Метод опорных векторов (Support Vector Machines). Нелинейная оптимизация.

Тверская Е. С.
e_tverskaya@bmstu.ru

Московский государственный технический университет им. Н. Э. Баумана
(национальный исследовательский университет)

Москва, 2025

ИУ-6
Компьютерные
системы и сети

Нелинейная оптимизация.
Прямая задача оптимизации.

Пусть заданы вещественные функции: $\boldsymbol{f(w)}$, $\boldsymbol{\varphi_i(w)}$ и $\boldsymbol{g_j(w)}$, где $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{i} = \overline{\boldsymbol{1, m}}$, $\boldsymbol{j} = \overline{\boldsymbol{1, k}}$. Для простоты и не нарушая общности изложения, положим $\boldsymbol{k = m}$.

Нелинейная оптимизация.
Прямая задача оптимизации.

ИУ-6 Компьютерные системы и сети

Пусть заданы вещественные функции: $f(w)$, $\varphi_i(w)$ и $g_j(w)$, где $w \in \mathbb{R}^n$, $i = \overline{1,m}$, $j = \overline{1,k}$. Для простоты и не нарушая общности изложения, положим $k = m$.

Постановка задачи.

Необходимо найти

$$\begin{aligned} &\inf_w f(w); \\ &g_i(w) \leqslant 0, \quad i = \overline{1,m}; \\ &\varphi_i(w) = 0, \quad i = \overline{1,m}. \end{aligned} \tag{1}$$

Нелинейная оптимизация.

Прямая задача оптимизации.

Пусть заданы вещественные функции: $\boldsymbol{f(w)}$, $\boldsymbol{\varphi_i(w)}$ и $\boldsymbol{g_j(w)}$, где $\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{n}}$, $\boldsymbol{i = \overline{1,m}}$, $\boldsymbol{j = \overline{1,k}}$. Для простоты и не нарушая общности изложения, положим $\boldsymbol{k = m}$.

Постановка задачи.

Необходимо найти

$$\begin{aligned} &\inf_{\boldsymbol{w}} \boldsymbol{f(w)}; \\ &\boldsymbol{g_i(w) \leqslant 0, \quad i = \overline{1,m};} \\ &\boldsymbol{\varphi_i(w) = 0, \quad i = \overline{1,m}.} \end{aligned} \tag{1}$$

Область допустимых решений

$$\boldsymbol{\Omega} = \{\boldsymbol{w} \in \mathbb{R}^{\boldsymbol{n}} : \ \boldsymbol{g(w) \leqslant 0, \ \varphi(w) = 0}\}.$$

Нелинейная оптимизация.
Прямая задача оптимизации.

Пусть заданы вещественные функции: $f(w)$, $\varphi_i(w)$ и $g_j(w)$, где $w \in \mathbb{R}^n$, $i = \overline{1, m}$, $j = \overline{1, k}$. Для простоты и не нарушая общности изложения, положим $k = m$.

Постановка задачи.

Необходимо найти

$$\begin{aligned} &\inf_w f(w); \\ &g_i(w) \leqslant 0, \quad i = \overline{1, m}; \\ &\varphi_i(w) = 0, \quad i = \overline{1, m}. \end{aligned} \tag{1}$$

Область допустимых решений

$$\Omega = \{w \in \mathbb{R}^n : \; g(w) \leqslant 0, \; \varphi(w) = 0\}.$$

Определение

Решение задачи оптимизации – это вектор $w^* \in \Omega \subset \mathbb{R}^n$ такой, что не существует $w \in \mathbb{R}^n$ такого, что $f(w) < f(w^*)$.

Данное определение дает понятие ***глобального минимума***. Если $w^* \in U \subset \Omega$, то тогда говорят о ***локальном минимуме***, где $U = U(w^*)$.

Функция f назвается ***целевой функцией***.

Определение

Функция f называется ***выпуклой***, если для всех $w, u \in \mathbb{R}^n$ и $0 \leqslant \lambda \leqslant 1$ выполнено

$$f(\lambda w + (1-\lambda)u) \leqslant \lambda f(w) + (1-\lambda) f(u)$$

Если имеются только ограничения типа равенств $\varphi(w) = 0$, то записывают функцию Лагранжа

$$L(w, \beta) = f(w) + \beta \varphi(w).$$

И необходимые условия минимума записываются как

$$\frac{\partial L(w,\beta)}{\partial w} = 0, \quad \frac{\partial L(w,\beta)}{\partial \beta} = 0$$

Если $L(w, \beta)$ – выпуклая функция, то эти условия являются и достаточными.
Если рассматривается задача общего вида (1) то функция Лагранжа записывается как

$$L(w, \beta) = f(w) + \beta \varphi(w) + \alpha g(w).$$

Нелинейная оптимизация.
Двойственная задача оптимизации.

Часто, ***двойственная задача оптимизации*** проще, чем прямая, так как имеет более простые граничные условия
Пусть

$$\boldsymbol{\Theta(\alpha, \beta) = \inf_{w} L(\alpha, \beta, w).}$$

Двойственная задача оптимизации заключается в том, чтобы найти

$$\boldsymbol{\max_{\alpha,\beta} \Theta(\alpha, \beta), \quad \alpha_i \geqslant 0.} \tag{2}$$

Теорема

Слабая теорема двойственности. Пусть вектор $\boldsymbol{w}$ удовлетворяет условиям $\boldsymbol{g(w) \leqslant 0}$ и $\boldsymbol{\varphi(w) = 0}$ прямой задачи оптимизации. В частности, он может быть решением прямой задачи оптимизации, а $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ – решение двойственной задачи (2). Тогда $\boldsymbol{f(w) \geqslant \Theta(\alpha, \beta)}$.

Доказательство.

$$\boldsymbol{\Theta(\alpha, \beta) = \inf_{u} L(u, \alpha, \beta) \leqslant L(w, \alpha, \beta) = f(w) + \alpha g(w) + \beta h(w) \leqslant f(w),}$$

так как $\boldsymbol{\alpha g(w) \leqslant 0}$ ($\boldsymbol{\alpha \geqslant 0}$) и $\boldsymbol{h(w) = 0}$. □

Нелинейная оптимизация.
Двойственная задача оптимизации.

Следствие 1.

Значение решения двойственной задачи не превосходи значения решения прямой задачи:

$$\sup\{\Theta(\alpha,\beta):\ \alpha \geqslant 0\} \leqslant \inf\{f(w):\ g(w) \leqslant 0,\ \varphi(w) = 0\}.$$

Следствие 2.

Достаточные условия того, что решения прямой и двойственной задач совпадают. Если $f(w^*) = \Theta(\alpha^*, \beta^*)$, где $\alpha^* \geqslant 0$, $g(w^*) \leqslant 0$, $\varphi(w^*) = 0$, то w^* и α^*, β^* – решения прямой и двойственной задачи, соответсвенно. Причем, в этом случае, также $\alpha^* g(w^*) = 0$.

Определение

Функция $h(w)$ называется ***аффинной***, если она имеет вид $Aw + b$, где A – некоторая матрица.

Теорема

Пусть $\Omega \subset \mathbb{R}$ – область допустимых решений (выпуклое множество). Функции h и g – аффинные. Тогда значение прямой и двойственной задач совпадают.

Нелинейная оптимизация.

Теорема Куна-Таккера. Условия Каруша-Куна-Таккера.

Теорема

Пусть $\boldsymbol{\Omega} \subset \mathbb{R}$ – область допустимых решений (выпуклое множество). Функции $\boldsymbol{h}$ и $\boldsymbol{g}$ – аффинные. Тогда $\boldsymbol{w}^*$ является решением прямой задачи

$$\inf f(w), \; w \in \Omega,$$
$$g(w) \leqslant 0, \quad h(w) = 0$$

тогда и только тогда, когда существует пара $\boldsymbol{\alpha}^*$ и $\boldsymbol{\beta}^*$ такая, что

$$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial w} = 0.$$

$$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial \beta} = 0, \tag{3}$$

$$\alpha_i^* g_i(w^*) = 0, \; i = \overline{1, m}, \tag{4}$$

$$g_i(w^*) \leqslant 0, \quad \alpha_i^* \geqslant 0, \quad i = \overline{1, m}.$$

Нелинейная оптимизация.

Теорема Куна-Таккера. Условия Каруша-Куна-Таккера.

Условие (3) задает условие достижения максимума по $\boldsymbol{\beta}$ линейной по $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ функции $\boldsymbol{L(w^*, \alpha, \beta)}$. Данное условие эквивалентно условию $\boldsymbol{h(\omega) = 0}$.
Условия максимума $\boldsymbol{L(w^*, \alpha, \beta)}$ по $\boldsymbol{\alpha}$ содержатся в (4), так как при $\boldsymbol{\alpha^* > 0}$, каждое такое условие превращается в $\boldsymbol{g_i(w^*) = 0}$ (это эквивалентно $\boldsymbol{\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial \alpha_i} = 0}$), а при $\boldsymbol{g(w^*) < 0}$ в точке максимума должно выполняться $\boldsymbol{\alpha^* = 0}$
Условие (4) называется ***условием Каруша-Куна-Таккера***.