§3 ЗАДАЧА ВЫПУКЛОЙ ОПТИМИЗАЦИИ

Рассмотрим следующую задачу условной оптимизации

$$\begin{cases} f_0(x) \to \min; \\ f_i(x) \le 0, \quad 1 \le i \le m; \\ x \in X; \end{cases} \tag{3.1}$$

где $f_j\colon \mathbb{R}^n \to \mathbb{R}$ — выпуклые функции, $0 \le j \le m$, а $X \subset \mathbb{R}^n$ — произвольное выпуклое множество. Задача (3.1) называется задачей выпуклой оптимизации (выпуклого программирования), а функция $f_0(x)$ — целевой функцией задачи (3.1). Множество $Y \stackrel{\text{def}}{=} X \cap \{x\colon f_i(x) \le 0, 1 \le i \le m\} \subset \mathbb{R}^n$ будем называть множеством допустимых векторов (точек). Очевидно, что Y — выпуклое множество. Допустимую точку $x \in Y$, для которой выполнены неравенства $f_i(x) < 0$, $1 \le i \le m$, будем называться строго допустимой. Ограничение $f_j(x) \le 0$ называется активным в допустимой точке $x \in Y$, если $f_j(x) = 0$. Множество индексов активных ограничений обозначим через $I(x) \stackrel{\text{def}}{=} \{j\colon f_j(x) = 0, 1 \le j \le m\}$.

Опр. 3.1. *Допустимый вектор $x^* \in Y$ называется решением задачи* (3.1), *если $f_0(x^*) \le f_0(x)$ при $x \in Y$.*

В общем случае, когда функции f_j не обязательно выпуклы, приводят определения локального и глобального экстремумов. Однако, очевидно, что для выпуклых задач эти понятия совпадают. Более того, если дополнительно известно, что f_0 — строго выпуклая функция, то задача (3.1) имеет не более одного решения.

Иногда к ограничением задачи (3.1) добавляют следующее $Ax = b$, где $A \in \in \mathbb{R}^{k\times n}$, $b \in \mathbb{R}^k$ (отметим, что поверхность уровня $\{x\colon f(x) = c\}$ произвольной выпуклой функции $f(x)$, вообще говоря, не является выпуклым множеством). Пусть $\widetilde{x}$ — какое-либо решение линейного уравнения $Ax = b$ и $K \in \mathbb{R}^{n\times d}$ — матрица, столбцы которой образуют базис $\ker A$, $\dim\ker A = d$. Тогда, $\{\widetilde{x} + Ky\colon y \in \mathbb{R}^d\}$ — множество всех решений системы $Ax = b$. Таким образом, исходная задача равносильна задаче (3.1) для функций $\widetilde{f}_i(y) \stackrel{\text{def}}{=} f_i(\widetilde{x} + Ky)$, $0 \le i \le m$, и множества $\widetilde{X} = K^{-1}(X - \widetilde{x})$.

Лемма 3.1. *Пусть функция f_0 из задачи* (3.1) *является дифференцируемой, тогда точка $x^* \in Y$ — решение задачи* (3.1), *если и только если для любой точки $x \in Y$ справедливо неравенство*

$$\nabla f_0(x^*)^{\mathsf{T}}(x - x^*) \ge 0. \tag{3.2}$$

$\triangleright$ Пусть x^* — решение задачи (3.1), докажем, что верно неравенство (3.2). Предположим противное, т.е., что нашлась такая допустимая точка $\widetilde{x} \in Y$, что $\nabla f(x^*)^{\mathsf{T}}(\widetilde{x} - x^*) < 0$. Так как f_0 — дифференцируемая функция, то имеет место равенство $f_0\big(x^* + t(\widetilde{x} - x^*)\big) = f_0(x^*) + t\nabla f_0(x^*)^{\mathsf{T}}(\widetilde{x} - x^*) + o(t)$, $t \in [0, 1]$. При достаточно малом $t > 0$ слагаемое $t\big(\nabla f_0(x^*)^\mathsf{T}(\widetilde{x} - x^*) + o(t)/t\big)$ отрицательное, а значит, $f_0\big(x^* + t(\widetilde{x} - x^*)\big) < f_0(x^*)$, что противоречит выбору x^*.

Предположим, что для некоторой точки $x^* \in Y$ выполнено неравенство (3.2). Так как $f_0(x) \geq f_0(x^*) + \nabla f_0(x^*)^\mathsf{T}(x - x^*)$, $x \in Y$, то $f_0(x) \geq f_0(x^*)$, а значит, x^* — решение задачи (3.1). $\lhd$

Опр. 3.2. *Говорят, что для задачи* (3.1) *выполнено условие Слейтера, если множество строго допустимых точек задачи* (3.1) *не пусто.*

Следуя [1, с. 52–58], перейдём к доказательству фундаментального результата, который является прямым аналогом метода множителей Лагранжа. Напомним, что функция $\mathcal{L}(x; \lambda_0, \lambda) \overset{\text{def}}{=} \sum\limits_{j=0}^{m} \lambda_j f_j(x)$, где $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, называется функцией Лагранжа задачи (3.1), а числа λ_0, λ_1, …, λ_m — множителями Лагранжа.

Теорема 3.1 (Кун, Таккер). *Пусть $f_j\colon \mathbb{R}^n \to \mathbb{R}$, $0 \leq j \leq m$, — выпуклые функции, а X — выпуклое множество. Если x^* является решением задачи* (3.1), *то найдутся такие множители Лагранжа λ_0^* и $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)$, что*

a) (условие невырожденности) числа λ_0^*, λ_1^*, …, λ_m^* не равны 0 одновременно;

b) (условие неотрицательности) $\lambda_j^* \geq 0$, $0 \leq j \leq m$;

c) (условия дополняющей нежёсткости) $\lambda_i^* f_i(x^*) = 0$, $1 \leq i \leq m$.

d) (принцип минимума) $\min\limits_{x \in X} \mathcal{L}(x; \lambda_0^*, \lambda^*) = \mathcal{L}(x^*; \lambda_0^*, \lambda^*)$;

Пусть для некоторых множителей λ_0^*, λ^* и допустимой точки $x^* \in Y$ выполнены условия a) – d), тогда

A) x^* — решение задачи (3.1), *если $\lambda_0^* \neq 0$;*

B) $\lambda_0^* \neq 0$, если для задачи (3.1) *справедливо условие Слейтера.*

$\rhd$ Пусть x^* — решение задачи (3.1). Без нарушения общности будем считать, что $f_0(x^*) = 0$. Действительно, если это не так, то определим новую функцию $\widetilde{f}_0(x) = f_0(x) - f_0(x^*)$. Рассмотрим множество $C \subset \mathbb{R}^{m+1}$, состоящее из таких векторов $\mu = (\mu_0, \mu_1, \ldots, \mu_m)^\mathsf{T}$, для которых найдётся точка $x_\mu \in X$, такая что выполнены неравенства

$$f_0(x_\mu) < \mu_0, \quad f_1(x_\mu) \leq \mu_1, \quad \ldots, \quad f_m(x_\mu) \leq \mu_m. \tag{3.3}$$

Установим ряд свойств множества C. Сперва докажем, что множество C непусто и выпукло. Действительно, любой вектор $\mu \in \mathbb{R}^{m+1}$ с положительными компонентами принадлежит C, так как в (3.3) для такого вектора достаточно положить $x = x^*$. Выпуклость множества C устанавливается аналогично доказательству выпуклости надграфика $\operatorname{epi} f$ произвольной выпуклой функции f.

Докажем, что нулевой вектор $\mathbf{0} \in \mathbb{R}^{m+1}$ не принадлежит C. Предположим противное. Тогда существует такая точка $\widetilde{x} \in X$, что $f_0(\widetilde{x}) < 0$ и $f_i(\widetilde{x}) \leq 0$, $1 \leq i \leq m$, а значит, x^* не является решением задачи (3.1).

Поскольку C — выпуклое множество и $\mathbf{0} \notin C$, то из теоремы 1.2 отделимости следует, что найдутся такие числа λ_0^*, λ_1^*, …, λ_m^*, неравные одновременно нулю, что $\sum\limits_{j=0}^{m} \lambda_j^* \mu_j \geq 0$, $\mu \in C$. Докажем, что λ_0^*, λ^* — искомые множители Лагранжа.

Множители λ_j^*, $0 \leq j \leq m$, неотрицательны. Действительно, очевидно, что вектор $(\delta, \dots, \delta, 1, \delta, \dots \delta)^{\mathsf{T}}$, где $\delta > 0$ и 1 стоит на j_0-м месте, принадлежит C, а значит, $\lambda_{j_0}^* \geq -\delta \sum\limits_{j \neq j_0} \lambda_j^*$. Так как $\delta > 0$ выбрано произвольно, то $\lambda_{j_0}^* \geq 0$.

Множители λ_i^*, $1 \leq i \leq m$, удовлетворяют условиям дополняющей нежёсткости. Выберем индекс i_0. Если $f_{i_0}(x^*) = 0$, то $\lambda_{i_0}^* f_{i_0}(x^*) = 0$. Предположим, что $f_{i_0}(x^*) < 0$. Очевидно, что вектор $(\delta, 0, \dots, 0, f_{i_0}(x^*), 0, \dots, 0)^{\mathsf{T}}$, где $\delta > 0$ и число $f_{i_0}(x^*)$ стоит на i_0-м месте, принадлежит C. Следовательно, $\lambda_{i_0}^* f_{i_0}(x^*) \geq -\lambda_{i_0}^* \delta$. Таким образом, $\lambda_{i_0}^* f_{i_0}(x^*) \geq 0$, а значит, $\lambda_{i_0}^* = 0$, т.е. $\lambda_{i_0}^* f_{i_0}(x^*) = 0$.

В точке x^* выполнен принцип минимума. Действительно, пусть $x \in X$. Тогда вектор $\big(f_0(x) + \delta, f_1(x), \dots, f_m(x)\big)^{\mathsf{T}}$, где $\delta > 0$, принадлежит множеству C. Следовательно, $\lambda_0^* f_0(x) + \sum\limits_{i=1}^{m} \lambda_i^* f_i(x) \geq -\lambda_0^* \delta$, а значит, $\mathcal{L}(x; \lambda_0^*, \lambda^*) \geq 0$. С другой стороны, $f_0(x^*) = 0$ и выполнены условия дополняющей нежёсткости, поэтому $\mathcal{L}(x^*; \lambda_0^*, \lambda^*) = 0$. Таким образом, $\mathcal{L}(x; \lambda_0^*, \lambda^*) \geq \mathcal{L}(x^*; \lambda_0^*, \lambda^*)$ для любого $x \in X$.

Пусть теперь для некоторых множителей λ_0^*, λ^* и допустимой точки $x^* \in Y$ выполнены условия a) – d). Предположим, что $\lambda_0^* \neq 0$. Без нарушения общности будем считать, что $\lambda_0^* = 1$. Тогда для любой допустимой точки $x \in Y$ получаем

$$f_0(x) \geq f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) = \mathcal{L}(x; 1, \lambda^*) \geq \mathcal{L}(x^*; 1, \lambda^*).$$

Так как $\mathcal{L}(x^*; 1, \lambda^*) = f_0(x^*) + \sum\limits_{i=1}^{m} \lambda_i^* f_i(x^*) = f_0(x^*)$, то x^* — решение задачи (3.1).

Предположим, что для задачи (3.1) выполнено условие Слейтера. Следовательно, существует точка $\widetilde{x} \in Y$, такая что $f_i(\widetilde{x}) < 0$, $1 \leq i \leq m$. Докажем, что $\lambda_0^* \neq 0$. Предположим противное. Тогда $\mathcal{L}(\widetilde{x}; 0, \lambda^*) = \sum\limits_{i=1}^{m} \lambda_i^* f_i(\widetilde{x}) < 0$, так как не все множители λ_i^*, $1 \leq i \leq m$, равны нулю. С другой стороны, $\mathcal{L}(x^*; 0, \lambda^*) = 0$, а значит, $\mathcal{L}(\widetilde{x}; 0, \lambda^*) < \mathcal{L}(x^*; 0, \lambda^*)$. Получено противоречие. $\lhd$

Функция $\mathcal{L}(x; \lambda) \overset{\text{def}}{=} f_0(x) + \sum\limits_{i=1}^{m} \lambda_i f_i(x)$, заданная на множестве $X \times \mathbb{R}_+^m$, где $\mathbb{R}_+^m \overset{\text{def}}{=} \{\lambda \in \mathbb{R}^m \colon \lambda_i \geq 0\}$, называется нормальной функцией Лагранжа.

Лемма 3.2. *Пусть $(x^*, \lambda^*) \in X \times \mathbb{R}_+^m$. Тогда точка x^* допустимая, т.е. $x^* \in Y$, и для пары (x^*, λ^*) выполнены условия a) – d) теоремы Куна – Таккера, если и только если (x^*, λ^*) — седловая точка нормальной функции Лагранжа, т.е.*

$$\min_{x \in X} \mathcal{L}(x; \lambda^*) = \mathcal{L}(x^*; \lambda^*) = \max_{\lambda \in \mathbb{R}_+^m} \mathcal{L}(x^*; \lambda). \tag{3.4}$$

$\rhd$ Действительно, пусть для пары $(x^*, \lambda^*) \in Y \times \mathbb{R}_+^m$ выполнены условия a) – d). Необходимо доказать только правое неравенство в (3.4). Для произвольных множителей $\lambda \in \mathbb{R}^m_+$ в силу условий b) и c) имеем

$$\mathcal{L}(x^*;\lambda^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) = f_0(x^*) \geq f_0(x^*) + \sum_{i=1}^m \lambda_i f_i(x^*) = \mathcal{L}(x^*;\lambda).$$

Пусть теперь пара $(x^*, \lambda^*) \in X \times \mathbb{R}^m_+$ — седловая точка функции $\mathcal{L}(x;\lambda)$. Докажем, только что $x^* \in Y$ и справедливо условие c), так как остальные условия, очевидно, выполнены. Если $f_{i_0}(x^*) > 0$ для некоторого $i_0 \geq 1$, то имеет место неравенство $\mathcal{L}(x^*;\lambda^*) < \mathcal{L}(x^*;\widetilde{\lambda})$, где $\widetilde{\lambda} \overset{\text{def}}{=} (\lambda_1^*, \ldots, \lambda_{i_0}^* + \delta, \ldots, \lambda_m^*)^{\mathsf{T}}$ и $\delta > 0$, которое противоречит (3.4). Следовательно, x^* — допустимая точка задачи (3.1). Так как $\mathcal{L}(x^*;\lambda^*) \geq \mathcal{L}(x^*;0)$, то $\sum\limits_{i=1}^m \lambda_i^* f_i(x^*) \geq 0$, а значит, $\lambda_i^* f_i(x^*) = 0$, $1 \leq i \leq m$. $\triangleleft$

Приведём критерий существования седловой точки, из доказательства которого, в частности, будет следовать способ её построения.

Теорема 3.2. *Пусть $f(x,y)\colon X \times Y \to \mathbb{R}$ — непрерывная функция, заданная на произведении компактных множеств X и Y. Тогда*

$$\min_x \max_y f(x,y) \geq \max_y \min_x f(x,y). \tag{3.5}$$

При этом равенство в (3.5) *достигается тогда и только тогда, когда существует седловая точка (x_0, y_0) функции f :*

$$f(x_0, y) \leq f(x_0, y_0) \leq f(x, y_0), \quad x \in X,\ y \in Y.$$

$\triangleright$ Выберем произвольную точку $\widetilde{y} \in Y$. Так как $\max\limits_y f(x,y) \geq f(x,\widetilde{y})$, $x \in X$, то $\min\limits_x \max\limits_y f(x,y) \geq \min\limits_x f(x,\widetilde{y})$, а значит, в силу произвольного выбора $\widetilde{y}$ выполнено неравенство (3.5).

Предположим, что $\min\limits_x \max\limits_y f(x,y) = \max\limits_y \min\limits_x f(x,y)$. Выберем точки $x_0 \in X$ и $y_0 \in Y$ из условий $\max\limits_y f(x_0,y) = \min\limits_x \max\limits_y f(x,y)$ и $\min\limits_x f(x,y_0) = \max\limits_y \min\limits_x f(x,y)$. Тогда $f(x_0,y_0) \geq \min\limits_x f(x,y_0) = \max\limits_y \min\limits_x f(x,y) = \min\limits_x \max\limits_y f(x,y)$. Поэтому, справедливо неравенство $f(x_0,y_0) \geq \max\limits_y f(x_0,y)$. Аналогично, получаем

$$f(x_0,y_0) \leq \max_y f(x_0,y) = \min_x \max_y f(x,y) = \max_y \min_x f(x,y) = \min_x f(x,y_0).$$

Таким образом, (x_0, y_0) — седловая точка.

Пусть теперь известно, что (x_0, y_0) — седловая точка. Тогда

$$\max_y \min_x f(x,y) \geq \min_x f(x,y_0) \geq f(x_0,y_0),$$
$$\min_x \max_y f(x,y) \leq \max_y f(x_0,y) \leq f(x_0,y_0).$$

Следовательно, $\max\limits_y \min\limits_x f(x,y) \geq \min\limits_x \max\limits_y f(x,y)$ и неравенство (3.5) обращается в равенство. $\triangleleft$

Отметим, что в доказательстве теоремы 3.2 условия компактности множеств X, Y и непрерывности функции $f(x, y)$ мы неявно использовали лишь для существования векторов x_0, y_0 при построении седловой точки.

Пример 3.1 (Метод опорных векторов, SVM)**.** Сформулируем задачу обучения с учителем. Пусть $f\colon X \to Y$ — отображение из пространства объектов X в множество ответов Y. Отображение f, вообще говоря, не известно, однако, дана обучающая выборка $S = \{(x_i, y_i)\}_{i=1}^N$ размера N, где $x_i \in X$ и $y_i = f(x_i) \in Y$, $1 \le i \le N$. Требуется построить отображение $\widetilde{f}\colon X \to Y$, аппроксимирующее f на всём пространстве X.

Рассмотрим частный случай задачи обучения с учителем — задачу бинарной классификации, в которой $Y = \{-1, 1\}$ и объекты описываются n-мерными вещественными векторами, т.е. $X = \mathbb{R}^n$. Далее будем считать, что в обучающей выборе содержатся объекты двух классов, а искомое отображение $\widetilde{f}$ будем строить в форме линейного порогового классификатора $\widetilde{f}(x) = \operatorname{sign}(\omega^\mathsf{T} x - \omega_0)$, где $\omega \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ и $\omega_0 \in \mathbb{R}$ — параметры, которые необходимо определить.

Предположим, что выборка S строго линейно разделима, т.е. существуют такие значения параметров ω и ω_0, при которых справедливы неравенства $y_i(\omega^\mathsf{T} x_i - \omega_0) > 0$, $1 \le i \le N$. В этом случае разделяющая гиперплоскость, вообще говоря, не единственна. Идея метода опорных векторов (support vector machine) состоит в выборе такой разделяющей гиперплоскости, которая максимально далеко отстоит от ближайших к ней точек обоих классов.

Заметим, что параметры ω и ω_0 линейного порогового классификатора $\widetilde{f}$ определяются с точностью до умножения на одну и ту же ненулевую константу. Поэтому, без ограничения общности будем считать, что $\min\limits_{1 \le i \le N} y_i(\omega^\mathsf{T} x_i - \omega_0) = 1$. Ориентированное расстояние от точки x_i до гиперплоскости, заданной уравнением $\omega^\mathsf{T} x = \omega_0$, равно $(\omega^\mathsf{T} x_i - \omega_0)/\|\omega\|$. Поэтому для определения параметров ω и ω_0 необходимо решить задачу

$$\begin{cases} \|\omega\| \to \min; \\ \min\limits_{1 \le i \le N} y_i(\omega^\mathsf{T} x_i - \omega_0) = 1; \end{cases}$$

которая эквивалентна следующей

$$\begin{cases} \dfrac{1}{2}\|\omega\|^2 \to \min; \\ y_i(\omega^\mathsf{T} x_i - \omega_0) \ge 1, \quad 1 \le i \le N. \end{cases} \tag{3.6}$$

В силу предположения о строгой линейной разделимости множество строго допустимых точек задачи (3.6) не пусто, а значит, выполнено условие Слейтера. Несложно показать, что задача (3.6) имеет решение. Действуя согласно теореме Куна – Таккера и лемме 3.2, найдём седловую точку нормальной функции Лагранжа:

$$\mathcal{L}(\omega, \omega_0; \lambda) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^N \lambda_i\big(y_i(\omega^\mathsf{T} x_i - \omega_0) - 1\big) \to \min_{\omega, \omega_0} \max_{\lambda}. \tag{3.7}$$

Фиксируя $\lambda \in \mathbb{R}^N_+$, решим задачу $\mathcal{L}(\omega, \omega_0; \lambda) \to \min\limits_{\omega, \omega_0}$. Из условия стационарности имеем

$$\frac{\partial \mathcal{L}}{\partial \omega} = \omega - \sum_{i=1}^N \lambda_i y_i x_i = 0, \quad \text{т.е.} \quad \omega = \sum_{i=1}^N \lambda_i y_i x_i; \quad \frac{\partial \mathcal{L}}{\partial \omega_0} = \sum_{i=1}^N \lambda_i y_i = 0. \tag{3.8}$$

Из (3.8), в частности, следует, что вектор ω является линейной комбинацией тех векторов обучающей выборки, для которых $\lambda_i \neq 0$. Согласно условиям дополняющей нежёсткости для этих векторов справедливы равенства $\omega^\mathsf{T} x_i - \omega_0 = y_i$. Такие векторы называются опорными. Используя равенства (3.8), преобразуем задачу (3.7) к задаче квадратичного программирования, содержащую только двойственные переменные:

$$\begin{cases} -\mathcal{L}(\lambda) \stackrel{\text{def}}{=} -\min\limits_{\omega,\omega_0} \mathcal{L}(\omega,\omega_0;\lambda) = -\sum\limits_{i=1}^{N} \lambda_i + \dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} \lambda_i\lambda_j y_i y_j x_i^\mathsf{T} x_j \to \min\limits_{\lambda}; \\ \sum\limits_{i=1}^{N} \lambda_i y_i = 0; \\ \lambda \geq 0. \end{cases} \tag{3.9}$$

Пусть λ — решение задачи (3.9). Тогда вектор ω вычисляется согласно (3.8). Для определения порога ω_0 достаточно взять произвольный опорный вектор x_i и положить $\omega_0 = \omega^\mathsf{T} x_i - y_i$. Однако, из-за возможных погрешностей вычислений рекомендуется брать такой опорный вектор x_i при определении ω_0, для которого двойственная переменная λ_i максимальна.

Рассмотрим общий случай, не делая предположений о линейной разделимости выборки. При любом выборе параметров ω и ω_0 линейный классификатор $\widetilde{f}$ может ошибаться на объектах выборки. Введём набор дополнительных переменных $\xi_i \geq 0$, характеризующих величину ошибки на объектах x_i, $1 \leq i \leq N$. Уточним задачу (3.6):

$$\begin{cases} \dfrac{1}{2}\|\omega\|^2 + C\sum\limits_{i=1}^{N} \xi_i \to \min\limits_{\omega,\omega_0,\xi}; \\ y_i(\omega^\mathsf{T} x_i - \omega_0) \geq 1 - \xi_i, \quad 1 \leq i \leq N; \\ \xi \geq 0; \end{cases} \tag{3.10}$$

где $C > 0$ — некоторый заданный гиперпараметр, определяющий компромисс между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки. Очевидно, что задача (3.10) имеет решение и множество строго допустимых точек не пусто, а значит, выполнено условие Слейтера. Рассмотрим функцию Лагранжа для задачи (3.10):

$$\mathcal{L}(\omega,\omega_0,\xi;\lambda,\eta) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{N} \lambda_i\big(y_i(\omega^\mathsf{T} x_i - \omega_0) - 1\big) - \sum_{i=1}^{N} \xi_i(\lambda_i + \eta_i - C),$$

где $\eta = (\eta_1, \eta_2, \ldots, \eta_N)^\mathsf{T}$ — вектор переменных, двойственных к вектору переменных $\xi = (\xi_1, \xi_2, \ldots, \xi_N)^\mathsf{T}$. Согласно теореме Куна – Таккера и лемме 3.2 задача (3.10) сводится к поиску седловой точки функции Лагрнажа: $\mathcal{L}(\omega,\omega_0,\xi;\lambda,\eta) \to \min\limits_{\omega,\omega_0,\xi} \max\limits_{\lambda,\eta}$. Зафиксируем произвольные векторы λ и η. Из условия стационарности по аргументам ω и ω_0 получаем

$$\frac{\partial\mathcal{L}}{\partial\omega} = \omega - \sum_{i=1}^{N} \lambda_i y_i x_i = 0, \quad \text{т.е.} \quad \omega = \sum_{i=1}^{N} \lambda_i y_i x_i; \quad \frac{\partial\mathcal{L}}{\partial\omega_0} = \sum_{i=1}^{N} \lambda_i y_i = 0; \tag{3.11}$$

Если для некоторого индекса $1 \leq i \leq N$ верно неравенство $\lambda_i + \eta_i > C$, то, очевидно, $\min\limits_{\omega,\omega_0,\xi} \mathcal{L}(\omega,\omega_0,\xi;\lambda,\eta) = -\infty$, а значит, такие множители Лагранжа λ и η не могут быть компонентами седловой точки. Поэтому будем предполагать, что $\lambda_i + \eta_i \leq C$. При этом, выполнено равенство $\xi_i(\lambda_i + \eta_i - C) = 0$, а именно, если $\lambda_i + \eta_i < C$, то мы должны положить $\xi_i = 0$. Используя (3.11), сведём задачу поиска седловой точки к задаче квадратичного программирования:

$$\begin{cases} -\mathcal{L}(\lambda) \stackrel{\text{def}}{=} -\min\limits_{\omega,\omega_0,\xi} \mathcal{L}(\omega,\omega_0,\xi;\lambda,\eta) = -\sum\limits_{i=1}^{N} \lambda_i + \dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} \lambda_i\lambda_j y_i y_j x_i^{\mathsf{T}} x_j \to \min\limits_{\lambda}; \\ \sum\limits_{i=1}^{N} \lambda_i y_i = 0; \\ 0 \leq \lambda_i \leq C, \quad 1 \leq i \leq N. \end{cases} \tag{3.12}$$

Пусть λ — решение задачи (3.12). Аналогично случаю линейной разделимой выборки, вектор ω вычисляется согласно (3.11). Порог ω_0 определим как $\omega_0 = \arg\min\limits_{\omega_0} \sum\limits_{i=1}^{N} \xi_i(\omega_0)$, где $\xi_i(\omega_0) = \max\bigl(0, 1 - y_i(\omega^{\mathsf{T}} x_i - \omega_0)\bigr)$. Несложно показать, что решение имеет вид $\omega_0 = \omega^{\mathsf{T}} x_i - y_i$ для некоторого i, а значит, ω_0 может быть найдено за не более чем $O(N \log N)$ арифметических операций. Однако, как правило, ω_0 удаётся определить гораздо быстрее следующим образом. Без нарушения общности будем считать, что $\eta_i = C - \lambda_i$, $1 \leq i \leq N$. Если для некоторого индекса i выполнено двойное неравенство $0 < \lambda_i < C$, то $\eta_i > 0$. Из теоремы Куна – Таккера следует, что $\xi_i = 0$ и $y_i(w^{\mathsf{T}} x_i - \omega_0) = 1$, а значит, $\omega_0 = \omega^{\mathsf{T}} x_i - y_i$.

Упражнения

9. Постройте эффективный алгоритм решения задачи $\sum\limits_{i=1}^{n} \max(0, a_i x + b_i) \to \min\limits_{x}$, где a_i и b_i, $1 \leq i \leq n$, — заданные действительные числа.
10. Докажите, что квадратичная функция $f(x) = x^{\mathsf{T}} A x + b^{\mathsf{T}} x$ либо достигает своей нижней грани на $\mathbb{R}^n$, либо не ограничена снизу.
11. Докажите, что для того, чтобы точка $x^* \in X$ была решением задачи выпуклого программирования (3.1), достаточно, чтобы для любого вектора v, удовлетворяющего системе неравенств $v^{\mathsf{T}} \nabla f_j(x^*) \leq 0$, $j \in I(x^*)$, было верно $v^{\mathsf{T}} \nabla f_0(x^*) \geq 0$.

Литература

1. Алексеев В.М., Тихомиров В.М., Фомин С.В. *Оптимальное управление* – М.: Наука, 1979. – 432 с.