Семинары по линейным классификаторам

Евгений Соколов

27 октября 2013 г.

2 Оптимизационные задачи и теорема Куна-Таккера

Рассмотрим задачу минимизации

$$\begin{cases} f_0(x) \to \min\limits_{x \in \mathbb{R}^d} \\ f_i(x) \leqslant 0, \quad i = 1, \dots, m, \\ h_i(x) = 0, \quad i = 1, \dots, p. \end{cases} \tag{2.1}$$

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.

§2.1 Лагранжиан

Задача условной оптимизации (2.1) 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 (2.1). 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.

§2.2 Двойственная функция

Двойственной функцией для задачи (2.1) 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 (2.1) 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:

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

§2.3 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:

$$\begin{cases} g(\lambda, \nu) \to \max\limits_{\lambda, \nu} \\ \lambda_i \geqslant 0, \quad i = 1, \ldots, m. \end{cases} \tag{2.2}$$

Данная задача называется *двойственной* к задаче (2.1). Заметим, что функционал в двойственной задаче всегда является вогнутым.

Задача 2.1. *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 2.2. *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 $-x \neq 0$. 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:

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

■

§2.4 Сильная и слабая двойственность

Пусть (λ^*, ν^*) — решение двойственной задачи. Значение двойственной функции в этой 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 $Ax - b$).

Вспомним, например, двойственную задачу линейного программирования:

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

Здесь максимизируемая функция и ограничения являются линейными, и по ослабленным условиям Слейтера для сильной двойственности достаточно потребовать, чтобы существовала допустимая точка.

§2.5 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 (2.1) 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^*):

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

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 2.1. *Пусть x_* — решение задачи* (2.1). *Тогда найдутся такие векторы λ^* и ν^*, что выполнены условия* (KKT).

Если задача (2.1) является выпуклой и удовлетворяет условию Слейтера, то условия Куна-Таккера становятся *необходимыми и достаточными*.

Задача 2.3. *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.

■

Список литературы

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