Машинное обучение — 2

Ядровые методы

Полина Гусева, Анастасия Дроздова, Дарья Сапожникова
БПМИ181

Ядровые методы — класс алгоритмов машинного обучения, использующих ядровую функцию. За счет ядровой функции методы неявным образом оперируют объектами в многомерном, возможно бесконечномерном, пространстве. Большинство методов из этого класса хорошо статистически обоснованы.

Ядровые методы получили широкое распространение в конце 1990-х, когда было показано, что в задаче распознавания изображений из набора данных MNIST ядровый SVM показывает результаты, аналогичные нейронным сетям того времени [1].

Теория ядровых методов

Как мы уже знаем из линейной алгебры, *евклидово пространство* — это конечномерное вещественное векторное пространство с введенным на нем скалярным произведением.

Скалярное произведение $\langle \cdot, \cdot \rangle$ порождает на пространстве метрику:

$$\rho(x, y) = \sqrt{\langle x - y, x - y \rangle}$$

Таким образом, евклидово пространство является метрическим. Будем называть метрическое пространство *полным*, если каждая фундаментальная последовательность на нем сходится к элементу этого пространства. Можно показать, что любое евклидово пространство является полным [2].

Гильбертово пространство является обобщением евклидова пространства. Гильбертово пространство допускает бесконечномерность, но при этом требуется, чтобы пространство обязательно было полным. Полнота пространства гарантирует существование ортонормированного базиса, за счет чего с гильбертовыми пространствами часто можно работать по аналогии с конечномерными пространствами [3].

В дальнейшем будем рассматривать гильбертово пространство на функциях.

Будем говорить, что $K : X \times X \to \mathbb{R}$ — *воспрозводящее ядро* гильбертова пространства $\mathbb{H}$, если выполняются следующие условия:

- $\forall x \in X$ верно, что $K_x(z) = K(x, z) \in \mathbb{H}$
- $\forall x \in X, \forall f \in \mathbb{H}$ верно $f(x) = \langle f, K_x \rangle$

Для любого ядра существует гильбертово пространство, для которого оно является воспроизводящим.

Примечательно, что в гильбертовых пространствах с вопроизводящим ядром оптимизационные задачи имеют решение в довольно простой форме. Следующая теорема будет именно об этом.

Теорема о представителе[1]. Пусть $\mathbb{H}$ — гильбертово пространство с воспроизводящим ядром K и нормой $\|h\| = \sqrt{\langle h, h \rangle}$. Пусть $G : \mathbb{R} \to \mathbb{R}$ — монотонно возрастающая функция, а L — функция потерь. Тогда любое решение оптимизационной задачи

$$\arg\min_{h \in \mathbb{H}} \left(G(\|h\|) + L(h(x_1), \ldots, h(x_\ell)) \right)$$

имеет вид

$$x \mapsto \alpha_1 K(x, x_1) + \ldots + \alpha_\ell K(x, x_\ell)$$

Доказательство теоремы ~~оставим любопытному читателю~~ можно найти в Википедии.

Все это говорит о том, что ядровый метод по сути подбирает в соответствующем гильбертовом пространстве функцию, при помощи которой отражает объекты в спрямляющее пространство.

Приложения

У ядровых методов есть множество преимуществ. Помимо очевидных плюсов, таких как нелинейная разделяющая поверхность и эффективная работа со скалярными произведениями в многомерных пространствах, есть еще возможность работать с данными, которые не имеют естественного представления в виде вектора вещественных чисел. Это позволяет эффективно работать, к примеру, с текстами или даже последовательностями белков. Рассмотрим несколько популярных ядер, которые мы не успели затронуть на занятиях, и их области применения.

1. *Строковое ядро* [4]. Это ядро часто применяется в биоинформатике, потому что позволяет работать с последовательностями вроде ДНК или белков. При помощи этого ядра замеряется схожесть двух строк.

   Пусть Σ — алфавит, а s и t — аргументы ядра. Суть подхода заключается в том, что для каждого слова в алфавите Σ считается его вес в каждом из аргументов, а результатом является сумма произведений таких весов. Чем чаще слово входит в строку как подпоследовательность, тем больше вес. Если между элементами подпоследовательности в слове большое расстояние, то вес уменьшается.

[1] Про эту теорему я нашла только тексты на английском, где она называется Representer theorem. Возможно, она все же как-то называется на русском, но гугл мне не раскрыл эту тайну.

2. *Графовые ядра* [5]. Аналогично строковому ядру, графовые ядра замеряет схожесть пары графов. При помощи графовых ядер удается работать, например, с молекулами, так как

   Одним из графовых ядер является ядро случайного блуждания. Суть этого ядра заключается в том, что сначала в обоих графах осуществляется много случайных блужданий, а затем считается количество путей, которые были получены в обоих графах.

Список литературы

[1] Yann LeCun. The MNIST Database. http://yann.lecun.com/exdb/mnist/.

[2] Walter Rudin. Principles of Mathematical Analysis. https://proofwiki.org/wiki/Euclidean_Space_is_Complete_Metric_Space.

[3] Math StackExchange: The role of completeness in Hilbert spaces. https://mathoverflow.net/a/35846.

[4] Huma Lodhi et al. Text classification using string kernels. The Journal of Machine Learning Research, 2002.

[5] S.V.N. Vishwanathan et al. Graph kernels. The Journal of Machine Learning Research, 2010.