Машинное обучение, ФКН ВШЭ
Семинар №13

1 Ядра

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 $\varphi : X \to H$ — отображение из исходного признакового пространства в некоторое *спрямляющее пространство*. Оказывается, ядро содержит в себе много информации о спрямляющем пространстве, и позволяет производить в нем различные операции, не зная самого отображения $\varphi(x)$.

Задача 1.1. *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 1.2. *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 $K(x, x) - 2K(x, z) + K(z, z)$.

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 $K = \big(K(x_i, x_j)\big)_{i,j=1}^{\ell}$ неотрицательно определена.

Задача 1.3. *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 $(x_1, \ldots, x_\ell)$ — выборка, а $K = \big(K(x_i, x_j)\big)_{i,j=1}^{\ell}$ — 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 1.1. *Пусть $K_1(x, z)$ и $K_2(x, z)$ — ядра, заданные на множестве X, $f(x)$ — вещественная функция на X, $\varphi : X \to \mathbb{R}^N$ — векторная функция на 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 1.2. *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 1.4. *Покажите, что произведение ядер является ядром (третий пункт теоремы 1.1).*

Решение. Пусть ядро K_1 соответствует отображению $\varphi_1 : X \to \mathbb{R}^{d_1}$, а ядро K_2 — отображению $\varphi_2 : X \to \mathbb{R}^{d_2}$. 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 1.5. *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 1.1, где $f(x) = \sqrt{a_0}$;
5. $\sum_{i=0}^{m} a_i \langle x,z \rangle^i$ — ядро как линейная комбинация ядер.

■