Машинное обучение, ФКН ВШЭ
Семинар №13

1 Условная оптимизация

§1.1 Двойственная задача

Рассмотрим задачу минимизации

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d}, \\ f_i(x) \leqslant 0, \quad i = 1, \dots, m, \\ h_i(x) = 0, \quad i = 1, \dots, p. \end{cases} \tag{1.1}$$

На лекции было введены понятия лагранжиана $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + + \sum_{i=1}^p \nu_i h_i(x)$ и двойственной функции $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$ и показано, что значение двойственной функции является нижней оценкой на минимум в задаче 1.1:

$$g(\lambda, \nu) \le f_0(x_*) \quad \forall (\lambda, \nu).$$

Также было введено понятие двойственной задачи:

$$\begin{cases} g(\lambda, \nu) \to \max\limits_{\lambda, \nu} \\ \lambda_i \geqslant 0, \quad i = 1, \dots, m. \end{cases} \tag{1.2}$$

Задача 1.1. *Постройте двойственную к оптимизационной задаче:*

$$\begin{cases} \|x\|^2 \to \min\limits_x \\ Ax = b. \end{cases}$$

Отметим, что это задача поиска решения системы линейных уравнений с наименьшей нормой.

Решение. Запишем лагранжиан:

$$L(x, \nu) = \|x\|^2 + \nu^T (Ax - b).$$

Найдем градиент:

$$\nabla_x L(x, \nu) = 2x + \nu^T A = 2x + A^T \nu.$$

Приравняв градиент нулю, найдем минимум лагранжиана при данном ν:

$$x = -\frac{1}{2}A^T\nu.$$

Значит, двойственная функция равна

$$g(\nu) = L\left(-\frac{1}{2}A^T\nu, \nu\right) = -\frac{1}{4}\nu^T AA^T\nu - b^T\nu.$$

Поскольку ограничений-неравенств в исходной задаче нет, в двойственной задаче не будет ограничений. Получаем двойственную задачу

$$-\frac{1}{4}\nu^T AA^T\nu - b^T\nu \to \max_{\nu}.$$

■

Задача 1.2. *Постройте двойственную к задаче линейного программирования в стандартном виде:*

$$\begin{cases} \langle c, x\rangle \to \min\limits_x \\ Ax = b, \\ x \geqslant 0. \end{cases}$$

Решение. Запишем лагранжиан:

$$L(x, \lambda, \nu) = \langle c, x\rangle - \lambda^T x + \nu^T(Ax - b).$$

Отметим, что ограничения-неравенства вошли с минусом, мы привели их к стандартному виду $-x \leqslant 0$. Немного преобразуем лагранжиан:

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T Ax - b^T\nu = -b^T\nu + (c + A^T\nu - \lambda)^T x.$$

Двойственная функция имеет вид

$$g(\lambda, \nu) = -b^T\nu + \inf_x (c + A^T\nu - \lambda)^T x.$$

Заметим, что выражение $(c + A^T\nu - \lambda)^T x$ линейно по x и не ограничено снизу, если $c + A^T\nu - \lambda \neq 0$, а потому точная нижняя грань не существует. Поскольку в двойственной задаче ищется максимум этой точной нижней грани, достаточно найти его для случая $c + A^T\nu - \lambda = 0$, поэтому оно будет являться ограничением в двойственная задаче.

Получаем, что двойственная задача имеет вид

$$\begin{cases} -b^T\nu \to \max\limits_\nu \\ c + A^T\nu - \lambda = 0, \\ \lambda \geqslant 0. \end{cases}$$

Ограничения можно объединить, избавившись от λ:

$$\begin{cases} -b^T\nu \to \max\limits_\nu \\ c + A^T\nu \geqslant 0. \end{cases}$$

■

§1.2 Сильная и слабая двойственность

Пусть (λ^*, ν^*) — решение двойственной задачи. Значение двойственной функции всегда не превосходит условный минимум исходной задачи:

$$g(\lambda^*, \nu^*) \leqslant f_0(x_*).$$

Это свойство называется *слабой двойственностью*. Разность $f_0(x_*) - g(\lambda^*, \nu^*)$ называется *зазором* между решениями прямой и двойственной задач.

Если имеет место равенство

$$g(\lambda^*, \nu^*) = f_0(x_*),$$

то говорят о *сильной двойственности*. Существует много достаточных условий сильной двойственности. Одним из таких условий для выпуклых задач является условие Слейтера. *Выпуклой* задачей оптимизации называется задача

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d} \\ f_i(x) \leqslant 0, \quad i = 1, \dots, m, \\ Ax = b. \end{cases}$$

где функции $f_0, f_1, \dots, f_m$ являются выпуклыми. Условие Слейтера требует, чтобы существовала такая допустимая точка x', в которой ограничения-неравенства выполнены строго:

$$\begin{cases} f_i(x) < 0, \quad i = 1, \dots, m, \\ Ax = b. \end{cases}$$

Условие Слейтера можно ослабить: достаточно, чтобы ограничения-неравенства были строгими только в том случае, если они не являются линейными (т.е. не имеют вид $Ax - b$).

§1.3 Условия Куна-Таккера

Пусть x_* и (λ^*, ν^*) — решения прямой и двойственной задач. Будем считать, что имеет место сильная двойственность. Тогда:

$$\begin{aligned} f_0(x_*) &= g(\lambda^*, \nu^*) \\ &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right) \\ &\leqslant f_0(x_*) + \sum_{i=1}^m \lambda_i^* f_i(x_*) + \sum_{i=1}^p \nu_i^* h_i(x_*) \\ &\leqslant f_0(x_*) \end{aligned}$$

Получаем, что все неравенства в этой цепочке выполнены как равенства. Отсюда можно сделать несколько выводов.

Во-первых, если подставить в лагранжиан решение двойственной задачи (λ^*, ν^*), то его минимум будет достигаться на решении прямой задачи x_*. Иными словами, решение исходной задачи (1.1) эквивалентно минимизации лагранжиана $L(x, \lambda^*, \nu^*)$ с подставленным решением двойственной задачи.

Во-вторых, из последнего неравенства получаем, что

$$\sum_{i=1}^{m} \lambda_i^* f_i(x_*) = 0.$$

Каждый член неположителен, поэтому

$$\lambda_i^* f_i(x_*) = 0, \quad i = 1, \ldots, m.$$

Эти условия называются *условиями дополняющей нежесткости*. Они говорят, что множитель Лагранжа при i-м ограничении может быть не равен нулю лишь в том случае, если ограничение выполнено с равенством (в этом случае говорят, что оно является *активным*).

Итак, мы можем записать следующие условия, которые выполнены для решений прямой и двойственной задач x_* и (λ^*, ν^*):

$$\begin{cases} \nabla f_0(x_*) + \sum\limits_{i=1}^{m} \lambda_i^* \nabla f_i(x_*) + \sum\limits_{i=1}^{p} \nu_i^* \nabla h_i(x_*) = 0 \\ f_i(x_*) \leqslant 0, \quad i = 1, \ldots m \\ h_i(x_*) = 0, \quad i = 1, \ldots p \\ \lambda_i^* \geqslant 0, \quad i = 1, \ldots m \\ \lambda_i^* f_i(x_*) = 0, \quad i = 1, \ldots m \end{cases} \tag{ККТ}$$

Данные условия называются *условиями Куна-Таккера* (в зарубежной литературе их принято называть условиями Каруша-Куна-Таккера) и являются необходимыми условиями экстремума. Их можно сформулировать несколько иначе:

Теорема 1.1. *Пусть x_* — решение задачи* (1.1). *Тогда найдутся такие векторы λ^* и ν^*, что выполнены условия* (ККТ).

Если задача (1.1) является выпуклой и удовлетворяет условию Слейтера, то условия Куна-Таккера становятся *необходимыми и достаточными*.

Задача 1.3. *Решите следующую задачу условной оптимизации:*

$$\begin{cases} (x-4)^2 + (y-4)^2 \to \min\limits_{x,y} \\ x + y \leqslant 4, \\ x + 3y \leqslant 9. \end{cases}$$

Решение. Выпишем лагранжиан:

$$L(x, y, \lambda_1, \lambda_2) = (x-4)^2 + (y-4)^2 + \lambda_1(x + y - 4) + \lambda_2(x + 3y - 9).$$

Условия Куна–Таккера запишутся в виде:

$$\begin{cases} 2(x-4)+\lambda_1+\lambda_2=0, \\ 2(y-4)+\lambda_1+3\lambda_2=0, \\ x+y\leqslant 4,\ \lambda_1\geqslant 0,\ \lambda_1(x+y-4)=0, \\ x+3y\leqslant 9,\ \lambda_2\geqslant 0,\ \lambda_2(x+3y-9)=0. \end{cases}$$

Решая их, рассмотрим 4 случая:

- $x+y=4,\ x+3y=9,\ \lambda_1\geq 0,\ \lambda_2\geq 0.$
  Два эти уравнения дают $(x=\frac{3}{2}, y=\frac{5}{2})$. После подстановки в первые два уравнения условий Куна–Таккера, получаем

  $$\begin{cases} 2(\frac{3}{2}-4)+\lambda_1+\lambda_2=0; \\ 2(\frac{5}{2}-4)+\lambda_1+3\lambda_2=0, \end{cases}$$

  откуда $\lambda_2=-1$, что противоречит принятым условиям.

- $x+y=4,\ x+3y\leq 9,\ \lambda_1\geq 0,\ \lambda_2=0.$
  Подстановка $\lambda_2=0$ в первые два уравнения условий Куна–Таккера вместе с уравнением $x+y=4$ дают решение $(x=2, y=2, \lambda_1=4, \lambda_2=0)$. Эти решения удовлетворяют всем условиям Куна–Таккера.

- Два оставшихся случая, как и первый, ведут к противоречиям.

Поскольку задача выпуклая и удовлетворяет ослабленным условиям Слейтера, найденная точка является решением.

■