Машинное обучение, ФКН ВШЭ
Семинар 13

29 января 2021 г.

1 Оптимизационные задачи и теорема Куна-Таккера

Рассмотрим задачу минимизации

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

Если ограничения в этой задаче отсутствуют, то имеет место *необходимое условие экстремума*: если в точке x функция f_0 достигает своего минимума, то ее градиент в этой точке равен нулю. Значит, для решения задачи безусловной оптимизации

$$f_0(x) \to \min$$

достаточно найти все решения уравнения

$$\nabla f_0(x) = 0,$$

и выбрать то, в котором достигается наименьшее значение. Для решения условных задач оптимизации требуется более сложный подход, который мы сейчас и рассмотрим.

§1.1 Лагранжиан

Задача условной оптимизации (1.1) эквивалентна следующей безусловной задаче:

$$f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)) \to \min_x,$$

где $I_-(x)$ — индикаторная функция для неположительных чисел:

$$I_-(x) = \begin{cases} 0, & x \leqslant 0 \\ \infty, & x > 0, \end{cases}$$

а $I_0(x)$ — индикаторная функция для нуля:

$$I_0(x) = \begin{cases} 0, & x = 0 \\ \infty, & x \neq 0, \end{cases}$$

Такая переформулировка, однако, не упрощает задачу — индикаторные функции являются кусочно-постоянными и могут быть оптимизированы лишь путем полного перебора решений.

Заменим теперь индикаторные функции на их линейные аппроксимации:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

где $\lambda_i \geqslant 0$. Полученная функция называется *лагранжианом* задачи (1.1). Числа λ_i и ν_i называются *множителями Лагранжа* или *двойственными переменными.*

Конечно, линейные аппроксимации являются крайне грубыми, однако их оказывается достаточно, чтобы получить необходимые условия на решение исходной задачи.

§1.2 Двойственная функция

Двойственной функцией для задачи (1.1) называется функция, получающаяся при взятии минимума лагранжиана по x:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$$

Можно показать, что данная функция всегда является вогнутой.

Зачем нужна двойственная функция? Оказывается, она дает нижнюю оценку на минимум в исходной оптимизационной задаче. Обозначим решение задачи (1.1) через x_*. Пусть x' — *допустимая* точка, т.е. $f_i(x') \leqslant 0$, $h_i(x') = 0$. Пусть также $\lambda_i > 0$. Тогда

$$L(x', \lambda, \nu) = f_0(x') + \sum_{i=1}^{m} \lambda_i f_i(x') + \sum_{i=1}^{p} \nu_i h_i(x') \leqslant f_0(x').$$

Если взять в левой части минимум по всем допустимым x, то неравенство останется верным; оно останется верным и в случае, если мы возьмем минимум по всем возможным x:

$$\inf_x L(x, \lambda, \nu) \leqslant \inf_{x\text{ — допуст.}} L(x, \lambda, \nu) \leqslant L(x', \lambda, \nu).$$

Итак, получаем

$$\inf_x L(x, \lambda, \nu) \leqslant f_0(x').$$

Поскольку решение задачи x_* также является допустимой точкой, получаем, что при $\lambda \geqslant 0$ двойственная функция дает нижнюю оценку на минимум:

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

§1.3 Двойственная задача

Итак, двойственная функция для любой пары (λ, ν) с $\lambda > 0$ дает нижнюю оценку на минимум в оптимизационной задаче. Попробуем теперь найти наилучшую нижнюю оценку:

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

Данная задача называется *двойственной* к задаче (1.1). Заметим, что функционал в двойственной задаче всегда является вогнутым.

Разберём несколько примеров построения двойственных задач.

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

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

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

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

■

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:

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

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, если $c + A^T \nu - \lambda \neq 0$. Таким образом, условие $c + A^T \nu - \lambda = 0$ является 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:

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

■

§1.4 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

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

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

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

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

§1.5 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 условия, которые выполнены для решений прямой и двойственной задач x_* и (λ^*, ν^*):

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

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 (KKT).*

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

Посмотрим на примере, как из условий Куна-Таккера можно найти решение задачи оптимизации.

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:

- $x + y = 4,\ x + 3y = 9,\ \lambda_1 > 0,\ \lambda_2 > 0.$
  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.

- $x + y = 4,\ x + 3y < 9,\ \lambda_1 > 0,\ \lambda_2 = 0.$
  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.

■

§1.6 Экономическая интерпретация двойственной задачи

Предположим, что мы хотим открыть фирму. В нее мы можем нанимать программистов и менеджеров — обозначим их количество через x_1 и x_2 соответственно. При этом каждый программист будет приносить c_1 рублей в месяц, а каждый менеджер — c_2 рублей. Труд каждого сотрудника должен оплачиваться. Наша фирма может платить в двух формах — акциями и картошкой, причем в месяц каждому программисту нужно выдать a_{11} акций и a_{21} килограммов картошки; для менеджеров эти числа обозначим через a_{12} и a_{22}. Разумеется, наши возможности ограничены: мы можем тратить не больше b_1 акций и b_2 килограммов картошки в месяц. Запишем формально все эти соотношения:

$$\begin{cases} c_1 x_1 + c_2 x_2 \to \max\limits_{x_1, x_2} \\ a_{11}x_1 + a_{12}x_2 \leqslant b_1 \\ a_{21}x_1 + a_{22}x_2 \leqslant b_2 \\ x_1 \geqslant 0, x_2 \geqslant 0 \end{cases}$$

Это задача линейного программирования, для которой легко найти двойственную:

$$\begin{cases} b_1 y_1 + b_2 y_2 \to \min\limits_{y_1, y_2} \\ a_{11}y_1 + a_{21}y_2 \geqslant c_1 \\ a_{12}y_1 + a_{22}y_2 \geqslant c_2 \\ y_1 \geqslant 0, y_2 \geqslant 0 \end{cases}$$

Двойственную задачу можно проинтерпретировать следующим образом. Допустим, что у нас появились другие дела, и вместо открытия фирмы мы решили продать все ресурсы (т.е. акции и картошку). Разумеется, наши покупатели будут стремиться установить максимально низкую цену — иными словами, они будут минимизировать общую сумму сделки $b_1 y_1 + b_2 y_2$, где через y_1 и y_2 обозначены цены на одну акцию и на один килограмм картошки соответственно. При этом у нас есть ограничение: мы не хотим продавать ресурсы дешевле, чем могли бы на них заработать, если бы все же открыли фирму. Это означает, что суммарная стоимость a_{11} акций и a_{21} килограммов картошки (т.е. размер оплаты одного программиста) не должна быть меньше, чем доход от одного программиста c_1. Это требование, вкупе с аналогичным требованием к размеру оплаты менеджера, как раз соответствует ограничениям в двойственной задаче.

Поскольку для данных задач имеет место сильная двойственность, их решения будут совпадать. Это означает, что оптимальная прибыль, которую можно получить при открытии фирмы, совпадает с оптимальной выгодой от продажи всех ресурсов.

Список литературы

[1] *Boyd, S., Vandenberghe, L.* Convex Optimization. // Cambridge University Press, 2004.