Методы оптимизации в машинном обучении, ВМК+Физтех, осень 2017

Домашнее задание 4: Субдифференциалы.

Срок сдачи: 22 ноября 2017 (среда), 23:59 для ВМК
24 ноября 2017 (пятница), 23:59 для Физтеха

Обязательная часть

1 Пусть $f : [0,1] \to \mathbb{R}$ — функция $f(t) := -\sqrt{t}$. Найдите субдифференциал ∂f.

2 Пусть $\delta_{\mathbb{R}^n_+} : \mathbb{R}^n_+ \to \mathbb{R}$ — индикаторная функция конуса $\mathbb{R}^n_+$ (таким образом, $\delta_{\mathbb{R}^n_+}(x) = 0$ для $x \in \mathbb{R}^n_+$). Найдите субдифференциал $\partial \delta_{\mathbb{R}^n_+}$.

3 Пусть E — непустое множество в евклидовом пространстве V, пусть $f : E \to \mathbb{R}$ — функция, и пусть $x_0 \in E$. Покажите, что

$$\partial f(x_0) = \{g \in D : \langle g, x_0 \rangle = f^*(g) + f(x_0)\},$$

где $D := \{g \in V : \sup_{x \in E}\{\langle g, x \rangle - f(x)\} < +\infty\}$ и $f^* : D \to \mathbb{R}$ — *сопряженная функция Фенхеля* $f^*(g) := \sup_{x \in E}\{\langle g, x \rangle - f(x)\}$. Другими словами, вектор $g \in D$ является субградиентом функции f в точке x_0, если и только если *неравенство Фенхеля–Юнга*

$$\langle g, x_0 \rangle \leq f^*(g) + f(x_0)$$

переходит в равенство.

4 Пусть $f : \mathbb{S}^n \to \mathbb{R}$ — функция $f(X) := \lambda_{\max}(X)$.

(a) Найдите субдифференциал ∂f.

(b) Покажите, что

$$\partial f(0) = \{G \in \mathbb{S}^n_+ : \mathrm{Tr}(G) = 1\}.$$

(c) Покажите, что f дифференцируема в точке $X_0 \in \mathbb{S}^n$ тогда и только тогда, когда максимальное собственное значение матрицы X_0 является простым (т. е. имеет кратность 1).

5 Для каждой из следующих функций $f : \mathbb{R}^n \to \mathbb{R}$ найдите субдифференциал ∂f:

(a) $f(x) := \|x\|_\infty$.

(b) $f(x) := \sum_{1 \leq i < j \leq n}^n c_{ij}|x_i - x_j|$, где $c_{ij} \geq 0$ для $1 \leq i < j \leq n$.

(c) $f(x) := \sum_{i=1}^m \|x - a_i\|_2$, где $a_1, \ldots, a_m \in \mathbb{R}^n$.

6 Для каждой из следующих задач минимизации выпишите решение в явном виде:

(a) (Проксимальное отображение для l^2-нормы)

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2}\|x - v\|_2^2 + \tau\|x\|_2 \right\},$$

где $v \in \mathbb{R}^n$, $\tau > 0$.

(b) (Проксимальное отображение для l^1-нормы)

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

где $v \in \mathbb{R}^n$, $\tau > 0$.

Бонусная часть

1 Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, f^* — минимальное значение функции f, и пусть $(x_k)_{k=0}^{\infty}$ — последовательность точек в $\mathbb{R}^n$, построенная *субградиентным методом Поляка*

$$x_{k+1} := x_k - \frac{f(x_k) - f^*}{\|g_k\|^2} g_k,$$

где $g_k \in \partial f(x_k)$ — произвольный субградиент функции f в точке x_k для $k \geq 0$. Пусть $x^* \in \mathbb{R}^n$ — точка минимума функции f, и пусть $R := \|x_0 - x^*\|$. Пусть также $T \geq 1$, и пусть $\|g_k\| \leq M$ для всех $0 \leq k \leq T - 1$. Покажите, что

$$\min_{0 \leq k \leq T-1} f(x_k) - f^* \leq \frac{MR}{\sqrt{T}},$$

(*Подсказка*: адаптируйте доказательство скорости сходимости классического субградиентного метода.)

2 Пусть $\|\|_{\mathrm{op}} : \mathbb{R}^{m \times n} \to \mathbb{R}$ — операторная норма $\|X\|_{\mathrm{op}} := \sigma_{\max}(X)$. Напомним, что сопряженная норма $\|\|_{\mathrm{op}}^* : \mathbb{R}^{m \times n} \to \mathbb{R}$ задается формулой $\|A\|_{\mathrm{op}}^* := \max_{\|X\|_{\mathrm{op}}=1} |\langle A, X \rangle|$. Пусть $A \in \mathbb{R}^{m \times n}$. Покажите, что

$$\|A\|_{\mathrm{op}}^* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A).$$

Таким образом, сопряженная норма к операторной норме равна сумме сингулярных чисел (эта норма называется *ядерной нормой*). (*Подсказка*: воспользуйтесь сингулярным разложением $A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T$ и оцените $|\langle A, X \rangle|$ сверху; затем покажите, что неравенство достигается для некоторого X.)

3* Пусть $A \in \mathbb{R}^{m \times n}$ — матрица ранга r, и пусть $A = U \Sigma V^T$ — сингулярное разложение матрицы A, где $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$ — матрицы с ортонормированными столбцами, $\Sigma \in \mathbb{R}^{r \times r}$ — диагональная матрица. Покажите, что

$$\partial \|\|_{\mathrm{op}}^*(A) = \{UV^T + W : W \in \mathbb{R}^{m \times n};\ U^T W = 0;\ WV = 0;\ \|W\|_{\mathrm{op}} \leq 1\}.$$

(*Подсказка*: воспользуйтесь результатом задачи **3** (из обязательной части).)