

Методы оптимизации, ФКН ВШЭ, зима 2017

Домашнее задание 4: Условная оптимизация.

Срок сдачи: 22 марта (среда) 2017, 23:59, после срока не принимается.

Царькова Анастасия

либо сдать 21 марта на семинаре в письменном виде.

1 Для каждой из следующих задач найдите оптимальное значение и множество оптимальных решений:

(a) (Линейное программирование с одним ограничением)

$$\min_{x \in \mathbb{R}^n} \{c^T x : a^T x \leq \beta\},$$

где $a, c \in \mathbb{R}^n$, $\beta \in \mathbb{R}$, $a \neq 0$, $c \neq 0$.

Решение:

$$L(x, \lambda) = c^T x + \lambda(a^T x - \beta)$$

$$L(x, \lambda) = c^T x + \lambda(a^T x - \beta)$$

$$\begin{cases} \nabla L = \nabla(c^T x + \lambda(a^T x - \beta)) \\ \lambda(a^T x - \beta) = 0 \end{cases}$$

Так как $a, c \neq 0$, следовательно $\lambda \neq 0$

$$\begin{cases} -c^T = \lambda a^T \\ a^T x = \beta \end{cases}$$

Получили что решение x – гиперплоскость проходящая через точки $(a_1/\beta, 0, \dots, 0), (0, \dots, 0, a_n/\beta)$, при условии что вектора a и c пропорциональны

(b) (Линейная функция на стандартном симплексе)

$$\min_{x \in \mathbb{R}^n} \left\{ c^T x : x \succeq 0, \sum_{i=1}^n x_i = 1 \right\},$$

где $c \in \mathbb{R}^n$.

Решение

Так как $c_i < \min(c)$ тогда, поскольку $\sum_{i=1}^n x_i = 1$, $c^T x \geq (\min(c), \dots, \min(c))x$ – ответ – x такие что $x = (0, \dots, 0, 1, 0, \dots, 0)$ где 1 на том месте где у c стоит минимальное значение.

(c) (Линейная функция с энтропийным регуляризатором)

$$\min_{x \in \mathbb{R}_{++}^n} \left\{ c^T x + \sum_{i=1}^n x_i \ln x_i : \sum_{i=1}^n x_i = 1 \right\},$$

где $c \in \mathbb{R}^n$.

Решение:

$$e = (1, \dots, 1)$$

$$L = c^T x + \sum_{i=1}^n x_i \ln x_i + \lambda(e^T x - 1)$$

$$\nabla_i = c_i + \ln x_i + 1 + \lambda = 0$$

$$\nabla_\lambda = e^T x - 1 = 0$$

$$\ln x_i = -c_i - 1 - \lambda$$

$$x_i = \exp(-c_i - 1 - \lambda)$$

$$\begin{cases} x_i = \exp(-c_i - 1 - \lambda) \\ e^T x - 1 = 0 \end{cases}$$

Теперь e это экспонента

$$\begin{cases} x_i = e^{-c_i} e^{-1-\lambda} \\ \sum x_i = 1 \end{cases}$$

$$\begin{cases} x_i = e^{-c_i} e^{-1-\lambda} \\ e^{-1-\lambda} \sum e^{-c_i} = 1 \end{cases}$$

$$\begin{cases} x_i = e^{-c_i} e^{-1-\lambda} \\ e^{-1-\lambda} \sum e^{-c_i} = 1 \end{cases}$$

$$x_i = \frac{e^{-c_i}}{\sum e^{-c_i}}$$

Множество значений

$$f(x^*) = \ln \sum e^{-c_i}$$

2 Для каждой из следующих задач оптимизации: 1) Построить двойственную задачу. 2) Выписать явные формулы, позволяющие по решению двойственной задачи восстановить (вычислить) решение прямой.

(a) (Гребневая регрессия)

$$\min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \left\{ \frac{1}{2} \|s - b\|_2^2 + \frac{\rho}{2} \|x\|_2^2 : s = Ax \right\},$$

где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\rho \in \mathbb{R}_{++}$.

(b) (SVM)

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}^m} \left\{ \sum_{i=1}^m t_i + \frac{\rho}{2} \|x\|_2^2 : Ax \succeq 1_m - t, t \succeq 0 \right\},$$

где $A \in \mathbb{R}^{m \times n}$, $1_m := (1, \dots, 1) \in \mathbb{R}^m$.

3 Свести эквивалентным образом следующие негладкие безусловные задачи к гладким условным:

(a) (Максимум из конечного числа гладких функций)

$$\min_{x \in \mathbb{R}^n} \max\{f_1(x), \dots, f_m(x)\},$$

где $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ — заданные гладкие функции.

Решение:

Рассмотрим любое t и любое x , для которых $t > f_i(x)$. Для них справедливо, что

$$t > \max\{f_1(x), \dots, f_m(x)\} \geq \min_{x \in \mathbb{R}^n} \max\{f_1(x), \dots, f_m(x)\}$$

Поскольку это неравенство верно для всех t , то оно верно и для минимума по x, t , то есть

$$\min_{x, t, t > f_i(x)} t \geq \min_{x \in \mathbb{R}^n} \max\{f_1(x), \dots, f_m(x)\}$$

С другой стороны, для любого $x, \varepsilon > 0$ рассмотрим

$$t = \max f_i(x) + \varepsilon$$

Это выражение больше, чем все $f_i(x)$. Поэтому, оно больше, чем $\min_{x, t, t > f_i(x)} t$. Но поскольку это происходит для любого сколь угодно малого ε , то

$$\max f_i(x) \geq \min_{x, t, t > f_i(x)} t$$

Поскольку это происходит для всех x , то

$$\min_{x \in \mathbb{R}^n} \max\{f_1(x), \dots, f_m(x)\} \geq \min_{x, t, t > f_i(x)} t$$

В итоге, получаем, что требуемые выражения равны.

(b) (Наилучшее решение линейной системы в ℓ_∞ -норме)

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_\infty,$$

где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Для произвольного вектора $y \in \mathbb{R}^m$: $\|y\|_\infty := \max_{i=1}^m |y_i|$.

(c) (Наилучшее решение линейной системы в ℓ_1 -норме)

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1,$$

где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Для произвольного вектора $y \in \mathbb{R}^m$: $\|y\|_1 := \sum_{i=1}^m |y_i|$.

(d) (Задача LASSO)

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \rho \|x\|_1 \right\},$$

где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\rho \in \mathbb{R}_{++}$.

4 Для каждой из следующих квадратичных задач (QCQP) найдите аналитическое решение:

(a) (Минимизация линейной формы на эллипсоиде)

$$\min_{x \in \mathbb{R}^n} \{c^T x : x^T A x \leq 1\},$$

где $c \in \mathbb{R}^n \setminus \{0\}$ и $A \in \mathbb{S}_{++}^n$.

Решение:

Существует матрица D такая, что

$$A = D^T D$$

Сделаем замену

$$y = Dx$$

Матрица D – обратима, поскольку имеет ненулевой определитель, поскольку матрица A строго-положительно определена. Поэтому

$$x = D^{-1}y$$

В итоге, получается, что нам нужно найти

$$\min_{x \in \mathbb{R}^n} \{c^T D^{-1}y : y^T y \leq 1\}$$

Иными словами, найти минимум выражения

$$k_1 y_1 + \dots + k_n y_n$$

при условии

$$y_1^2 + \dots + y_n^2 \leq 1$$

Выписав производные Лагранжиана, получаем условия

$$k_1 = 2\lambda y_1$$

то есть оптимальные значения y_i пропорциональны k_i . Из условия на сумму квадратов, получаем такой ответ

$$y_i = \frac{-k_i}{\sqrt{k_1^2 + \dots + k_n^2}}$$

где k_i – коэффициенты вектора $c^T D^{-1}$.

(b) (Минимизация квадратичной формы на эллипсоиде)

$$\min_{x \in \mathbb{R}^n} \{x^T B x : x^T A x \leq 1\},$$

где $A \in \mathbb{S}_{++}^n$, $B \in \mathbb{S}_+^n$.

Очевидно, что

$$x^T B x \geq 0$$

Кроме того, ноль достигается при $x = 0$, поэтому $x = 0$ и есть ответ.

5 Для каждого из следующих множеств $Q \subseteq \mathbb{R}^n$ найти евклидову проекцию заданной точки $v \in \mathbb{R}^n$ на множество Q (т. е. найти $\Pi_Q(v) := \operatorname{argmin}_{x \in Q} \|x - v\|_2^2$):

- (a) (Короб) $Q = \{x \in \mathbb{R}^n : x_i \in [l_i, r_i], i = 1, \dots, n\}$, где $-\infty \leq l_i \leq r_i \leq +\infty$. (Замечание: Допускается, что $l_i = -\infty$ и/или $r_i = +\infty$, т. е. короб может быть неограниченным вдоль некоторых направлений.)

Рассмотрим произвольную точку v . Надо найти для нее самую ближайшую в коробе, для этого надо минимизировать сумму

$$|x_1 - v_1|^2 + \dots + |x_n - v_n|^2$$

Если

$$v_i \in [l_i, r_i]$$

то возьмем

$$x_i = v_i$$

Если же

$$v_i > r_i$$

то возьмем

$$x_i = r_i$$

И, наконец, если

$$v_i < l_i$$

возьмем

$$x_i = l_i$$

Понятно, что таким образом мы минимизируем каждое слагаемое $|x_i - v_i|$ по-отдельности. И, плюс к этому, построенная точка действительно лежит в коробе, поэтому она и минимизирует всю сумму квадратов модулей

- (b) (Аффинное многообразие) $Q = \{x \in \mathbb{R}^n : Ax = b\}$, где $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\text{Rank}(A) = m$.
 (c) (Полупространство) $Q = \{x \in \mathbb{R}^n : a^T x \leq \beta\}$, где $a \in \mathbb{R}^n$, $\beta \in \mathbb{R}$, $a \neq 0$.

Если v лежит в полуплоскости, то $x = v$. Если нет, то надо выбрать ортогональную проекцию, то есть

$$x = v - \lambda a$$

Найдем λ :

$$\begin{aligned} a^T v - a^T a \lambda &= b \\ \lambda &= \frac{a^T v - b}{a^T a} \end{aligned}$$

В итоге, ответ

$$x = v - \frac{a^T v - b}{a^T a} a$$

Воспользуйтесь полученными выше результатами и выпишите ответ для следующих случаев:

- (Неотрицательный ортант) $Q = \mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_i \geq 0, i = 1, \dots, n\}$. Нужно взять такой вектор x , что

$$x_i = \min(0, v_i)$$

- (Единичный L_∞ -шар) $Q = B_\infty(0, 1) = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$. Нужно взять

$$x_i = \operatorname{sgn}(v_i) \cdot \max(1, |v_i|)$$

- (Гиперплоскость) $Q = \{x \in \mathbb{R}^n : a^T x = \beta\}$, где $a \in \mathbb{R}^n, \beta \in \mathbb{R}, a \neq 0$.

$$x = v - \frac{a^T v - \beta}{a^T a} a$$

Бонусная часть (6 баллов)

6 Рассмотрим QCQP:

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} x^T A x - b^T x : \|x\|_2 \leq 1 \right\},$$

где $A \in \mathbb{S}_{++}^n$ и $b \in \mathbb{R}^n$. Докажите, что оптимальное решение в этой задаче равно $(A + \lambda I_n)^{-1} b$, где $\lambda := \max\{0, \bar{\lambda}\}$ и $\bar{\lambda}$ — это наибольшее из решений нелинейного уравнения

$$b^T (A + \lambda I_n)^{-2} b = 1.$$

Решение:

x — оптимально если

1. $x^T x < 1$ — — $Ax + b = 0$ тогда $Ax = -b$ и если $\|A^{-1}b\|_2 \leq 1$
2. $x^T x = 1$ — — $Ax + b = -\lambda x \forall \lambda \geq 0$. Из условий $\|x\|_2 = 1$, и пусть λ_i — собственные числа матрицы A — тогда определим $f(\lambda) = \|(A + \lambda I)^{-1} b\|_2^2 = \sum_{i=1}^n \frac{b_i^2}{(\lambda + \lambda_i)^2}$ тогда для этой функции $f(0) = \|A^{-1}b\|_2^2 > 1$. Полученная функция монотонно убывает при стремлении λ к нулю — поэтому уравнение $Ax + b = -\lambda x$ эквивалентное $f(\lambda) = 0$ имеет только одно единственное решение а значит $x = -(A + \lambda I)^{-1} b$ — оптимальное решение

7 Рассмотрим задачу поиска евклидовой проекции заданной точки $v \in \mathbb{R}^n$ на стандартный симплекс:

$$\Pi_{\Delta_n}(v) := \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ \|x - v\|_2 : x \succeq 0, \sum_{i=1}^n x_i = 1 \right\}.$$

Докажите, что $\Pi_{\Delta_n}(v) = (v - \nu 1_n)_+$, где $\nu \in \mathbb{R}$ — корень нелинейного уравнения

$$1_n^T (v - \nu 1_n)_+ = 1. \quad (1)$$

Здесь $1_n := (1, \dots, 1) \in \mathbb{R}^n$, а $(u)_+$ обозначает поэлементную положительную срезку $(u_i)_+ := \max\{0, u_i\}$. Нарисуйте схематичный график левой части уравнения (1) как функции от ν .

Подсказка. Удобно рассмотреть упорядоченные компоненты $v_{[1]} \geq \dots \geq v_{[n]}$.

8 Пусть $\Sigma, \Sigma_0 \in \mathbb{S}_{++}^n$. Обозначим через $D(\Sigma; \Sigma_0)$ дивергенцию Кульбака–Лейблера между нормальными распределениями $\mathcal{N}(0, \Sigma)$ и $\mathcal{N}(0, \Sigma_0)$:

$$D(\Sigma; \Sigma_0) = \frac{1}{2} (\operatorname{Tr}(\Sigma_0^{-1} \Sigma) - \ln \operatorname{Det}(\Sigma_0^{-1} \Sigma) - n).$$

Пусть $H \in \mathbb{S}_{++}^n$. Пусть также $y, s \in \mathbb{R}^n$, причем $y^T s > 0$. Рассмотрим задачу поиска матрицы $X \in \mathbb{S}_{++}^n$, минимизирующую дивергенцию $D(X^{-1}; H^{-1})$ при условии $Xy = s$:

$$\min_{X \in \mathbb{S}_{++}^n} \{D(X^{-1}; H^{-1}) : Xy = s\}.$$

Решите эту задачу и убедитесь, что ее решение выражается по формуле обновления обратной матрицы в схеме BFGS:

$$X = (I_n - \rho s y^T) H (I_n - \rho y s^T) + \rho s s^T,$$

где $\rho := 1/(y^T s)$.