Методы оптимизации

Домашнее задание 1

Арутюнян Саро, Б05-924

Выпуклые множества.

Задача 1. Является ли выпуклым множество дискретных случайных величин с ограниченной дисперсией ($\mathbb{V}x \leq \alpha$)?

Решение. Пусть для дискретных случайных величин x, y верно $\mathbb{V}x \leq \alpha$, $\mathbb{V}y \leq \alpha$. Тогда $\forall\theta \in [0, 1]$

$$\mathbb{V}(\theta x + (1-\theta)y) = \mathbb{E}(\theta x + (1-\theta)y)^2 - \left(\mathbb{E}(\theta x + (1-\theta)y)\right)^2 =$$

$$= \mathbb{E}(\theta^2 x^2 + (1-\theta)^2 y + 2\theta(1-\theta)xy) - \theta^2(\mathbb{E}x)^2 - (1-\theta)^2(\mathbb{E}y)^2 - 2\theta(1-\theta)\mathbb{E}x\mathbb{E}y =$$

$$= \theta^2(\mathbb{E}x^2 - (\mathbb{E}x)^2) + (1-\theta)^2(\mathbb{E}y^2 - (\mathbb{E}y)^2) + 2\theta(1-\theta)(\mathbb{E}xy - \mathbb{E}x\mathbb{E}y) =$$

$$= \theta^2\mathbb{V}x + (1-\theta)^2\mathbb{V}y + 2\theta(1-\theta)\operatorname{cov}(x, y) \leq$$

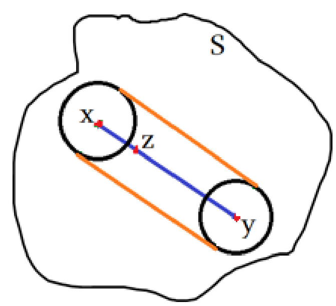
$$\leq \theta^2\mathbb{V}x + (1-\theta)^2\mathbb{V}y + 2\theta(1-\theta)\sqrt{\mathbb{V}x\mathbb{V}y} =$$

$$= \left(\theta\sqrt{\mathbb{V}x} + (1-\theta)\sqrt{\mathbb{V}y}\right)^2 \leq \alpha.$$

Значит рассматриваемое множество выпукло.

Задача 2. Докажите, что если множество выпуклое, то его внутренность тоже выпуклое множество. Верно ли обратное утверждение?

Решение. Пусть S — данное множество. Нужно доказать, что для любых двух точек x, y из $\operatorname{Int} S$ содержит любую точку z отрезка xy. Из $x, y \in \operatorname{Int} S$ следует, что $\exists\delta > 0$: $\overline{B_\delta(x)} \subset S$, $\overline{B_\delta(y)} \subset S$, где $B_r(a)$ — шар с центром a и радиусом r. По выпуклости S для любых $u \in \overline{B_\delta(x)}$ и $v \in \overline{B_\delta(x)}$ отрезок uv лежит в S. Значит отрезок xy лежит в S со своей δ-окрестностью (т.е. с точками, на расстоянии меньше чем δ хотя бы с одной из точек отрезка xy), т.е. все точки отрезка xy внутренние для S. Значит $\operatorname{Int} S$ — выпуклое.



Обратное утверждение неверно. Рассмотрим, например множество

$$Q = \{(x, y) | -1 \leq x \leq 1, -1 \leq y < 1\} \cup \{(1, 1), (-1, 1)\}$$

(иными словами — квадрат без внутренности одной стороны). Тогда ясно, что его внутренность выпукла. Но само Q не выпукло: $(1, 1), (-1, 1) \in Q$, но $\frac{1}{2}(1, 1) + \frac{1}{2}(-1, 1) = (0, 1) \notin Q$.

Задача 3. Докажите, что гиперболическое множество $H = \{x \in \mathbb{R}^n | \prod_{i=1}^n x_i \geq 1\}$ выпуклое.

Решение (в предположении, что $x_i \geq 0$). $\forall x, y \in H$, $\forall\theta \in [0, 1]$

$$\prod_{i=1}^n(\theta x_i + (1-\theta)y_i) \geq \prod_{i=1}^n x_i^\theta y_i^{1-\theta} = \left(\prod_{i=1}^n x_i\right)^\theta \left(\prod_{i=1}^n y_i\right)^{1-\theta} \geq 1.$$

На самом деле, если разрешить координатам векторов быть также отрицательными, то H не выпукло. Действительно, в случае $n = 2$ имеем $(1, 1) \in H$, $(-1, -1) \in H$, но $\frac{1}{2}(1, 1) + \frac{1}{2}(-1, -1) = (0, 0) \notin H$.

Задача 4. Докажите, что множество $S \subseteq \mathbb{R}^n$ выпуклое тогда и только тогда, когда

$$(\alpha+\beta)S = \alpha S + \beta S \quad \forall \alpha, \beta \geq 0.$$

Решение. Имеем

$$\alpha S = \{\alpha x | x \in S\}, \qquad \beta S = \{\beta y | y \in S\},$$

$$\alpha S + \beta S = \{\alpha x + \beta y | x, y \in S\}, \qquad (\alpha+\beta)S = \{(\alpha+\beta)x | x \in S\},$$

поэтому $(\alpha+\beta)S \subseteq \alpha S + \beta S$.

Пусть S выпукло. Докажем, что $(\alpha+\beta)S = \alpha S + \beta S \quad \forall \alpha, \beta \geq 0$. Для этого достаточно доказать, что $(\alpha+\beta)S \supseteq \alpha S + \beta S$. Пусть $z \in \alpha S + \beta S$. Тогда $z = \alpha x + \beta y$ для некоторых $x, y \in S$, $\alpha, \beta \geq 0$. Нужно найти такой $w \in S$, чтобы $\alpha x + \beta y = (\alpha+\beta)w$. Можно взять, например, $w = \frac{\alpha}{\alpha+\beta}x + \frac{\beta}{\alpha+\beta}y$. Тогда по выпуклости S будет верно, что $w \in S$. Таким образом $(\alpha+\beta)S \supseteq \alpha S + \beta S$, что и нужно было доказать.

Наоборот, пусть $(\alpha+\beta)S = \alpha S + \beta S \; \forall \alpha, \beta \geq 0$. Тогда докажем, что S выпукло. Имеем $\forall x, y \in S$, $\forall \alpha, \beta \geq 0$ $\exists z \in S \colon (\alpha+\beta)z = \alpha x + \beta y \iff \forall x, y \in S \; \forall \alpha, \beta \geq 0 \; \frac{\alpha}{\alpha+\beta}x + \frac{\beta}{\alpha+\beta}y \in S$, что и есть определение выпуклости.

Задача 5. Пусть $x \in \mathbb{R}^n$ случайная величина с заданным распределением вероятностей $P(x_i = a_i) = p_i$, где $i = 1, \dots, n$ и $a_1 < \dots < a_n$. Говорят, что вектор вероятностей исходов $p \in \mathbb{R}^n$ принадлежит вероятностному симплексу, т.е.

$$P = \{p | 1^T p = 1, p \succcurlyeq 0\}.$$

Являются ли следующие множества выпуклыми?

(a) $P(x > \alpha) \leq \beta$;
(b) $\mathbb{E}(|x|^{2021}) \leq \alpha \mathbb{E}|x|$;
(c) $\mathbb{E}x^2 \geq \alpha$;
(d) $\mathbb{V}x \geq \alpha$.

Матрично-векторное дифференцирование.

Задача 1. Вычислить $\nabla \|Ax\|_2$.

Решение. $d\|Ax\|_2 = d\langle Ax, Ax\rangle^{1/2} = \frac{1}{2}\langle Ax, Ax\rangle^{-1/2} d\langle Ax, Ax\rangle = \frac{\|Ax\|_2}{2} \cdot 2\langle Ax, Adx\rangle = \|Ax\|_2 \langle A^T Ax, dx\rangle$,

$$\nabla \|Ax\|_2 = \|Ax\|_2 A^T Ax.$$

Задача 2. Вычислить $\nabla^2 f(x)$, где $f(x) = \log\det X$.

Решение. Имеем

$$df(x) = \frac{d\det X}{\det X} = \frac{\det X \cdot \langle X^{-T}, dX\rangle}{\det X} = \langle X^{-T}, dX\rangle = \operatorname{tr}(X^{-1}dX).$$

$$d^2 f(x) = d\langle X^{-T}, dX_1\rangle = \langle d(X^T)^{-1}, dX_1\rangle = \langle -X^{-T} d(X^T) X^{-T}, dX_1\rangle = \langle -X^{-T}(dX_2)^T X^{-T}, dX_1\rangle \dots$$

По формуле Тейлора

$$f(X+H) = f(X) + Df(X)[H] + \frac{1}{2}D^2 f(X)[H,H] + o(\|H\|^2)$$

$$\frac{1}{2}D^2 f(X)[H,H] + o(\|H\|^2) = \log\det(X+H) - \log\det X - \operatorname{tr}(X^{-1}H) =$$

$$= \log\det X(I_n + X^{-1}H) - \log\det X - \operatorname{tr}(X^{-1}H) = \log\det(I_n + X^{-1}H) - \operatorname{tr}(X^{-1}H).$$

Пусть $\lambda_1, \dots, \lambda_n$ — собственные значения $X^{-1}H$. Тогда $1+\lambda_1, \dots, 1+\lambda_n$ — собственные значения $I_n + X^{-1}H$ и $\det(I_n + X^{-1}H) = \prod_i (1+\lambda_i)$. Отсюда

$$\log\det(I_n + X^{-1}H) - \operatorname{tr}(X^{-1}H) = \sum_{i=1}^{n} \log(1+\lambda_i) - \operatorname{tr}(X^{-1}H) = \sum_{i=1}^{n} \left(\lambda_i + \frac{\lambda_i^2}{2} + o(|\lambda|^2)\right) - \operatorname{tr}(X^{-1}H) =$$

$$= \operatorname{tr}(X^{-1}H) + \frac{\operatorname{tr}(X^{-1}HX^{-1}H)}{2} + o(\|H\|^2) - \operatorname{tr}(X^{-1}H) = \frac{\operatorname{tr}(X^{-1}HX^{-1}H)}{2} + o(\|H\|^2)$$

Значит

$$D^2 f(X)[H,H] = \operatorname{tr}(X^{-1}HX^{-1}H).$$

Задача 3. Вычислить $\frac{\partial}{\partial X}\|X\|_F^2$, $X \in \mathbb{R}^{n\times m}$, где $\|X\|_F = \sqrt{\langle X, X\rangle} = \sqrt{\operatorname{tr} X^T X}$.

Решение. Имеем

$$\frac{\partial}{\partial X}\|X\|_F^2 = \lim_{t\to 0+} \frac{\|X + t\cdot dX\|_F^2 - \|X\|_F^2}{t} = \lim_{t\to 0+} \frac{\langle X + t\cdot dX, X + t\cdot dX\rangle - \langle X, X\rangle}{t} =$$

$$= \lim_{t\to 0+} (2\langle X, dX\rangle + t\langle dX, dX\rangle) = 2\langle X, dX\rangle.$$

Задача 4. В машинном обучении очень часто используют следующую функцию потерь: $L = \frac{1}{n}\sum_{i=1}^{n}\|y_i - \hat{y}\|_2^2$, где $\hat{y} = Wx_i + b$.

- n — количество $x_i \in \mathbb{R}^l$ — данные обучающей выборки;
- $y_i \in \mathbb{R}^m$ — истинное предсказание на векторе x_i;
- $\hat{y} \in \mathbb{R}^m$ — предсказание модели на векторе x_i;
- В данном случае у нас линейная модель, где $W \in \mathbb{R}^{l\times m}$, $b \in \mathbb{R}^m$.

Вычислить $\frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}$ для $n = 1$.

Решение. Имеем

$$\frac{\partial L}{\partial W} = \lim_{t\to 0+} \frac{\|y_1 - (W + t\cdot dW)x_1 - b\|^2 - \|y_1 - Wx_1 - b\|^2}{t} =$$

$$= \lim_{t\to 0+} \frac{\langle y_1 - (W + t\cdot dW)x_1 - b, y_1 - (W + t\cdot dW)x_1 - b\rangle - \langle y_1 - Wx_1 - b, y_1 - Wx_1 - b\rangle}{t} =$$

$$= \lim_{t\to 0}(-2\langle y_1 - Wx_1 - b, dW\rangle + t\langle dW, dW\rangle) = -2\langle y_1 - Wx_1 - b, dW\rangle.$$

Аналогично

$$\frac{\partial L}{\partial b} = -2\langle y_1 - Wx_1 - b, db\rangle.$$