Выпуклость. Сильная выпуклость.

Семинар

Оптимизация для всех! ЦУ

Напоминание с лекции

Выпуклые множества

Отрезок

Пусть x_1, x_2 два вектора в $\mathbb{R}^n$. Тогда отрезок, проходящий через них, определяется следующим образом:

$$x = \theta x_1 + (1 - \theta) x_2, \ \theta \in [0, 1]$$

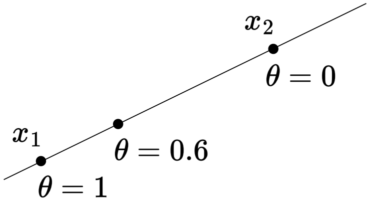


Рис. 1: Иллюстрация отрезка между точками x_1, x_2

Выпуклое множество

Множество S называется **выпуклым**, если для любых x_1, x_2 из S отрезок между ними также лежит в S, т.е.

$$\forall \theta \in [0,1], \ \forall x_1, x_2 \in S : \theta x_1 + (1-\theta) x_2 \in S$$

Example

Любое аффинное множество, луч, отрезок - все они являются выпуклыми множествами.

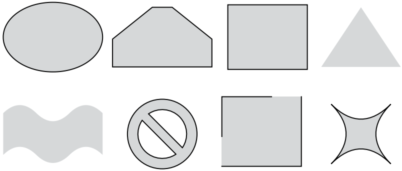


Рис. 2: Верх: примеры выпуклых множеств. Низ: примеры невыпуклых множеств.

Задача 1

Question

Докажите, что шар в $\mathbb{R}^n$ (т.е. множество $\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$) - является выпуклым.

Задача 2

Question

Является ли полоса $\{x \in \mathbb{R}^n \mid \alpha \leq a^\top x \leq \beta\}$ выпуклой?

Задача 3

Question

Пусть S такое, что $\forall x, y \in S \to \frac{1}{2}(x+y) \in S$. Является ли это множество выпуклым?

Задача 4

Question

Является ли множество $S = \{x \mid x + S_2 \subseteq S_1\}$, где $S_1, S_2 \subseteq \mathbb{R}^n$ с выпуклым S_1, выпуклым?

Функции

Выпуклая функция

Функция $f(x)$, **определенная на выпуклом множестве** $S \subseteq \mathbb{R}^n$, называется **выпуклой** на S, если:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

для любых $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$.

Если вышеуказанное неравенство выполняется как строгое неравенство $x_1 \neq x_2$ и $0 < \lambda < 1$, то функция называется **строго выпуклой** на S.

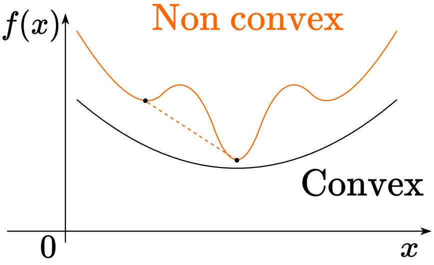


Рис. 3: Разница между выпуклой и невыпуклой функцией

Сильная выпуклость

$f(x)$, **определенная на выпуклом множестве** $S \subseteq \mathbb{R}^n$, называется μ-сильно выпуклой (сильно выпуклой) на S, если:

$$
f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{\mu}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2
$$

для любых $x_1, x_2 \in S$ и $0 \leq \lambda \leq 1$ для некоторого $\mu > 0$.

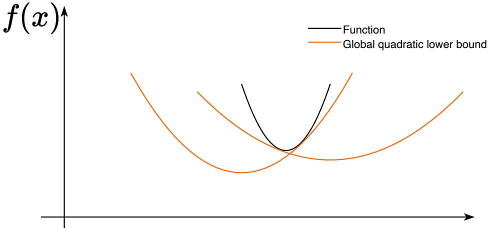


Рис. 4: Сильно выпуклая функция больше или равна квадратичной аппроксимации Тейлора в любой точке

Критерии выпуклости

Дифференциальный критерий выпуклости первого порядка

Дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является выпуклой тогда и только тогда, когда $\forall x, y \in S$:

$$f(y) \geq f(x) + \nabla f^T(x)(y - x)$$

Пусть $y = x + \Delta x$, тогда критерий станет более удобным:

$$f(x + \Delta x) \geq f(x) + \nabla f^T(x)\Delta x$$

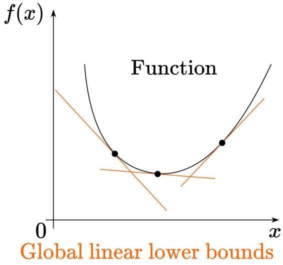


Рис. 5: Выпуклая функция больше или равна линейной аппроксимации Тейлора в любой точке

Дифференциальный критерий выпуклости второго порядка

Дважды дифференцируемая функция $f(x)$ определенная на выпуклом множестве $S \subseteq \mathbb{R}^n$ является μ-сильно выпуклой тогда и только тогда, когда $\forall x \in \mathbf{int}(S) \neq \emptyset$:

$$\nabla^2 f(x) \succeq \mu I$$

Другими словами:

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

Эксперимент с JAX

Почему выпуклость и сильная выпуклость важны? Проверьте простой код.

Задача 5

Question

Докажите, что $f(x) = \|x\|$ является выпуклой на $\mathbb{R}^n$.

Question

Докажите, что $f(x) = x^\top Ax$, где $A \succeq 0$ - является выпуклой на $\mathbb{R}^n$.

Задача 6

Question

Докажите, что если $f(x)$ является выпуклой на $\mathbb{R}^n$, то $\exp(f(x))$ является выпуклой на $\mathbb{R}^n$.

Задача 7

Question

Доказите, что если $f(x)$ является выпуклой неотрицательной функцией и $p \geq 1$, то $g(x) = f(x)^p$ является выпуклой.

Задача 8

Question

Докажите, что если $f(x)$ является вогнутой положительной функцией над выпуклым S, то $g(x) = \frac{1}{f(x)}$ является выпуклой.

Question

Докажите, что следующая функция является выпуклой на множестве всех положительных знаменателей

$$f(x) = \cfrac{1}{x_1 - \cfrac{1}{x_2 - \cfrac{1}{x_3 - \cfrac{1}{\dots}}}}, x \in \mathbb{R}^n$$

Задача 9

Question

Пусть $S = \{x \in \mathbb{R}^n \mid x \succ 0, \|x\|_\infty \leq M\}$. Докажите, что $f(x) = \sum_{i=1}^n x_i \log x_i$ является $\frac{1}{M}$-сильно выпуклой.

Условие Поляка - Лоясиевича

Условие Поляка - Лоясиевича

Условие Поляка - Лоясиевича выполняется, если выполняется следующее условие для некоторого $\mu > 0$,

$$
\|\nabla f(x)\|^2 \geq \mu(f(x) - f^*) \forall x
$$

Пример функции, которая удовлетворяет условию Поляка - Лоясиевича, но не является выпуклой.

$$
f(x, y) = \frac{(y - \sin x)^2}{2}
$$

Пример невыпуклой функции, удовлетворяющей условию Поляка - Лоясиевича Open in Colab.

Практические примеры

Логистическая регрессия

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n.$

Логистическая регрессия

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n.$

Найти

Найти функцию, которая переводит объект x в вероятность $p(y = 1|x)$:
$p : \mathbb{R}^m \to (0, 1)$, $p(x) \equiv \sigma(x^T w) = \frac{1}{1+\exp(-x^T w)}$

Логистическая регрессия

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{0,1\}^n.$

Найти

Найти функцию, которая переводит объект x в вероятность $p(y=1|x)$:
$p : \mathbb{R}^m \to (0,1)$, $p(x) \equiv \sigma(x^T w) = \frac{1}{1+\exp(-x^T w)}$

Критерий

Двоичная кросс-энтропия (лог-потери): $L(p, X, y) = -\sum_{i=1}^n y_i \log p\left(X_i\right) + \left(1-y_i\right) \log \left(1-p\left(X_i\right)\right),$ которая минимизируется относительно w.

Логистическая регрессия

Дано

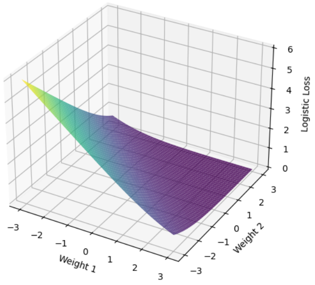
$X \in \mathbb{R}^{m \times n}, y \in \{0, 1\}^n$.

Найти

Найти функцию, которая переводит объект x в вероятность $p(y = 1|x)$:
$p : \mathbb{R}^m \to (0, 1)$, $p(x) \equiv \sigma(x^T w) = \frac{1}{1+\exp(-x^T w)}$

Критерий

Двоичная кросс-энтропия (лог-потери): $L(p, X, y) = -\sum_{i=1}^n y_i \log p(X_i) + (1 - y_i) \log (1 - p(X_i))$, которая минимизируется относительно w.



Мы можем сделать эту задачу μ-сильно выпуклой, если рассмотрим регуляризованную логистическую потерю как критерий: $L(p, X, y) + \frac{\mu}{2}\|w\|_2^2$.

Рассотрите эксперименты по логистической регрессии.

Метод опорных векторов (SVM)

Дано

$$X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n.$$

Метод опорных векторов (SVM)

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n.$

Найти

Найти гиперплоскость, которая максимизирует маржу между двумя классами:
$f : \mathbb{R}^m \to \{-1, 1\}, \; f(x) = \text{sign}(w^T x + b).$

Метод опорных векторов (SVM)

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n.$

Найти

Найти гиперплоскость, которая максимизирует маржу между двумя классами:
$f : \mathbb{R}^m \to \{-1, 1\},\ f(x) = \text{sign}(w^T x + b).$

Критерий

Шарнирная функция потерь:
$L(w, X, y) = \frac{1}{2}\|w\|_2^2 + C \sum_{i=1}^n \max(0, 1 - y_i(X_i^T w + b)),$
которая минимизируется относительно w и b.

Эта задача является сильно выпуклой из-за квадратичной евклидовой нормы.

Метод опорных векторов (SVM)

Дано

$X \in \mathbb{R}^{m \times n}, y \in \{-1, 1\}^n.$

Найти

Найти гиперплоскость, которая максимизирует маржу между двумя классами:
$f : \mathbb{R}^m \to \{-1, 1\}, \; f(x) = \text{sign}(w^T x + b).$

Критерий

Шарнирная функция потерь:
$L(w, X, y) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \max(0, 1 - y_i(X_i^T w + b)),$
которая минимизируется относительно w и b.

Эта задача является сильно выпуклой из-за квадратичной евклидовой нормы.

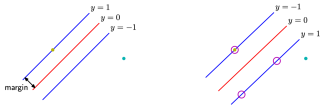


Рис. 6: Метод опорных векторов

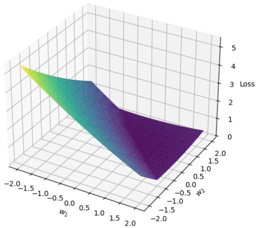


Рис. 7: L_2-регуляризованная шарнирная потеря в пространстве параметров для $x = (1, 1), y = 1$

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$
\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.
$$

Некоторые любопытные примеры

* **Приближение низкоранговой матрицы**

$$
\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.
$$

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.$$

Question

Является ли это выпуклым?

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.$$

> **Question**
>
> Является ли это выпуклым?

По теореме Эккарта-Янга это можно решить с помощью SVD: $X^* = U_k \Sigma_k V_k^T$, где $A = U\Sigma V^T$.

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.$$

> **Question**
>
> Является ли это выпуклым?

По теореме Эккарта-Янга это можно решить с помощью SVD: $X^* = U_k \Sigma_k V_k^T$, где $A = U \Sigma V^T$.

- **Выпуклая релаксация через ядерную норму**

$$\min_X rank(X), \text{ s.t. } X_{ij} = M_{ij}, \ (i,j) \in I.$$

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.$$

Question

Является ли это выпуклым?

По теореме Эккарта-Янга это можно решить с помощью SVD: $X^* = U_k \Sigma_k V_k^T$, где $A = U \Sigma V^T$.

- **Выпуклая релаксация через ядерную норму**

$$\min_X rank(X), \text{ s.t. } X_{ij} = M_{ij}, \ (i,j) \in I.$$

Некоторые любопытные примеры

- **Приближение низкоранговой матрицы**

$$\min_X \|A - X\|_F^2 \text{ s.t. } rank(X) \leq k.$$

Question

Является ли это выпуклым?

По теореме Эккарта-Янга это можно решить с помощью SVD: $X^* = U_k \Sigma_k V_k^T$, где $A = U \Sigma V^T$.

- **Выпуклая релаксация через ядерную норму**

$$\min_X rank(X), \text{ s.t. } X_{ij} = M_{ij}, \ (i,j) \in I.$$

NP-сложная задача, но $\|A\|_* = trace(\sqrt{A^T A}) = \sum_{i=1}^{rank(A)} \sigma_i(A)$ является выпуклой оболочкой ранга матрицы.