Хмелёв Алексей, MADE-DS-22

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

a) $X \cup Y$ - не сохраняет.
Пусть $X = \{x \mid x \in \mathbb{R}^n,\ \|x\|_2 \le 1\}$, $Y = \{y \mid y \in \mathbb{R}^n,\ \|y - 10 \cdot \mathbf{1}\|_2 \le 1\}$, где $\mathbf{1}$ - вектор из единиц. Простыми словами, X - сфера единичного радиуса с центром в начале координат, а Y - единичная сфера с центром в точке с координатами $(10, 10, ..., 10)$. Тогда очевидно, что при объединении множеств, если рассмотреть отрезок с одним концом в X и другим в Y, то только часть отрезка будет находится в объединении, что противоречит определению выпуклого множества

b) $X \times Y = \{(x, y) \mid x \in X,\ y \in Y\}$ - сохраняет
$\forall\ x_i, x_j \in X,\ \alpha \in [0; 1] :\ \alpha x_i + (1 - \alpha)x_j = x_k \in X$
$\forall\ y_i, y_j \in Y,\ \beta \in [0; 1] :\ \beta y_i + (1 - \beta)y_j = y_m \in Y$

$(x_i, y_j),\ \gamma \in [0; 1] : \gamma(x_i, y_j) + (1 - \gamma)(x_i, y_j) = (\gamma x_i + (1 - \gamma)x_i, \gamma y_j + (1 - \gamma)y_j) = (x_k, y_m) \in X \times Y$

c) $\alpha X + \beta Y = \{\alpha x + \beta y \mid x \in X,\ y \in Y,\ \alpha, \beta \in \mathbb{R}\}$ - сохраняет
$\forall\ x_i, x_j \in X,\ \alpha \in [0; 1] :\ \alpha x_i + (1 - \alpha)x_j = x_k \in X$
$\forall\ y_i, y_j \in Y,\ \beta \in [0; 1] :\ \beta y_i + (1 - \beta)y_j = y_m \in Y$

$\gamma \in [0; 1] : \gamma(\alpha x_i + \beta y_j) + (1 - \gamma)(\alpha x_i + \beta y_j) = \alpha(\gamma x_i + (1 - \gamma)x_i) + \beta(\gamma y_j + (1 - \gamma)y_j) = \alpha x_k + \beta y_m \in \alpha X + \beta Y$

d) $\alpha X = \{\alpha x \mid x \in X,\ \alpha \in \mathbb{R}\}$ - сохраняет, в силу утверждения (c) при $\beta = 0$

e) $X^c = \{x \in \mathbb{R}^n \mid x \notin X\}$ - не сохраняет
Пусть $X = \{x \mid x \in \mathbb{R}^n,\ \|x\|_2 \le 1\}$, тогда $X^c = \{x \mid x \in \mathbb{R}^n,\ \|x\|_2 > 1\}$. Если выбрать точку $a \in X^c$ и $-a$, которая тоже принадлежит множеству X^c, то часть отрезка, соединяющего эти точки, будет находиться во множестве X, что противоречит определению выпуклого множества
Ответ: a, e

2. Матрично-векторное дифференцирование
2.1 Посчитайте $df(x)$ и $\nabla f(x)$ для функции $f(x) = \log(x^\top Ax)$ и выберите правильный вариант ответа

$$g(y) = \log(y),\ dg(y) = \frac{dy}{y}$$

$$h(x) = x^\top A x = (A^\top x)^\top x = \langle A^\top x, x\rangle,\ dh(x) = \langle (A^\top + A)x, dx\rangle$$

$$df(x) = \frac{\langle (A^\top + A)x, dx\rangle}{x^\top A x} = \frac{((A^\top + A)x)^\top dx}{x^\top A x} = \frac{x^\top (A + A^\top) dx}{x^\top A x}, \quad \nabla f(x) = \frac{(A + A^\top)x}{x^\top A x}$$

Ответ: b

2.2 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

$$df(x) = d\left(\frac{1}{p}\|x\|_2^p\right) = \frac{1}{p} d\langle x, x\rangle^{\frac{p}{2}} = \frac{1}{p}\cdot\frac{p}{2}\langle x, x\rangle^{\frac{p}{2}-1} d\langle x, x\rangle = \frac{1}{2}\|x\|_2^{p-2} 2\langle x, dx\rangle = \langle \|x\|_2^{p-2} x, dx\rangle$$

$$\nabla f(x) = \|x\|_2^{p-2} x, \quad df(x) = \|x\|_2^{p-2} x^\top dx$$

$$d^2 f(x) = d(\|x\|_2^{p-2}\langle x, dx_1\rangle) = d(\|x\|_2^{p-2})\langle x, dx_1\rangle + \|x\|_2^{p-2} d\langle x, dx_1\rangle =$$

$$= d(\langle x, x\rangle^{\frac{p}{2}-1})\langle x, dx_1\rangle + \|x\|_2^{p-2}\langle dx_2, dx_1\rangle =$$

$$= (p-2)\|x\|_2^{p-4}\langle x, dx_2\rangle\langle x, dx_1\rangle + \|x\|_2^{p-2}\langle dx_2, dx_1\rangle =$$

$$= \langle ((p-2)\|x\|_2^{p-4} x x^\top + \|x\|_2^{p-2} I) dx_1, dx_2\rangle$$

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

Ответ: b, e

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

Для начала найдем гессиан одного слагаемого

$$g(x) = \log(1 + \exp(a_i^\top x)) = \log(1 + \exp\langle a_i, x\rangle)$$

$$h(x) = \log(1 + \exp(x)),\ dh(x) = \frac{\exp(x)}{1 + \exp(x)} dx = \sigma(x) dx$$

$$dg(x) = \sigma\langle a_i, x\rangle \cdot d\langle a_i, x\rangle = \sigma\langle a_i, x\rangle\langle a_i, dx\rangle = \langle a_i \cdot \sigma\langle a_i, x\rangle, dx\rangle$$

$$d^2 g(x) = d(\langle a_i \cdot \sigma\langle a_i, x\rangle, dx_1\rangle) = d(\sigma\langle a_i, x\rangle)\langle a_i, dx_1\rangle = \sigma'(\langle a_i, x\rangle)\langle a_i, dx_2\rangle\langle a_i, dx_1\rangle$$

$$\nabla^2 g(x) = \sigma'(\langle a_i, x\rangle) a_i a_i^\top$$

$$\sigma'(x) = \sigma(x)(1-\sigma(x)) = \frac{\exp(x)}{1+\exp(x)} - \frac{\exp(x)^2}{(1+\exp(x))^2}$$

Гессиан для слагаемого с μ найдем из формулы, доказанной в предыдущем упражнении

$$\mu \cdot \nabla^2 \frac{1}{2}\|x\|_2^2 = \mu I$$

Так как гессиан обладает свойством линейности, то итоговый гессиан равен

$$\nabla^2 f(x) = \frac{1}{n}\sum_{i=1}^{n}\left(-\frac{\exp(\langle a_i, x\rangle)^2}{(1+\exp(\langle a_i, x\rangle))^2}a_i a_i^\top + \frac{\exp(\langle a_i, x\rangle)}{1+\exp(\langle a_i, x\rangle)}a_i a_i^\top\right) + \mu I =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(-\frac{\exp(a_i^\top x)^2}{(1+\exp(a_i^\top x))^2}a_i a_i^\top + \frac{\exp(a_i^\top x)}{1+\exp(a_i^\top x)}a_i a_i^\top\right) + \mu I$$

Данная функция отличается от функции в вопросе квадратом у экспоненты в числителе первой дроби

Ответ: нет, не верно

3. Выпуклые функции. Докажите выпуклость следующих функций

а) $f(x) = \sum\limits_{i=1}^{n} e^{x_i}$

Т.к. экспонента функция выпуклая на $\mathbb{R}$ и линейная комбинация с положительными коэффициентами выпуклых функций сохраняет выпуклость, то $f(x)$ выпукла

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

$$\nabla^2 f(x) = \sum_{i=1}^{n} w_i \sigma'(a_i^\top x) a_i a_i^\top + \mu I$$

$$a_i a_i^\top \succeq 0,\ w_i \geq 0,\ \sigma'(x) > 0,\ \mu > 0 \Rightarrow \nabla^2 f(x) \succeq 0 \Rightarrow f(x) - \text{выпуклая}$$

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

$$\partial f(x) = \partial(\max(-c^\top x, c^\top x)) = \begin{cases} -c, \quad c^\top x < 0 \\ \operatorname{conv}(-c; c), \quad c^\top x = 0 \\ c, \quad c^\top x > 0 \end{cases}$$

$$= \begin{cases} -c, \quad c^\top x < 0 \\ -\alpha c + (1-\alpha)c, \quad c^\top x = 0, \ \alpha \in [0;1] \\ c, \quad c^\top x > 0 \end{cases}$$

$$= \begin{cases} -c, \quad c^\top x < 0 \\ (1-2\alpha)c, \quad c^\top x = 0, \ \alpha \in [0;1] \\ c, \quad c^\top x > 0 \end{cases}$$

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

$$\nabla f(x) = 2x$$

$$\|\nabla f(x) - \nabla f(y)\|_2 = \|2x - 2y\|_2 = 2\|x-y\|_2 \leq L\|x-y\|_2$$

$$f(x) - 2\text{-гладкая функция}$$

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle = 2\langle x-y, x-y\rangle = 2\|x-y\|_2^2 \geq \mu\|x-y\|_2^2$$

$$f(x) - \text{сильно выпуклая функция с константой } \mu = 2$$

Ответ: 2, 2

Дополнительные задачи

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

$$f(x) = \begin{cases} 1-x & x \leq 0 \\ e^x & x > 0 \end{cases} \quad \partial f(x) = \begin{cases} -1 & x < 0 \\ [-1, 1] & x = 0 \\ e^x & x > 0 \end{cases} \quad (\partial f(0) = \mathrm{conv}\{-1, e^0\} = [-1, 1])$$

5. Какие условия верны для сильно выпуклой функции $f(x)$ с константой μ? Ответы обосновать.

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

По определению сильно выпуклой функции $\alpha \in [0;1]$, а не $\alpha \in \mathbb{R}$. Так что это условие неверно

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

Это эквивалентное определение сильно выпуклой функции для дифференцируемых функций, оно верно

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

Перед нормой разницы градиентов должен стоять коэффициент $\frac{1}{2\mu}$, условие не верно

- если $\nabla f(x^*) = 0$, то $\|\nabla f(x)\|_2^2 \leq 2\mu(f(x) - f(x^*)) \quad \forall x \in \mathbb{R}^n$

Если бы в неравенсте был знак $\geq$, то это было бы условие Поляка-Лоясевича. Условие неверно

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

Условие верно (теорема 2.1.10 из книги Нестерова)

- если $f(x)$ - дважды непрерывно дифференцируема, то $\nabla^2 f(x) \succeq \mu I_n \quad \forall x \in \mathbb{R}^n$

Условие верно (теорема 2.1.11 из книги Нестерова)

- $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n$

Условие верно (теорема 2.1.9 из книги Нестерова)

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

Должен быть знак $\geq$ (Теорема 2.1.8 из книги Нестерова)

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

Если в задании под $\|\cdot\|_2$ имеется ввиду спектральная норма (часто для матриц ее так обозначают), то данное неравенство напрямую следует из того, что $\nabla^2 f(x) \preceq L I_n$

Если же подразумевается евклидова норма (то есть для матрицы норма Фробениуса $\|\cdot\|_F$), то используя неравенство $\|A\|_F \leq \sqrt{n}\|A\|_2$ получим $\|\nabla^2 f(x)\|_F \leq \sqrt{n}\|\nabla^2 f(x)\|_2 \leq \sqrt{n} L$