Методы оптимизации в ML, весна 2023. Домашнее задание 1

1. Выпуклые множества (1 балл)

a) $X \cup Y$: Не сохраняет выпуклость; Контрпример:
$X = \{x : 0 \leqslant x \leqslant 1\}$ – отрезок от 0 до 1,
$Y = \{y : 2 \leqslant x \leqslant 3\}$ – отрезок от 2 до 3,
Пусть $x = 1, y = 2$, тогда $z = \alpha \cdot 1 + (1 - \alpha) \cdot 2 = 2 - \alpha \notin X \cup Y$ при $\alpha = 0.5$

b) $X \times Y$: Сохраняет выпуклость; Док-во:
Пусть $a = (x_1, y_1) \in X \times Y$, $b = (x_2, y_2) \in X \times Y$
Тогда $\alpha a + (1 - \alpha)b = (\alpha x_1 + (1 - \alpha)x_2, \alpha y_1 + (1 - \alpha)y_2) \in X \times Y$, тк X и Y выпуклые.

c) $aX + bY$: Сохраняет выпуклость; Док-во:
Пусть $x_1, x_2 \in X, y_1, y_2 \in Y$.
Тогда
$x_3 = \alpha x_1 + (1 - \alpha)x_2 \in X$,
$y_3 = \alpha y_1 + (1 - \alpha)y_2 \in Y$,
$ax_1 + by_1 \in aX + bY$,
$ax_2 + by_2 \in aX + bY$,
$\alpha(ax_1 + by_1) + (1 - \alpha)(ax_2 + by_2) = \alpha ax_1 + (1 - \alpha)ax_2 + \alpha by_1 + (1 - \alpha)by_2) = ax_3 + by_3 \in aX + bY$.

d) aX: Сохраняет выпуклость; Док-во: как частный случай c) при $b = 0$.

e) X^c: Не сохраняет выпуклость; Контрпример:
Пусть
$X = [0; 1]$ – отрезок от 0 до 1,
Тогда
$X^c = (-\inf; 0) \cup (1; \inf)$,
при $x = -1 \in X^c, y = 2 \in X^c$,
$z = \alpha \cdot -1 + (1 - \alpha) \cdot 2 = 2 - 3\alpha \notin X^c$ при $\alpha = 0.5$

2. Матрично-векторное дифференцирование (3 балла)

1. $f(x) = \log(x^T Ax)$:
$d(x^T Ax) = \langle (A + A^T)x, dx\rangle$ – док-во на лекции.
$df(x) = d(\log(x^T Ax)) = \frac{d(x^T Ax)}{x^T Ax} = \frac{\langle (A+A^T)x, dx\rangle}{x^T Ax}$,
$\nabla f(x) = \frac{(A+A^T)x}{x^T Ax}$.
Следовательно, правильный вариант b.

2. $f(x) = \frac{1}{p}||x||_2^p$, $p > 1$:
$d||x||_2 = ||x||_2^{-1}\langle x, dx\rangle$ – док-во на лекции.
$df(x) = ||x||_2^{p-1} d||x||_2 = ||x||_2^{p-2}\langle x, dx\rangle$,

$\nabla f(x) = ||x||_2^{p-2}x.$
$d^2 f(x) = d(||x||_2^{p-2}\langle x, dx_1\rangle) =$
{По правилу Лейбница}
$= d(||x||_2^{p-2})\langle x, dx_1\rangle + ||x||_2^{p-2} d(\langle x, dx_1\rangle) = (p-2)||x||_2^{p-3}||x||_2^{-1}\langle x, dx_2\rangle\langle x, dx_1\rangle +$
$||x||_2^{p-2}\langle dx_2, dx_1\rangle = \langle((p-2)||x||_2^{p-4}xx^T + ||x||_2^{p-2}I_n)dx_1, dx_2\rangle,$
$\nabla^2 f(x) = ((p-2)||x||_2^{p-4}xx^T + ||x||_2^{p-2}I_n).$
Следовательно, правильные варианты: b, e.

3. $f(x) = \frac{1}{n}\sum \log(1 + e^{a_i^T x}) + \frac{\mu}{2}||x||_2^2$:
   $\nabla \frac{\mu}{2}||x||_2^2 = \mu\nabla\frac{1}{2}||x||_2^2 = \mu I_n$ – как частный случай 2 при $p = 2$.
   $\sigma(x)$ – сигмоида, $\sigma'(x) = \sigma(x)(1 - \sigma(x))$.
   $d\langle a, x\rangle = \langle a, dx\rangle$ – док-во на лекции.
   $d\log(1 + e^{ax}) = \frac{d(1+e^{ax})}{1+e^{ax}} = \frac{e^{ax}\langle a,dx\rangle}{1+e^{ax}} = \sigma(ax)\langle a, dx\rangle.$
   $d^2\log(1 + e^{ax}) = d\sigma(ax)\langle a, dx_1\rangle = \sigma(ax)(1 - \sigma(ax))\langle a, dx_2\rangle\langle a, dx_1\rangle =$
   $\langle(\sigma(ax)(1 - \sigma(ax))aa^T)dx_1, dx_2\rangle,$
   $\nabla^2\log(1 + e^{ax}) = \sigma(ax)(1 - \sigma(ax))aa^T.$
   $\nabla^2 f(x) = \frac{1}{n}\sum \sigma(ax)aa^T - \sigma^2(ax)aa^T + \mu I_n = \frac{1}{n}\sum \frac{e^{ax}}{1+e^{ax}}aa^T - \frac{e^{2ax}}{(1+e^{ax})^2}aa^T +$
   $\mu I_n.$
   Следовательно, не верно.

3. Выпуклые функции (3 балла)

(a) $f(x) = \sum e^{x_i}$; Док-во:
e^x – выпукла, тк $(e^x)'' = e^x > 0$.
$f(x)$ – выпукла, как сумма выпуклых функций.

(b) $f(x) = \frac{||Ax-b||^2}{1-||x||^2}, x \in \{x : ||x||^2 < 1\}$; Док-во:
$||Ax - b||$ – выпукла как афинная подстановка в выпуклую функцию (норма выпукла, док-во на лекции), неотрицательна по свойству нормы, бесконечно дифференцируема.
$1 - ||x||^2$ – вогнута, тк норма выпукла, но после умножения на -1 становится вогнутой, положительна по условию.
$f(x)$ – выпукла, как частный случай (c).

(c) $F(x) = \frac{f^2(x)}{g(x)}$,
f(x) – выпукла, неотрицательна, дважды дифф-ма,
g(x) – вогнута, положительна;
Док-во:
$F(\alpha x + (1-\alpha)y) = \frac{(f(\alpha x+(1-\alpha)y))^2}{g(\alpha x+(1-\alpha)y)} \leqslant \frac{(\alpha f(x)+(1-\alpha)f(y))^2}{\alpha g(x)+(1-\alpha)g(y)}$
{По нер-ву Коши-Буняковского https://en.wikipedia.org/wiki/Sedrakyan%27s_inequality}
$\leqslant \alpha\frac{f^2(x)}{g(x)} + (1-\alpha)\frac{f^2(y))}{g(y)}$, функция $F(x)$ выпукла.

(d) aX:

(e) $f(x) = \sum w_i \ln(1 + e^{a_i^T x}) + \frac{\mu}{2}||x||_2^2, \mu, w_i > 0$; Док-во:
$(\ln(1 + e^x))'' = \frac{e^x}{(1+e^x)^2} > 0$, следовательно функ-я выпукла.
$a_i^T x$ – скалярное произведение выпукло(док-во на лекции).
$\ln(1 + e^{a_i^T x})$ – выпукла как суперпозиция выпуклых функ-й.
$||x||_2^2$ – выпукла(док-во на лекции).
$f(x)$ – выпукла как сумма выпуклых функций с полож. весами.

(f) $f(x) = \ln \sum e^{\max\{0,x_i\}^2}$; Док-во:
$\max\{0, x_i\}$ – выпукла как максимум выпуклых функ-й.
x^2 – выпукла, тк вторая производная неотрицательна.
$g(\alpha x + (1 - \alpha)y) = \ln \sum e^{\alpha x_i + (1-\alpha)y_i} = \ln \sum (e^{x_i})^\alpha (e^{y_i})^{(1-\alpha)}$
По нер-ву Гельдера при $p = \frac{1}{\alpha}, q = \frac{1}{1-\alpha}$
$\leqslant \alpha \ln \sum e^{x_i} + (1 - \alpha) \ln \sum e^{y_i}$ – $g(x)$ выпукла.
$f(x)$ – выпукла как суперпозиция выпуклых функ-й.

4. (1 балл) Посчитайте субдифференциал функции $f(x) = |c^T x|, x \in R^n$

Аналогично примеру из лекции с функцией $f(x) = \sum_1^m |a^T x - b|$:
$f(x) = |c^T x| = \max\{c^T x, -c^T x\}$ – выпуклая функ-я.

$$\partial f(x) = \begin{cases} \text{c}, & \text{if } c^T x > 0. \\ \text{-c}, & \text{if } c^T x < 0. \\ \text{[-c, c] – выпуклая оболочка -c и c} & \text{if } c^T x = 0. \end{cases} \tag{1}$$

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

Аналогично примеру из лекции с функцией $f(x) = \sum_1^m |a^T x - b|$:
e_i – нулевой вектор с 1 на iм месте.
$f(x) = ||x||_1 = \sum |x_i| = \sum |e_i^T x| = \sum \max\{e_i^T x, -e_i^T x\}$ – выпуклая функ-я.

$$\partial |e_i^T x| = \begin{cases} e_i, & i \in \{i : e_i^T x > 0\} = I_+(x). \\ -e_i, & i \in \{i : e_i^T x < 0\} = I_-(x). \\ [-e_i, e_i] \text{ – вып. оболочка} & i \in \{i : e_i^T x = 0\} = I_0(x). \end{cases} \tag{2}$$

$\partial f(x) = \sum_{i \in I_+(x)} e_i - \sum_{i \in I_-(x)} e_i + \sum_{i \in I_0(x)} [-e_i, e_i]$.

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

$f(x) = ||x||_2^2 = x^T x$,
$\nabla f(x) = 2x$ – док-во на лекции.
$||\nabla f(x) - \nabla f(y)|| = 2||x - y||$ – получаем по определению $L = 2$.
Для определения μ используем Теорему 6 из лекции 2:
$\langle \nabla f(x) - \nabla f(y), x - y\rangle = 2\langle x - y, x - y\rangle = 2||x - y||^2$ – получаем $\mu = 2$.