# Машинное обучение, ФКН ВШЭ
# Семинар №14

## 1 Условные задачи оптимизации

**Задача 1.1.** *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.

■

## 2 Построение ядер

Напомним, что ядром мы называем функцию $K(x, z)$, представимую в виде скалярного произведения в некотором пространстве: $K(x, z) = \langle \varphi(x), \varphi(z) \rangle$, где $\varphi : \mathbb{X} \to H$ — отображение из исходного признакового пространства в некоторое *спрямляющее пространство* $H$.

Вспомним, какие функции в принципе могут быть ядрами — по теореме Мерсера функция $K(x, z)$ является ядром тогда и только тогда, когда:

1. Она симметрична: $K(x, z) = K(z, x)$.

2. Она неотрицательно определена, то есть для любой конечной выборки $(x_1, \dots, x_\ell)$ матрица $K = \bigl(K(x_i, x_j)\bigr)_{i,j=1}^{\ell}$ неотрицательно определена.

**Задача 2.1.** *Покажите, что если $K(x, z)$ — ядро, то оно симметрично и неотрицательно определено.*

**Решение.** Функция $K(x, z)$ — ядро, то есть она определяет скалярное произведение в некотором пространстве: $K(x, z) = \langle \varphi(x), \varphi(z) \rangle$. Симметричность этой функции вытекает из симметричности скалярного произведения.

Покажем неотрицательную определенность. Пусть $(x_1, \dots, x_\ell)$ — выборка, а $K = \bigl(K(x_i, x_j)\bigr)_{i,j=1}^{\ell}$ — матрица ядра, соответствующая ей. Тогда для произвольного вектора $v$:

$$
\begin{aligned}
\langle Kv, v \rangle &= \sum_{i,j=1}^{\ell} v_i v_j K(x_i, x_j) = \\
&= \sum_{i,j=1}^{\ell} v_i v_j \langle \varphi(x_i), \varphi(x_j) \rangle = \\
&= \sum_{i,j=1}^{\ell} \langle v_i \varphi(x_i), v_j \varphi(x_j) \rangle = \\
&= \left\langle \sum_{i=1}^{\ell} v_i \varphi(x_i), \sum_{j=1}^{\ell} v_j \varphi(x_j) \right\rangle = \\
&= \left\| \sum_{i=1}^{\ell} v_i \varphi(x_i) \right\|^2 \geqslant 0.
\end{aligned}
$$

Мы доказали неотрицательную определенность матрицы $K$, а значит и ядра $K(x, z)$.

■

Вместо того, чтобы проверять эти свойства, можно сразу составлять ядра по фиксированным правилам. Вспомним две следующие теоремы.

**Теорема 2.1.** *Пусть $K_1(x, z)$ и $K_2(x, z)$ — ядра, заданные на множестве $\mathbb{X}$, $f(x)$ — вещественная функция на $\mathbb{X}$, $\varphi : \mathbb{X} \to \mathbb{R}^N$ — векторная функция на $\mathbb{X}$, $K_3$ — ядро, заданное на $\mathbb{R}^N$. Тогда следующие функции являются ядрами:*

1. $K(x, z) = K_1(x, z) + K_2(x, z)$,

2. $K(x,z) = \alpha K_1(x,z)$, $\alpha > 0$,

3. $K(x,z) = K_1(x,z)K_2(x,z)$,

4. $K(x,z) = f(x)f(z)$,

5. $K(x,z) = K_3(\varphi(x), \varphi(z))$.

**Теорема 2.2.** *Пусть $K_1(x,z), K_2(x,z), \ldots$ — последовательность ядер, причем предел*

$$K(x,z) = \lim_{n\to\infty} K_n(x,z)$$

*существует для всех $x$ и $z$. Тогда $K(x,z)$ — ядро.*

**Задача 2.2.** *Покажите, что произведение ядер является ядром (третий пункт теоремы 2.1).*

**Решение.** Пусть ядро $K_1$ соответствует отображению $\varphi_1 : \mathbb{X} \to \mathbb{R}^{d_1}$, а ядро $K_2$ — отображению $\varphi_2 : \mathbb{X} \to \mathbb{R}^{d_2}$. Определим новое отображение, которое соответствует всевозможным произведениям признаков из первого и второго спрямляющих пространств:

$$\varphi_3(x) = \Big( (\varphi_1(x))_i \, (\varphi_2(x))_j \Big)_{i,j=1}^{d_1,d_2}.$$

Соответствующее этому спрямляющему пространству ядро примет вид

$$\begin{aligned} K_3(x,z) = \langle \varphi_3(x), \varphi_3(z) \rangle &= \\ &= \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} (\varphi_3(x))_{ij} \, (\varphi_3(z))_{ij} = \\ &= \sum_{i=1}^{d_1} (\varphi_1(x))_i \, (\varphi_1(z))_i \sum_{j=1}^{d_2} (\varphi_2(x))_j \, (\varphi_2(z))_j = \\ &= K_1(x,z) K_2(x,z). \end{aligned}$$

Мы показали, что произведение двух ядер соответствует скалярному произведению в некотором спрямляющем пространстве, а значит является ядром.

∎

**Задача 2.3.** *Пусть $p(x)$ — многочлен с положительными коэффициентами. Покажите, что $K(x,z) = p(\langle x, z\rangle)$ — ядро.*

**Решение.** Пусть многочлен имеет вид

$$p(x) = \sum_{i=0}^{m} a_i x^i.$$

Будем доказывать требуемое утверждение по шагам.

1. $\langle x, z\rangle$ —ядро по определению ($\varphi(x) = x$);
2. $\langle x, z\rangle^i$ — ядро как произведение ядер;
3. $a_i\langle x, z\rangle^i$ — ядро как произведение положительной константы на ядро;
4. константный член $a_0$ — ядро по пункту 4 теоремы 2.1, где $f(x) = \sqrt{a_0}$;
5. $\sum_{i=0}^{m} a_i\langle x, z\rangle^i$ — ядро как линейная комбинация ядер.

■

## §2.1 Спрямляющие пространства

Иногда может оказаться полезным знать не только вид ядра $K(x, z)$, но и вид преобразования $\varphi(x)$, и наоборот. Рассмотрим данный переход на нескольких примерах.

**Задача 2.4.** *Рассмотрим ядро на пространстве всех подмножеств конечного множества $D$:*

$$K(A_1, A_2) = 2^{|A_1\cap A_2|}.$$

*Покажите, что оно соответствует отображению в $2^{|D|}$-мерное пространство*

$$(\varphi(A))_U = \begin{cases} 1, U \subseteq A, \\ 0, \textit{иначе}, \end{cases}$$

*где $U$ пробегает по всем подмножествам множества $D$.*

**Решение.** Покажем, что при использовании указанного отображения $\varphi(A)$ скалярное произведение в спрямляющем пространстве действительно имеет указанный вид:

$$\langle\varphi(A_1), \varphi(A_2)\rangle = \sum_{U\subseteq D} (\varphi(A_1))_U\, (\varphi(A_2))_U\, .$$

Заметим, что $(\varphi(A_1))_U\,(\varphi(A_2))_U = 1$ только в том случае, если $(\varphi(A_1))_U = 1$ и $(\varphi(A_2))_U = 1$, т.е. если $U \subseteq A_1$ и $U \subseteq A_2$. Таким образом,

$$\langle\varphi(A_1), \varphi(A_2)\rangle = \left|\{U \subseteq D\,|\, U \subseteq A_1, U \subseteq A_2\}\right|.$$

Подсчитаем количество таких множеств. Рассмотрим некоторое $U \subseteq A_1 \cap A_2$. Заметим, что все прочие подмножества $D$ не будут удовлетворять хотя бы одному из условий, в то время как для таким образом выбранного $U$ выполняются оба, поэтому необходимое число — число различных подмножеств $A_1 \cap A_2$. Оно, в свою очередь, равно $2^{|A_1\cap A_2|}$.

■

**Задача 2.5.** *Рассмотрим ядро*

$$K(x, z) = \prod_{j=1}^{d}(1 + x_j z_j).$$

*Какому спрямляющему пространству оно соответствует?*

**Решение.** Раскроем скобки в выражении для $K(x, z)$. Заметим, что итоговое выражение будет включать мономы всех чётных степеней от 0 до $2d$ включительно. При этом мономы степени $2k$, $k \in \{0, \ldots, d\}$, формируются следующим образом: из $d$ скобок, входящих в произведение, случайным образом выбираются $k$, после чего входящие в них слагаемые вида $x_j z_j$ умножаются на единицы, входящие в состав остальных $d - k$ скобок. Таким образом, в итоговое выражение входят все мономы степени $2k$ над всеми наборами из $k$ различных исходных признаков, и только они. Запишем это формально:

$$K(x, z) = (1 + x_1 z_1)(1 + x_2 z_2) \ldots (1 + x_d z_d) = \sum_{k=0}^{d} \sum_{\substack{D \subseteq \{1, \ldots, d\} \\ |D| = k}} \prod_{j \in D} x_j z_j.$$

Для простоты понимания приведем вид итогового выражения для $d = 2, 3$ (несложно убедиться в его справедливости путём раскрытия скобок):

$$\begin{aligned} K((x_1, x_2), (z_1, z_2)) &= 1 + x_1 z_1 + x_2 z_2 + x_1 x_2 z_1 z_2, \\ K((x_1, x_2, x_3), (z_1, z_2, z_3)) &= 1 + x_1 z_1 + x_2 z_2 + x_3 z_3 + x_1 x_2 z_1 z_2 + \\ &\qquad x_1 x_3 z_1 z_3 + x_2 x_3 z_2 z_3 + x_1 x_2 x_3 z_1 z_2 z_3. \end{aligned}$$

Таким образом, объект $x$ в спрямляющем пространстве представим в следующем виде:

$$\varphi(x) = (1, x_1, \ldots, x_d, x_1 x_2, \ldots, x_1 x_d, \ldots, x_{d-1} x_d, \ldots, x_1 x_2 \ldots x_d) = \left( \prod_{j \in D} x_j \right)_{D \subseteq \{1, \ldots, d\}},$$

то есть в виде вектора мономов всех степеней над наборами различных признаков в исходном пространстве.

■

**Задача 2.6.** *Пусть* $\{(x_i, y_i)\}_{i=1}^{\ell}$, $y_i \in \{-1, +1\}$ *— произвольная выборка, а* $\varphi(x)$ *— отображение в спрямляющее пространство, соответствующее гауссову ядру. Покажите, что в данном спрямляющем пространстве существует линейный классификатор, безошибочно разделяющий выборку* $\varphi(x_1), \dots, \varphi(x_\ell)$.

**Решение.** Покажем, что вектор весов $w$ в спрямляющем пространстве может быть найден как линейная комбинация объектов выборки $\varphi(x_1), \dots, \varphi(x_\ell)$, т.е. $w = \sum_{i=1}^{\ell} \alpha_i \varphi(x_i)$. Запишем условие верной классификации каждого из объектов выборки в спрямляющем пространстве:

$$\langle w, \varphi(x_i) \rangle = y_i,\ i = \overline{1, \ell}.$$

Заметим, что записанное нами условие является более строгим, чем необходимо, однако в дальнейшем мы покажем существование $w$, удовлетворяющего этим более строгим ограничениям. Преобразуем:

$$\left\langle \sum_{j=1}^{\ell} \alpha_j \varphi(x_j), \varphi(x_i) \right\rangle = y_i,\ i = \overline{1, \ell},$$

$$\sum_{j=1}^{\ell} \alpha_j \langle \varphi(x_j), \varphi(x_i) \rangle = y_i,\ i = \overline{1, \ell},$$

$$\sum_{j=1}^{\ell} \alpha_j K(x_i, x_j) = y_i,\ i = \overline{1, \ell}.$$

Таким образом, мы получили систему из $\ell$ линейных уравнений на $\alpha_1, \dots, \alpha_\ell$, при этом матрицей системы является матрица Грама, являющаяся невырожденной (согласно утв. 1.3 лекции 13), а потому система имеет решение, и соответствующий вектор $w$ существует.

■

## §2.2 Ядра в метрических методах

Теперь, когда у нас есть общее представление о природе ядер, попробуем использовать их для усовершенствования уже известных нам методов — например, метрических. Как вы знаете, для использования данного класса алгоритмов необходимо задать функцию расстояния на пространстве объектов — однако при использовании ядер у нас не всегда есть возможность выразить $\varphi(x)$ в явном виде. Тем не менее, оказывается, ядро содержит в себе много информации о спрямляющем пространстве, и позволяет производить в нем различные операции, не зная самого отображения $\varphi(x)$.

**Задача 2.7.** *Как вычислить норму вектора* $\varphi(x)$*, зная лишь ядро* $K(x, z)$*?*

**Решение.**

$$\|\varphi(x)\| = \sqrt{\|\varphi(x)\|^2} = \sqrt{\langle \varphi(x), \varphi(x) \rangle} = \sqrt{K(x, x)}.$$

■

**Задача 2.8.** *Как вычислить расстояние между векторами $\varphi(x)$ и $\varphi(z)$, зная лишь ядро $K(x,z)$?*

**Решение.**

$$\begin{aligned}\rho^2(\varphi(x),\varphi(z)) &= \|\varphi(x)-\varphi(z)\|^2 = \langle\varphi(x)-\varphi(z),\varphi(x)-\varphi(z)\rangle = \\ &= \langle\varphi(x),\varphi(x)\rangle - 2\langle\varphi(x),\varphi(z)\rangle + \langle\varphi(z),\varphi(z)\rangle = \\ &= K(x,x) - 2K(x,z) + K(z,z).\end{aligned}$$

■

Таким образом, ядра можно использовать и в метрических методах (например, kNN) — достаточно подставить в них в качестве функции расстояния величину $\sqrt{K(x,x)-2K(x,z)+K(z,z)}$.

# 3 Метод опорных векторов

**Задача 3.1.** *Рассмотрим задачу с линейно разделимой выборкой. Допустим, мы решили двойственную задачу SVM и нашли вектор двойственных переменных $\lambda$. Покажите, что половина ширины разделяющей полосы $\rho$ может быть вычислена по следующей формуле:*

$$\frac{1}{\rho^2} = \sum_{i=1}^{\ell}\lambda_i.$$

**Решение.** Поскольку выборка линейно разделима, то все объекты, для которых $\lambda_i \neq 0$, окажутся на границе разделяющей полосы. Для них будет выполнено равенство

$$y_i\left(\langle w,x_i\rangle + b\right) = 1,$$

из которого можно выразить $b$:

$$b = y_i - \langle w,x_i\rangle.$$

Домножим обе стороны на $\lambda_i y_i$ и просуммируем по $i$ (заметим, что для объектов не на границе разделяющей полосы выполняется $\lambda_i y_i = 0$):

$$b\sum_{i=1}^{\ell}\lambda_i y_i = \sum_{i=1}^{\ell}\lambda_i y_i^2 - \sum_{i=1}^{\ell}\lambda_i y_i\langle w,x_i\rangle.$$

Поскольку $w$, $b$ и $\lambda$ здесь — решения прямой и двойственной задач, то для них выполнены условия Куна-Таккера. В частности,

$$\sum_{i=1}^{\ell}\lambda_i y_i = 0,$$

$$w = \sum_{i=1}^{\ell}\lambda_i y_i x_i.$$

Заметим также, что $y_i^2 = 1$. Воспользовавшись этими тремя равенствами, получаем:

$$0 = \sum_{i=1}^{\ell} \lambda_i - \|w\|^2.$$

Ранее мы доказали, что в SVM ширина разделяющей полосы равна $\frac{2}{\|w\|}$, поэтому

$$0 = \sum_{i=1}^{\ell} \lambda_i - \frac{1}{\rho^2}.$$

Отсюда получаем требуемое равенство.

■

**Задача 3.2.** *Пусть $(w, b, \xi_1, \dots, \xi_\ell)$ — оптимальное решение прямой задачи SVM. Предположим, что $\xi_3 > 0$. Выразите отступ объекта $x_3$ для обученного линейного классификатора через значения $(\xi_1, \dots, \xi_\ell)$.*

**Решение.** Заметим, что, поскольку $\xi_3 > 0$, то объект $x_3$ является опорным нарушителем. Отсюда следует, что $\lambda_3 = C$. Напомним, что для двойственной задачи можно записать условия дополняющей нежесткости:

$$\lambda_3[y_3 (\langle w, x_3 \rangle + b) - 1 + \xi_3] = 0,$$

откуда можно получить, что $y_3 (\langle w, x_3 \rangle + b) - 1 + \xi_3 = 0 \Leftrightarrow M_3 = y_3 (\langle w, x_3 \rangle + b) = 1 - \xi_3$.

■

**Задача 3.3.** *Пусть мы решили двойственную задачу SVM и получили решение $(\lambda_1, \dots, \lambda_\ell)$. Пусть мы также восстановили оптимальный порог $b$. Выразите:*

1. *Квадрат нормы $\|w\|^2$ оптимального вектора $w$ для прямой задачи;*
2. *Сумму $\sum_{i=1}^{\ell} \xi_i$ оптимальных значений параметров $\xi_1, \dots, \xi_\ell$ для прямой задачи.*

**Решение.**

1. Напомним, что из условий Куна-Таккера для двойственной задачи имеем $w = \sum_{i=1}^{\ell} \lambda_i y_i x_i$. Отсюда
$$\|w\|^2 = \langle w, w \rangle = \left\langle \sum_{i=1}^{\ell} \lambda_i y_i x_i, \sum_{j=1}^{\ell} \lambda_j y_j x_j \right\rangle = \sum_{i,j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle.$$
2. Напомним, что имеет место
$$\mu_i \xi_i = 0 \Leftrightarrow (\mu_i = 0) \text{ или } (\xi_i = 0),$$

поэтому имеет смысл рассматривать лишь те объекты, для которых $\mu_i = 0$. Из $\lambda_i + \mu_i = C$ имеем $\lambda_i = C \neq 0$. Отсюда и из $\lambda_i[y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i] = 0$ имеем

$$y_i\left(\langle w, x_i\rangle + b\right) - 1 + \xi_i = 0 \Leftrightarrow \xi_i = 1 - y_i\left(\langle w, x_i\rangle + b\right) =$$
$$= 1 - y_i\left(\left\langle \sum_{j=1}^{\ell} \lambda_j y_j x_j, x_i \right\rangle + b\right) = 1 - y_i\left(\sum_{j=1}^{\ell} \lambda_j y_j \langle x_i, x_j\rangle + b\right).$$

Отсюда имеем:

$$\sum_{[i=1]}^{\ell} \xi_i = \sum_{i=1}^{\ell}\left(1 - y_i\left(\sum_{j=1}^{\ell} \lambda_j y_j \langle x_i, x_j\rangle + b\right)\right) =$$
$$= \ell - \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \lambda_j \langle x_i, x_j\rangle - b\sum_{i=1}^{\ell} y_i.$$

$\blacksquare$