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

Семинар 8: Стандартные классы выпуклых задач и двойственность

1 Стандартные классы выпуклых задач

1.1 Линейное программирование (LP)

Стандартная форма LP:

$$\min_x \langle c, x\rangle \qquad \text{s.t.} \quad Ax \preceq b,$$

где $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$.

Пример 1.1 (Максимум из аффинных функций)**.** Пусть $a_1, \dots, a_m \in \mathbb{R}^n$ и $b_1, \dots, b_m \in \mathbb{R}$. Рассмотрим задачу

$$\min_{x\in\mathbb{R}^n} \max_{1\le i\le m} \{\langle a_i, x\rangle - b_i\}.$$

Эта задача является задачей *негладкой* безусловной минимизации (из-за присутствия максимума). Тем не менее, эта задача эквивалентна следующей задаче линейного программирования:

$$\min_{x,t} t \qquad \text{s.t.} \quad t \ge \langle a_i, x\rangle - b_i, \quad 1 \le i \le m.$$

В стандартной форме это соответствует тому, что $\tilde{n} = n+1$, $\tilde{m} = m$, $\tilde{x} = (x, t) \in \mathbb{R}^{n+1}$, $\tilde{c} := (0_n, 1) \in \mathbb{R}^{n+1}$, $\tilde{b} := b$ и $\tilde{A} := (A, 1_n) \in \mathbb{R}^{m\times(n+1)}$, где $A \in \mathbb{R}^{m\times n}$ — матрица со строками $a_1, \dots, a_m$.

1.2 Квадратичное программирование (QP)

Стандартная форма QP:

$$\min_x \left\{\frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle\right\} \qquad \text{s.t.} \quad Gx \preceq h,$$

где $x, b \in \mathbb{R}^n$, $A \in \mathbb{S}^n_+$, $G \in \mathbb{R}^{m\times n}$, $h \in \mathbb{R}^m$.

Пример: линейная регрессия $\|Ax - b\|_2^2$ на положительном ортанте: $x \succeq 0$.
Пример: расстояние между полиэдрами: $\min \|x - y\|_2^2$, $Ax \preceq b$, $Cy \preceq d$.

1.3 Квадратичное программирование с квадратичными ограничениями (QCQP)

Стандартная форма QCQP:

$$\min_x \left\{\frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle\right\} \qquad \text{s.t.} \quad \frac{1}{2}\langle P_i x, x\rangle + \langle q_i, x\rangle + c_i \le 0, \quad 1 \le i \le m,$$

где $x, b, q_1, \dots, q_m \in \mathbb{R}^n$, $A, P_1, \dots, P_m \in \mathbb{S}^n_+$, $c_1, \dots, c_m \in \mathbb{R}$.

Пример 1.2 (Квадратичная функция на единичном шаре)**.** Рассмотрим задачу

$$\min_x \left\{\frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle\right\} \qquad \text{s.t.} \quad \|x\|_2 \le 1,$$

где $x, b \in \mathbb{R}^n$ и $A \in \mathbb{S}^n_+$. Эта задача эквивалентна следующей задаче QCQP:

$$\min_x \left\{ \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle \right\} \qquad \text{s. t.} \quad \langle x, x\rangle \le 1.$$

В этом случае $m = 1$, $P_1 = I_n$, $q_1 = 0$ и $c_1 = -1$.

1.4 Коническое программирование второго порядка (SOCP)

Стандартная форма SOCP:

$$\min_x \langle q, x\rangle \qquad \text{s. t.} \quad \|A_i x + b_i\|_2 \le \langle c_i, x\rangle + d_i, \quad 1 \le i \le m,$$

где $x, q, c_1, \ldots, c_m \in \mathbb{R}^n$, $A_1 \in \mathbb{R}^{s_1 \times n}$, $b_1 \in \mathbb{R}^{s_1}$, $\ldots$, $A_m \in \mathbb{R}^{s_m \times n}$, $b_m \in \mathbb{R}^{s_m}$, $d_1, \ldots, d_m \in \mathbb{R}$.

Пример 1.3 (Задача Lasso)**.** Пусть $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Рассмотрим задачу

$$\min_x \{\|Ax - b\|_2^2 + \|x\|_2\}. \tag{1.1}$$

Эта задача является негладкой из-за присутствия слагаемого $\|x\|_2$. Эта задача является эквивалентной следующей задаче SOCP:

$$\min_{x, t_1, t_2} \{t_1 + t_2\} \qquad \text{s. t.} \quad \|Ax - b\|_2 \le t_1, \qquad \|x\|_2 \le t_2.$$

(Почему эту задачу (1.1) нельзя было похожим образом переформулировать как QCQP?)

1.5 Полуопределённое программирование (SDP)

Стандартная форма SDP:

$$\min_x \langle c, x\rangle \qquad \text{s. t.} \quad A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0, \quad Cx = d,$$

где $x, c \in \mathbb{R}^n$, $A_0, \ldots, A_n \in \mathbb{S}^n$, $C \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$.

Пример 1.4 (Минимизация спектральной нормы)**.** Рассмотрим задачу

$$\min_x \|A(x) := A_0 + x_1 A_1 + \cdots + x_n A_n\|_2,$$

где $x \in \mathbb{R}^n$, $A_0, \ldots, A_n \in \mathbb{S}^n$. Представим эту задачу в виде SDP. Для этого сначала перейдем эквивалентной формулировке через надграфик:

$$\min_{x,t} t \qquad \text{s. t.} \quad t \ge \|A(x)\|_2.$$

Заметим, что

$$\|A(x)\|_2 \le t \qquad \Leftrightarrow \qquad \lambda_{\max}(A(x)A(x)^T) \le t^2 \qquad \Leftrightarrow \qquad A(x)A(x)^T \preceq t^2 I_n.$$

По *лемме о дополнении Шура*, получаем

$$A(x)A(x)^T \leq t^2 I_n \qquad \Leftrightarrow \qquad \begin{pmatrix} tI_n & A(x)^T \\ A(x) & tI_n \end{pmatrix} \succeq 0.$$

Таким образом, исходная задача эквивалентна следующей SDP:

$$\min_{x,t} t \qquad \text{s. t.} \quad \begin{pmatrix} tI_n & A(x)^T \\ A(x) & tI_n \end{pmatrix} \succeq 0.$$

1.6 Обсуждение

Можно показать, что LP $\subset$ QP $\subset$ QCQP $\subset$ SOCP $\subset$ SDP. Таким образом, SDP являются наиболее общей постановкой выпуклых задач.

Замечание 1.1. Полезно заметить, что целевую функцию всегда можно перенести в ограничения и получить эквивалентную задачу с линейной целевой функцией.

2 Двойственность

Ещё одно важное понятие, возникающее в условной оптимизации — двойственная задача.

Мы предложим универсальный способ: как для любой условной задачи оптимизации $\min_{x \in Q} f(x)$ записать новую *двойственную* задачу $\max_{y \in \Omega} g(y)$, при этом, двойственная задача будет оценивать прямую задачу снизу:

$$g(y) \;\leqslant\; f(x) \qquad \text{для всех} \quad y \in \Omega,\; x \in Q.$$

Поскольку неравенство справедливо для всех допустимых x и y, значит:

$$\max_{y \in \Omega} g(y) \;\leqslant\; \min_{x \in Q} f(x).$$

Прежде чем перейти к конкретному (одному из возможных) способов построения двойственной задачи, обсудим, зачем это может быть полезно.

(a) *Построение оценки снизу на решение прямой задачи.* Решить исходную задачу может быть очень сложно. Но если у нас есть двойственная к ней, то мы можем взять произвольный $y \in \Omega$ и подставить его в $g(y)$ — получим некоторую оценку снизу:

$$g(y) \;\leqslant\; \min_{x \in Q} f(x).$$

(b) *Проверка на допустимость задачи и ограниченность решения.* Из неравенства $\max_{y \in \Omega} g(y) \;\leqslant\; \min_{x \in Q} f(x)$ следует: если $\min_{x \in Q} f(x) = -\infty$, значит $\Omega = \emptyset$ и наоборот.

(c) *Двойственную задачу бывает решить легче, чем прямую.* При этом, если выполнена *сильная двойственность*: $g(y^*) = f(x^*)$ то мы ничего не теряем.

(d) *Получение оценки сверху на невязку по функции:* $f(x) - f^* \;\leqslant\; f(x) - g(y)$ для произвольного $y \in \Omega$.

Опишем возможный способ построения двойственной задачи для случая, когда Q задаётся функциональными ограничениями.

Заметим, что в этом случае, исходная условная задача $\min_{x\in Q} f(x)$ эквивалента следующей безусловной задаче:

$$\min_{x\in\mathbb{R}^n} \max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k} \mathcal{L}(x;\lambda,\mu).$$

Поменяем минимум и максимум местами — получим оценку снизу (это универсальное правило, которое всегда можно использовать):

$$\max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k} \min_{x\in\mathbb{R}^n} \mathcal{L}(x;\lambda,\mu) \;\leqslant\; \min_{x\in R^n} \max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k} \mathcal{L}(x;\lambda,\mu).$$

Обозначим внутреннюю задачу оптимизации в оценке снизу за $g(\lambda,\mu)$:

$$g(\lambda,\mu) \;:=\; \min_{x\in\mathbb{R}^n} \mathcal{L}(x;\lambda,\mu).$$

Получили *двойственную задачу*:

$$\max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k} g(\lambda,\mu).$$

Замечание 2.1. Заметим, что двойственная функция $g(\lambda,\mu)$ всегда будет вогнутой, как максимум аффинных функций.

По построению всегда выполнена *слабая двойственность*: $g^* \leqslant f^*$.

В случае, если для исходной выпуклой задачи выполнено *условие Слейтера*, справедлива *сильная двойственность*: $g^* = f^*$.

Замечание 2.2. Если выполнена сильная двойственность ($g^* = f^*$), то решения прямой задачи x^* и двойственной (λ^*,μ^*) тоже связаны:

$$x^* \in \operatorname*{Argmin}_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda^*,\mu^*),$$

— решения прямой задачи принадлежат множеству минимумов лагранжиана по прямым переменным, при фиксированных оптимальных двойственных.