягких дедлайнов!

1. **(2 балла)** Пусть $\mathrm{T}(x, \omega)$ обозначает тригонометрический полином
$$\mathrm{T}(x, \omega) = \sum_{i=1}^{n} x_i \cos((i-1)\omega).$$
Покажите, что функция
$$f(x) = -\int_0^{2\pi} \log \mathrm{T}(x, \omega) d\omega$$
выпукла на $\{x \in \mathbb{R}^n \mid \mathrm{T}(x, \omega) > 0,\ 0 \leq \omega \leq 2\pi\}$.

2. **(1 балл)** Пусть $p < 1$ и $p \neq 0$. Покажите, что функция
$$f(x) = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}$$
с $\mathrm{dom} f = \mathbb{R}^n_{++}$ (область опеределения функции f – вектора в $\mathbb{R}^n$ с положительными компонентами) является вогнутой.

3. **(1 балл)** Посчитайте субдифференциал функции $f(x) = \sin x$ на множестве $X = [0, \frac{3}{2}\pi]$.

4. **(1 балл)** Посчитайте субдифференциал функции $f(x) = \max\{e^x, 1 - x\},\ x \in \mathbb{R}$.

5. **(1 балл)** Какие условия верны для сильно выпуклой функции $f(x)$ с константной μ? Ответы обосновать.

   - $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|x - y\|_2^2 \quad \forall x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}$
   - $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2 \quad \forall x, y \in \mathbb{R}^n$
   - $f(y) \leq f(x) + \langle \nabla f(y), y - x\rangle + \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \forall x, y \in \mathbb{R}^n$
   - если $\nabla f(x^*) = 0$, то $\|\nabla f(x)\|_2^2 \leq 2\mu(f(x) - f(x^*)) \quad \forall x \in \mathbb{R}^n$
   - $\langle \nabla f(x) - \nabla f(y), x - y\rangle \leq \frac{1}{\mu}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \forall x, y \in \mathbb{R}^n$

[3]Данные задачи можно решать даже если что-то не получилось из обязательных задач. Баллы за дополнительные задачи будут прибавляться к баллам за основные задачи.

[4]Указанная дата – мягкий дедлайн, т.е. работы можно присылать некоторое время после дедлайна. Но количество дней просрочки влияет на финальное количество набранных баллов: **если прислать задание на k-й день после дедлайна, то заработанные баллы будут домножены на коэффициент** $\max\{0, 1 - 0.1 \cdot k\}$. Таким образом, если прислать задание на первый день после дедлайна, то баллы будут домножены на 0.9, если на второй день – на 0.8 и так далее. Отсюда следует, что на 10-й день и позднее после дедлайна за задание можно набрать только 0 баллов. Задание можно присылать частями в разные дни. Тогда баллы за каждую часть, присланную после мягкого дедлайна, будут домножены на свой коэффициент.

- если $f(x)$– дважды непрерывно дифференцирема, то $\nabla^2 f(x) \succeq \mu \mathrm{I}_n \quad \forall x \in \mathbb{R}^n$
- $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n$
- если $\nabla f(x^*) = 0$, то $f(x) \leq f(x^*) + \frac{\mu}{2}\|x - x^*\|^2 \quad \forall x \in \mathbb{R}^n$

6. **(2 балла)** Докажите, что если функция L-гладкая, то выполнено $\|\nabla^2 f(x)\|_2 \leq L$.

7. **(1 балл)** Докажите, что если функция L-гладкая и выпуклая, то выполнено неравенство

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y) \quad \forall x, y \in \mathbb{R}^n$$