АЛГОРИТМ ОЦЕНКИ ВОСПРОИЗВОДЯЩЕГО ЯДРА ДЛЯ РЕШЕНИЯ ЗАДАЧИ МАШИННОГО ОБУЧЕНИЯ МЕТОДОМ КОЛЛОКАЦИИ

Бывшев В.А.

Финансовый университет при Правительстве Российской Федерации,
г. Москва, Ленинградский проспект, д.49
VByvshev@fa.ru

Аннотация: Обсуждается алгоритм оценки воспроизводящего ядра, необходимого для решения задачи машинного обучения методом коллокации. Метод коллокации восходит к Л.В. Канторовичу и состоит в восстановлении искомых функций по значениям функционалов. Алгоритм базируется на теореме Лоэва об интерпретации воспроизводящего ядра как автоковариационной функции и концепции Винера оценки автоковариационной функции стационарного ряда по индивидуальной реализации

Ключевые слова: Задача и метод машинного обучения, задача и метод коллокации, искомая функция и функционалы, воспроизводящее ядро.

Введение

Метод коллокации восходит к лауреату Нобелевской премии Л.В. Канторовичу [1] и состоит в восстановлении искомых функций по определённым на них значениям функционалов. В данной работе сначала излагаются задача и метод коллокации во взаимосвязи с постановкой задачи и методом машинного обучения. Затем строится алгоритм оценки воспроизводящего ядра, необходимого для решения методом коллокации задачи машинного обучения.

1 Задача машинного обучения

Задачи машинного обучения по прецедентам состоят в следующем [2], [3]. Задано множество X объектов $x \in X$ и задано множество Y допустимых ответов у $\in Y$. Каждый объект $x \in X$ описывается набором признаков $a(x) = (a_1(x), \dots, a_q(x))$, и с этим набором x отождествляется: $x \sim a(x)$. Каждый признак $a_j(x)$ имеет смысл некоторой характеристики объекта x и формально может интерпретироваться как отображение множества объектов X в множество A_j возможных значений признака a_j, то есть $a_j: X \to A_j$. В зависимости от типа множества A_j признак a_j может быть бинарным, категориальным или количественным. Отметим, что категориальные признаки можно квантифицировать, поэтому будем предполагать, что все признаки в наборе $a(x)$ являются количественными, то есть $a(x) \in R_q$.

По предположению, существует неизвестная функция $f: X \to Y$, значения которой $y_i = f(x_i)$ известны на конечном множестве объектов $(x_1, x_2, \dots, x_n) \subset X$. Пары $(a(x_i), y_i)$ именуются прецедентами; совокупность $S = (a(x_i), y_i)_{i=1}^{N}$ прецедентов образуют выборку. Выборку разделяют на две части: на обучающую выборку $S_l = (a(x_i), y_i)_{i=1}^{n}$ и контролирующую выборку $S_c = \left(a(x_{(i)}), y_{(i)}\right)_{(i)=1}^{(n)}$. Требуется по выборке S построить функцию $\hat{f}: X \to Y$, приближающую неизвестную функцию $f: X \to Y$ на всём множестве X.

Схема решения задачи машинного обучения состоит из трёх шагов [2].

Шаг 1. Задаётся некоторое параметрическое семейство Ф функций $\varphi(a; \theta)$:

(1.1) $$\Phi = \{\varphi(a; \theta) \big| \varphi: R_q \times \Theta \to \mathrm{Y}\}.$$

Аргументом функции $\varphi(a;\theta) \in \Phi$ служит вектор $a = a(x)$ признакового описания объекта $x \in X$, компоненты вектора θ являются искомыми параметрами функции $\varphi(a;\theta)$, при этом $\theta \in \Theta \subset R_m$, где $\Theta -$ множество допустимых значений параметров. Областью изменения функции $\varphi(a;\theta)$ служит множество ответов Y. Пусть $\theta -$ некоторый произвольный фиксированный вектор параметров, $a(x) -$ признаковое описание некоторого объекта $x \in X$. Значение

(1.2) $$\tilde{y} = \varphi(a(x);\theta)$$

имеет смысл оценки истинного ответа $y = f(x)$ на объекте x; разность

(1.3) $$\tilde{y} - y = e(a(x), \varphi, \theta)$$

является истинной ошибкой оценки $\tilde{y}$ ответа $y = f(x)$ на объекте $x \in X$ при векторе параметров θ. Качество вектора параметров θ измеряется значением некоторой функции ошибки $e(a(x), \varphi, \theta)$. Такую функцию принято именовать функцией потерь и обозначать $Loss$. Функция потерь является функцией аргументов $(a(x), \varphi, \theta)$. Пусть $Loss(a(x), \varphi, \theta)$ – некоторая функция потерь. Её значения зависят от типа функций $\varphi \in \Phi$, вектора параметров $\theta \in \Theta$ и объекта $x \in X$. Усреднённое по множеству X значение функции потерь обозначим $Loss(\varphi, \theta)$, то есть

(1.4) $$Loss(\varphi, \theta) = E_x(Loss(a(x), \theta)).$$

Значение $Loss(\varphi, \theta)$ зависит от типа функций $\varphi \in \Phi$ и вектора параметров $\theta \in \Theta$. Величина $Loss(\varphi, \theta)$ интерпретируется как мера риска прогноза (1.4) ответа $y = f(x)$ при заданном векторе параметров θ. На практике значение $Loss(\varphi, \theta)$, как правило, не известно, но всегда доступна его оценка, например, по обучающей выборке S_l:

(1.5) $$Loss(\widetilde{\varphi, \theta}, S_l) = \frac{1}{n}\sum_{i=1}^{n} Loss(a(x_i), \varphi, \theta),$$

именуемая эмпирическим риском.

Шаг 2. На обучающей выборке решается вариационная задача на условный экстремум

(1.6) $$\begin{cases} Loss(\widetilde{\varphi, \theta}, S_l) = \frac{1}{n}\sum_{i=1}^{n} Loss(a(x_i), \varphi, \theta) \to min \\ \theta \in \Theta \end{cases}.$$

То есть отыскивается такой вектор параметров $\hat{\theta} \in \Theta$, который минимизирует эмпирический риск на обучающей выборке.

Шаг 3. При найденном векторе параметров $\hat{\theta}$ вычисляется эмпирический риск на контролирующей выборке:

(1.7) $$Loss\left(\widetilde{\varphi, \hat{\theta}}, S_c\right) = \frac{1}{(n)}\sum_{(i)=1}^{(n)} Loss\left(a\left(x_{(i)}\right), \varphi, \hat{\theta}\right).$$

Если величина $Loss\widetilde{\left(\varphi, \hat{\theta}, S_c\right)}$ значительно не превосходит количество $Loss\widetilde{\left(\varphi, \hat{\theta}, S_l\right)}$, то есть если имеет место приближённое равенство

(1.8) $$Loss\widetilde{\left(\varphi, \hat{\theta}, S_c\right)} \approx Loss\widetilde{\left(\varphi, \hat{\theta}, S_l\right)},$$

то задача машинного обучения считается решённой; это означает, что найдена функция

(1.9) $$\tilde{y} = \hat{f}(x) = \varphi(a(x); \hat{\theta}),$$

приближающая неизвестную функцию $f: X \to Y$ на всём множестве X. Функцию (1.9) используют для прогноза ответов. В противном случае, когда величина $Loss\widetilde{\left(\varphi, \hat{\theta}, S_c\right)}$ значительно превосходит $Loss\widetilde{\left(\varphi, \hat{\theta}, S_l\right)}$, говорят [2], что произошло переобучение модели (1.14), меняют состав обучающей и контролирующей выборок и повторяют шаги 1, 2 и 3 до достижения условия (1.8).

2 Задача коллокации

Вернёмся к постановке задачи машинного обучения о восстановлении неизвестной функции $f: X \to Y$. Под символом X будем теперь понимать признаковый образ множества объектов в задаче машинного обучения, так что $X \subset R_q$. В задаче коллокации о восстановлении функции $f: X \to Y$ предполагается, что

искомая функция $y = f(x)$ определена на ограниченном замкнутом множестве X и принадлежит некоторому гильбертову пространству H с воспроизводящим ядром $K(x, x')$, где $(x, x') \in X \times X$;

известны значения $(l_i)_{i=1}^{n}$ ограниченных линейно-независимых функционалов $(L_i)_{i=1}^{n}$ на искомой функции $f(x)$:

(2.1) $$L_1(f) = l_1, L_2(f) = l_2, \ldots, L_n(f) = l_n.$$

Систему (2.1) можно интерпретировать как систему уравнений наблюдений над косвенными проявлениями искомой функции $f \in H$; эту систему удобно записать компактно

(2.1)′ $$L(f) = l,$$

где $L = (L_1, \dots, L_n)^T -$ векторный функционал, $l = (l_1, \dots, l_n)^T \in R_n$.

В задаче коллокации требуется по уравнениям наблюдений (2.1) восстановить функцию $f: X \to Y$, то есть найти такую аналитически заданную функцию $\hat{f}: X \to Y$, которая приближала бы f на X и удовлетворяла бы системе уравнений наблюдений (2.1)′:

(2.2) $$L(\hat{f}) = l.$$

Решение $\hat{f}$ задачи коллокации базируется на следующей теореме.

Теорема [4]. Если в системе уравнений наблюдений (2.1)′ компоненты $L_1, \dots, L_n$ векторного функционала L являются ограниченными и линейно-независимыми функционалами на гильбертовом пространстве H с воспроизводящим ядром $K(x, x')$, то

1) гильбертово пространство H разлагается в прямую сумму ортогональных подпространств $KerL = \{\varphi : \varphi \in H,\ L(\varphi) = 0\}$ и $Ker^{\perp}L$:

(2.3) $$H = KerL \oplus Ker^{\perp}L,$$

где бесконечномерное подпространство $KerL$ является ядром векторного функционала L, размерность подпространства $Ker^{\perp}L$ совпадает с количеством n линейно-независимых компонентов векторного функционала L, базис подпространства $Ker^{\perp}L$ образуют функции

(2.4) $$(K(x, L_1),\ K(x, L_2), \dots, K(x, L_n)) = K(x, L)^T ,$$

где символом $K(x, L_i)$ обозначен результат действия функционала L_i на воспроизводящее ядро $K(x, x')$ по аргументу $x' \in X$;

2) решение $\hat{f}$ задачи коллокации является элементом подпространства $Ker^{\perp}L$, то есть

(2.5) $$\hat{f} \in Ker^{\perp}L;$$

3) решение задачи коллокации $\hat{f}(x)$ вычисляется по правилу

(2.6) $$\hat{f}(x) = K(x, L)^T \cdot K^{-1} \cdot l ,$$

где символом $K = \left(K\left(L_i, L_j\right)\right)$ обозначена симметричная положительно определённая матрица

Предположим теперь, что $\hat{f}(x)$ имеет смысл восстановленной функции в задаче машинного обучения, где всегда

(2.7) $$L_i(f) = \delta_{x_i}(f) = f(x_i) , \quad F(f) = \delta_x(f) = f(x),$$

Тогда, согласно (2.4) и (2.6), оценка ответа для любого объекта $x \in X$ будет вычисляться методом коллокации по правилу

(2.8) $$\tilde{y} = \hat{f}(x) = \sum_{j=1}^{n} c_j \cdot K\left(x, x_j\right).$$

Чтобы применять метод коллокации, нужно иметь оценку воспроизводящего ядра $K(x, x')$.

3 Алгоритм оценки воспроизводящего ядра $\boldsymbol{K(x, x')}$

Ниже представлен алгоритм оценки $K(x, x')$ по выборке $S = (a(x_i), y_i)_{i=1}^{N}$, базирующийся на теореме Лоэва [5, стр 512] об интерпретации воспроизводящего ядра как автоковариационной функции и концепции Винера [6, стр. 17] оценки автоковариационной функции стационарного ряда по индивидуальной реализации. Мы принимаем традиционную в приложениях [4] гипотезу об изотропности функции $K(x, x')$, то есть гипотезу, что $K(x, x') = K(\rho(x, x'))$.

1. Вычислить $\bar{y} = \frac{1}{N} \sum_1^N y_i$ и $y_i = y_i - \bar{y}$ при $i = 1, 2, \dots, N$.

2. Вычислить $\rho_{ij} = \rho\left(x_i, x_j\right)$ при $1 \le i < j \le N$ и вычислить $\rho_{max} = max\rho_{ij}$.

3. Задать количество M узлов сетки $(\rho_1, \rho_2, \dots, \rho_M)$, в которых будут вычислены оценки $\big(K_1 = K(\rho_1), K_2 = K(\rho_2), \dots, K_M = K(\rho_M)\big)$; вычислить шаг сетки $\Delta\rho = \frac{\rho_{max}}{M}$ и радиус окрестности узла $r = \frac{\Delta\rho}{2}$.

4. Вычислить $K(0) = \frac{1}{N} \cdot \sum_1^N y_i^2$.

5. Принять $K_m = 0,\ N_m = 0$ при $m = 1,2, \dots, M$.

6. Цикл по $i = 1,2, \dots, N - 1$.

7. Цикл по $j = i + 1, \dots, N$
8. Принять ρ=ρ_{ij}
8. Цикл по $m = 1,2, \dots, M$
9. Принять $\rho_m = \Delta\rho \cdot m$
10. Проверить справедливость неравенства $(\rho_m - r \leq \rho < \rho_m + r)$. (3.1)
Если (3.1) несправедливо, то идти на конец цикла по m.
Если (3.1) справедливо, то вычислить $K_m = K_m + y_i \cdot y_j$ и вычислить $N_m = N_m + 1$.
11. Вычислить $K(\rho_m) = K_m/N_m$ при $m = 1,2, \dots, M$ и $N_m \neq 0$.
12. Конец

Литература

1. *Канторович Л.В*. Об одном методе приближённого решения дифференциальных уравнений в частных производных. – ДАН СССР, 1931, 2, № 9.
2. *Воронцов К.В.* Математические методы обучения по прецедентам (теория обучения машин), http://www.ccas.ru/voron..
3. *Соловьёв В.И.* Анализ данных в экономике. КНОРУС, М., 2019.
4. *Мориц Г.* Современная физическая геодезия. – М, «Недра», 1983.
5. *Лоэв М.* Теория вероятностей. – М., ИЛ, 1962.
6. *Wiener N.* Time Series. M.I.T. Press, Cambridge, Mass., 1949.