

Решение регрессионной задачи машинного обучения на основе теории случайных функций

Бахвалов Ю.Н., к.т.н., независимый исследователь, г. Череповец, Россия

Аннотация.

В статье исследуется регрессионная задача машинного обучения как задача многомерной аппроксимации с использованием аппарата теории случайных функций. Предложен первопринципный вывод метода регрессии из постулатов индифферентности. Показано, что если вероятностная мера в бесконечномерном функциональном пространстве обладает естественными симметриями (инвариантность к сдвигу, повороту, масштабированию и гауссовость), то вся решающая схема — включая вид ядра, форму регуляризации и параметризацию шума — выводится аналитически из этих постулатов. Полученное ядро соответствует обобщённому полигармоническому сплайну, однако в отличие от существующих подходов, оно здесь не выбирается эмпирически, а является следствием принципа индифферентности. Результат даёт теоретическое обоснование широкому классу методов сглаживания и интерполяции, показывая их оптимальность в условиях отсутствия априорной информации.

Ключевые слова: машинное обучение, регрессия, случайная функция, симметрия, корреляционная функция, спектральная плотность, полигармонический сплайн.

Solving a machine learning regression problem based on random function theory

Bakhvalov Y.N., Ph.D. Independent Researcher, Cherepovets, Russia

Abstract.

This article examines a machine learning regression problem as a multivariate approximation problem using the apparatus of random function theory. An ab initio derivation of the regression method from the indifference postulates is proposed. It is shown that if a probability measure in an infinite-dimensional function space possesses natural symmetries (invariance to translation, rotation, scaling, and Gaussianity), then the entire solution scheme—including the kernel type, regularization form, and noise parameterization—is analytically derived from these postulates. The resulting kernel corresponds to a generalized polyharmonic spline; however, unlike existing approaches, it is not chosen empirically, but is a consequence of the indifference principle. The result provides a theoretical

justification for a wide class of smoothing and interpolation methods, demonstrating their optimality in the absence of a priori information.

Keywords: machine learning, regression, random function, symmetry, correlation function, spectral density, polyharmonic spline.

Предшествующая версия данного исследования была опубликована автором в виде [19], что в свою очередь является развитием идей [1]. Настоящая работа существенно расширяет и перерабатывает исходный материал: переформулирована задача как задача аппроксимации, введено строгое описание через вероятностную меру, исправлены математические неточности, добавлены новые преобразования и обоснования. Математический аппарат и структура изложения полностью переработаны.

Пусть есть обучающая выборка в виде набора векторов на входе $x_1, x_2, \dots, x_k (x_i \in R^n)$ размерностью n и набор значений на выходе $y_1, y_2, \dots, y_k (y_i \in R)$.

Решение регрессионной задачи машинного обучения в этом случае можно представить, как решение задачи аппроксимации в виде:

$$y_i = f(x_i) + u_i \quad (1)$$

где $f(x)$ – функция, связывающая входные и выходные значения x_i и y_i , искомая модель, которую нужно определить, основываясь на обучающей выборке;

$u_1, u_2, \dots, u_k \in R$ – независимые случайные величины, с нормальным законом распределения и нулевым математическим ожиданием, которые, можно считать, что были добавлены к $f(x_i)$, и поэтому ее значения могут не совпадать точно с y_i (тем самым моделируем погрешности или неоднозначности, которые могут присутствовать в обучающей выборке)

Задачу также можно интерпретировать следующим образом:

Пусть изначально существовала некоторая неизвестная функция $f(x)$. Затем над ней были выполнены эксперименты. Для некоторой выборки $x_1, x_2, \dots, x_k$ были найдены значения $f(x_i)$ (но которые нам неизвестны), к которым были прибавлены случайные, неизвестные нам величины $u_1, u_2, \dots, u_k$. В результате к уже известным $x_1, x_2, \dots, x_k$ дополнительно были получены $y_1, y_2, \dots, y_k$.

Теперь же задача состоит в том, чтобы по последовательностям x_i и y_i предположить какая могла быть $f(x)$.

Первоначально может показаться (даже если известны характеристики величин u_i), что такая формулировка ничего не дает для решения, поскольку ничего неизвестно о природе $f(x)$ и то, какая она может быть в (1) никак не раскрывается.

Но рассмотрим сначала сильно упрощенную версию задачи.

Допустим, что у нас есть “подсказка”. Предположим, что кто-то заранее сообщил нам варианты правильных ответов, какая может быть функция $f(x)$. Пусть подсказка была в виде конечной последовательности вариантов функций $f_j(x)$ и соответствующих им вероятностей p_j , что это правильный ответ. Но вероятности этот кто-то в подсказке дал нам априорные, не зная, что затем над $f(x)$ были проведены эксперименты и у нас еще есть независимая дополнительная информация в виде последовательностей x_i и y_i .

Предположим, что нам известны дисперсии случайных величин $u_1, u_2, \dots, u_k$. Допустим, что они все одинаковы и равны σ^2 . Тогда вероятности того, что j -тый вариант функции в подсказке $f_j(x)$ окажется правильным ответом, будут пропорциональны $\hat{p}_j$:

$$\hat{p}_j = p_j \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^k u_i^2} \quad (2)$$

Где u_i можем определить как:

$$u_i = y_i - f_j(x_i)$$

Перебрав последовательно все $\hat{p}_j$ и найдя максимальное из них, тем самым выберем наилучший вариант $f_j(x)$ как наилучшее решение задачи аппроксимации в (1).

Но, что представляет собой множество вариантов функции $f(x)$ с заданными на нем вероятностями (или заданной на нем функции распределения вероятностей)? Это ни что иное как описание случайной функции.

Как видно выше, при наличии такой подсказки (заданной случайной функции с конечным множеством реализаций), сразу же было получено решение.

Но аналогичные по смыслу рассуждения (выполненные в этом простом примере) можно провести на гораздо более “крупном” множестве функций и без “подсказки” в явном виде.

Возьмем в качестве возможных вариантов функции $f(x)$ множество $\mathcal{F}$ всех непрерывных вещественных функций, существующих в R^n . Это

множество можно представить в бесконечномерном гильбертовом пространстве, каждое измерение которого это возможные значения функции для каждого некоторого конкретного значения $x \in R^n$.

На этом множестве в свою очередь может быть задана вероятностная мера μ (что в совокупности с $\mathcal{F}$ дает описание случайной функции)..

Следует отметить, что на этом множестве, будучи несчётным бесконечномерным пространством (в котором не существует меры Лебега), плотность вероятности в классическом смысле не определена. Однако на нём может быть задана вероятностная мера μ как предел согласованных конечномерных распределений.

Корректность такого подхода – стандартный предмет теории гауссовых мер в бесконечномерных пространствах (Bogachev, 1998 [3]; Vakhania et al., 1987 [16]). Существование μ при заданных согласованных конечномерных распределениях обеспечивается теоремой Колмогорова (Kallenberg, 2002 [11]; Гихман И. И., Скороход А. В. 1971 [9])

Эту вероятностную меру μ можно считать неизвестной. Однако, относительно этой меры, исходя из принципов индифферентности, можно допустить существование некоторых симметрий (которые вполне можно было предположить перед тем, как над $f(x)$ были выполнены эксперименты).

1. Инвариантность вероятностной меры, относительно преобразований сдвига и поворота системы координат.

Для любых M , t и преобразования $T: f_1 \rightarrow f_2$ связанных между собой:

$$f_2(x) = f_1(Mx + t), x, t \in R^n \quad (3)$$

где M - некоторая матрица поворота

И любого измеримого подмножества функций $A \subset \mathcal{F}$ должно быть выполнено $\mu(A) = \mu(T(A))$.

Т.е. для любых подмножеств функций, которые преобразуются друг в друга путем поворота или параллельного переноса вероятностная мера μ должна быть одинакова.

2. Инвариантность вероятностной меры относительно преобразований масштаба.

Для любого k и преобразования $T: f_1 \rightarrow f_2$ таких что:

$$f_2(x) = kf_1(x/k), x \in R^n \quad (4)$$

где $k \in R$ – некоторый коэффициент

И любого измеримого подмножества функций $A \subset \mathcal{F}$ должно быть выполнено $\mu(A) = \mu(T(A))$.

Т.е. для любых подмножеств функций, которые преобразуются друг в друга изменением масштаба вероятностная мера μ должна быть одинакова.

3. Вероятностная мера μ является бесконечномерной гауссовской мерой с нулевым средним.

Это означает, что любая конечномерная проекция случайной функции имеет многомерное нормальное распределение, а сама мера согласована по Колмогорову (Гихман И. И., Скороход А. В. [9]).

Для лучшего понимания содержания пункта 3, сформулируем два его следствия, имеющих наглядную интерпретацию.

Следствие 3.1. Из гауссовости вероятностной меры μ следует, что если какое-то подмножество функций порождается масштабированием одной базовой функции $f_1(x) \rightarrow kf_1(x)$ то условная плотность вероятности на этом множестве пропорциональна нормальному закону по параметру k с нулевым средним.

Т.е. если предположить, что решением является одна из функций на подмножестве функций, которые описываются как:

$$f_k(x) = kf_1(x), k \in R \quad (5)$$

где k – некоторое вещественное число (назовем его в данном контексте амплитудой $f_1(x)$)

$f_1(x)$ – некоторая произвольная непрерывная функция, взятая чтобы породить множество в (5)

Тогда условная плотность вероятности на этом подмножестве функций будет определяться нормальным законом от амплитуды k , с математическим ожиданием при $k = 0$.

Т.е. если мы берем любую непрерывную функцию (как одну из реализаций) и сравним ее с другой, которая повторяет первую, но отличается от нее только амплитудой, то функция с меньшей амплитудой будет более вероятна, чем с большей. А закон распределения вероятностей будет иметь характер нормального закона, зависящего от амплитуды (в (5) обозначенной через k) с математическим ожиданием равным нулю. В данном случае мы не уточняем, какая может быть дисперсия (подразумевая, что не нулевая, иначе говорить о нормальном законе распределения не имело бы смысла), а говорим лишь о характере закона распределения вероятностей.

Следствие 3.2. Если к каждой такой функции из множества сл.3.1 прибавить произвольную фиксированную $f_2(x)$ то распределение по k сохранит гауссов характер (хотя среднее может сместиться).

Прибавим ко всем функциям в множестве, что описано в сл.3.1 одну и ту же функцию $f_2(x)$ и получим новое множество функций, элементы которого можно описать как:

$$f_k^*(x) = kf_1(x) + f_2(x), k \in R \quad (6)$$

где k – некоторое вещественное число (амплитуда $f_1(x)$)

$f_1(x)$ и $f_2(x)$ – любые непрерывные функции

Какую бы мы функцию $f_2(x)$ не взяли, плотность вероятности на этом подмножестве функций (порожденная как конечномерная проекция меры μ) также как и в сл.3.1 сохранит нормальный закон распределения от амплитуды k (но математическое ожидание уже может не равняться нулю).

То есть, если мы возьмем произвольную непрерывную функцию $f_2(x)$, а затем также возьмем произвольную непрерывную функцию $f_1(x)$ и умножая на различные варианты амплитуды k будем прибавлять ее к $f_2(x)$, порождая таким образом множество функций $f_k^*(x)$ (как бы совершающие колебания вокруг $f_2(x)$), то плотность вероятности на этом множестве функций будет соответствовать нормальному закону распределения от амплитуды k (но математическое ожидание не обязательно будет соответствовать $k = 0$).

Мотивацию выбора именно таких видов симметрий п.1 – п.3 вероятностной меры μ можно обосновать принципом индифферентности (principle of indifference) при полном отсутствии априорной информации об задаче до получения обучающей выборки. В такой ситуации наиболее естественно предположить, что должны быть:

1. **Пространственная инвариантность (1):** Мера μ не должна зависеть от выбора начала системы координат (сдвиг t) и их ориентации (поворот M). Не существует выделенной точки или направления в пространстве R^n .
2. **Масштабная инвариантность (2):** Не существует выделенного масштаба наблюдения. Законы, связывающие x и y , должны иметь одинаковый вид при изменении единиц измерения (масштабирование k).
3. **Отсутствие выделенных амплитуд (3):** В рамках любого заданного "паттерна" функции $f_1(x)$ не существует априорно предпочтительной амплитуды его проявления. Наиболее общим и естественным допущением о распределении амплитуд является нормальный закон

(гауссовость), вытекающий из принципа максимальной энтропии при заданной дисперсии.

Таким образом, постулаты п.1 – п.3 представляют собой минимальный набор естественных предположений о симметриях функционального пространства в условиях полного отсутствия иной априорной информации.

Рассмотрим следствия из п.1-3, какие будут свойства у случайной функции, для которой они выполняются.

Любое невырожденное (что выполняется в нашем случае) многомерное нормальное распределение можно свести к множеству независимых нормальных случайных величин. Это означает, что обязательно найдется такая последовательность функций (образующих базис, по которому может быть разложена любая другая непрерывная функция), с помощью которой можно выразить рассматриваемую нами случайную функцию как линейную комбинацию этой последовательности функций и независимых случайных нормальных величин. А это не что иное как каноническое разложение случайной функции. Т.е. это значит, что у такой случайной функции обязательно должно существовать каноническое разложение.

Каноническое разложение случайной функции (В.С. Пугачев [13], стр.248-249) в случае счетного базиса:

$$F(x) = m_f(x) + \sum_{j=1}^{\infty} V_j \varphi_j(x), \quad (7)$$

где $m_f(x)$ – функция математического ожидания,

V_j – некоррелированные случайные величины, математические ожидания которых равны нулю (коэффициенты канонического разложения),

$\varphi_j(x)$ – координатные функции канонического разложения

Из п.3. следует, что $m_f(x) = 0$. А также, поскольку количество функций $\varphi_j(x)$ в нашем случае будет несчетным, (7) обобщается до интегрального канонического представления случайной функции:

$$F(x) = \int_{\Lambda} V(\lambda) \varphi(x, \lambda) d\lambda, \quad (8)$$

где $V(\lambda)$ – белый шум параметра λ ,

$\varphi(x, \lambda)$ – некоторая (неслучайная) функция аргумента x и параметра λ

Таким образом получается разложение всех возможных реализаций (непрерывных функций) по некоррелированным бесконечно малым

элементарным случайным функциям $V(\lambda)\varphi(x,\lambda)d\lambda$. Но пока мы лишь сделали вывод, что из п.3 о гауссовости меры следует существование такого разложения, а не то, каким оно должно быть.

Рассмотрим теперь факт существования канонического интегрального разложения (8), следующий из п.3. в свете того, что выполняется п.1.

Из инвариантности меры относительно операций всех возможных сдвигов и поворотов, применяемых к любым подмножествам функций (реализаций), следует строгая стационарность и изотропия случайной функции. А значит, она не просто представима в виде (8), а должна иметь спектральное разложение.

Интегральное каноническое представление стационарной случайной функции представлено В.С. Пугачевым в [13], стр.333. Но используемый там белый шум $V(\omega)$, интенсивность которого может быть различной для разных частот, мы можем представить напрямую как произведение составляющих

$$V(\omega) \rightarrow \frac{1}{2} V(\omega) \sqrt{\frac{S(\omega)}{d\omega}}, \quad (9)$$

где $V(\omega)$ будет уже иметь одинаковую интенсивность на всех частотах (почему использование (9) будет корректным, будет еще дополнительно показано ниже).

Вообще выражение (9) представляет собой формальный переход к новому представлению белого шума $V(\omega)$, которое обеспечивает постоянную единичную интенсивность (дисперсию) на всех частотах. Такое представление является стандартным методом в спектральной теории случайных процессов (Cramér, 1942 [7]) и эквивалентно более строгому определению через ортогональные стохастические меры. Множитель $1/2$ учитывает последующий переход к области интегрирования R_1^n и условию вещественности (11).

Таким образом, мы можем представить нашу случайную функцию в следующем виде:

$$F(x) = \frac{1}{2} \int_{R^n} V(\omega) \sqrt{\frac{S(\omega)}{d\omega}} e^{i\omega x} d\omega, \quad x, \omega \in R^n \quad (10)$$

где $S(\omega)$ – спектральная плотность, неотрицательная вещественная симметричная функция,

$V(\omega)$ – комплексная случайная функция, каждое значение которой для промежутка, соответствующего $d\omega$, является независимой случайной величиной (кроме дополнительного условия, о котором ниже), действительная и мнимая части которой распределены по нормальному закону с

математическим ожиданием равным нулю и дисперсией равной единице. Т.е. $V(\omega)$ это белый шум с одинаковой дисперсией на всех частотах.

Дополнительное условие: поскольку все возможные варианты реализаций $f(x)$ должны быть вещественными, значит для любых частот ω значения $V(\omega)$ и $V(-\omega)$ должны быть комплексно сопряженными.

Чтобы отразить это условие поделим пространство частот R^n на две части R_1^n и R_2^n какой-либо гиперплоскостью размерности $n - 1$, проходящей через начало координат.

Тогда $V(\omega)$ можно записать:

$$V(\omega) = \begin{cases} V_R(\omega) + iV_I(\omega), & \text{если } \omega \in R_1^n \\ V_R(\omega) - iV_I(\omega), & \text{если } \omega \in R_2^n \end{cases}, \quad (11)$$

где $V_R(\omega)$ и $V_I(\omega)$ – вещественные симметричные случайные функции, каждое значение которых для промежутка, соответствующего $d\omega$ из (10), одинаково для частот ω и $-\omega$, является независимой случайной величиной с математическим ожиданием равным нулю и дисперсией равной единице (симметричный вещественный белый шум).

Тогда (10) можно записать как:

$$F(x) = \frac{1}{2} \int_{R_1^n} (V_R(\omega) + iV_I(\omega)) \sqrt{\frac{S(\omega)}{d\omega}} e^{i\omega x} d\omega + \\ + \frac{1}{2} \int_{R_2^n} (V_R(\omega) - iV_I(\omega)) \sqrt{\frac{S(\omega)}{d\omega}} e^{i\omega x} d\omega \quad (12)$$

Поскольку для любой ω , если $\omega \in R_1^n$ то $-\omega \in R_2^n$ и наоборот, то во втором интеграле можно перейти к области интегрирования R_1^n и одновременно заменить везде ω на $-\omega$ в подынтегральном выражении.

$$F(x) = \frac{1}{2} \int_{R_1^n} (V_R(\omega) + iV_I(\omega)) \sqrt{\frac{S(\omega)}{d\omega}} e^{i\omega x} d\omega \\ + \frac{1}{2} \int_{R_1^n} (V_R(-\omega) - iV_I(-\omega)) \sqrt{\frac{S(-\omega)}{d\omega}} e^{-i\omega x} d\omega =$$

$$\begin{aligned}
&= \frac{1}{2} \int_{R_1^n} \sqrt{\frac{S(\omega)}{d\omega}} \left(V_R(\omega)(e^{i\omega x} + e^{-i\omega x}) + iV_I(\omega)(e^{i\omega x} - e^{-i\omega x}) \right) d\omega = \\
&= \int_{R_1^n} \sqrt{\frac{S(\omega)}{d\omega}} (V_R(\omega) \cos(\omega x) - V_I(\omega) \sin(\omega x)) d\omega \quad (13)
\end{aligned}$$

Таким образом мы получили интегральное каноническое представление нашей случайной функции (соответствующее записи (8)), выраженное через действительные функции и независимые действительные нормальные случайные величины с дисперсией равной единице.

Как показали преобразования, выражение (13) полностью соответствует комплексной форме (10), в которой белый шум был представлен как

$$\frac{1}{2} V(\omega) \sqrt{\frac{S(\omega)}{d\omega}}$$

Для более удобного наглядного представления дальнейших преобразований заменим интеграл (13) бесконечной суммой. Обозначим через v_j^R бесконечную последовательность значений $V_R(\omega)$ каждое из которых соответствует своему промежутку области интегрирования $d\omega$ в R_1^n (ему будет соответствовать частота ω_j). Введем аналогичную последовательности v_j^I для $V_I(\omega)$ и s_j для $S(\omega)$.

Тогда (13) можно записать:

$$F(x) = \sum_{j=0}^{\infty} \sqrt{\frac{s_j}{d\omega}} (v_j^R \cos(\omega_j x) - v_j^I \sin(\omega_j x)) d\omega \quad (14)$$

Подразумевая, что $d\omega$ это некоторая очень малая область (в пределе бесконечно малая, тогда (14) превращается в (13)).

Запишем выражение (14) чуть иначе:

$$F(x) = \sum_{j=0}^{\infty} v_j^R \sqrt{s_j d\omega} \cos(\omega_j x) - \sum_{j=0}^{\infty} v_j^I \sqrt{s_j d\omega} \sin(\omega_j x) \quad (15)$$

Но выражение (15), это то же самое каноническое разложение случайной функции вида (7), где функция математического ожидания равна нулю, v_j^R и v_j^I это независимые случайные величины с единичной дисперсией а роль координатных функций выполняют $\sqrt{s_j d\omega} \cos(\omega_j x)$ и $-\sqrt{s_j d\omega} \sin(\omega_j x)$.

Тогда мы можем выразить корреляционную функцию (каноническое разложение через координатные функции):

$$\begin{aligned}
K_f(x_1, x_2) &= \sum_{j=0}^{\infty} s_j (\cos(\omega_j x_1) \cos(\omega_j x_2) + \sin(\omega_j x_1) \sin(\omega_j x_2)) d\omega \\
&= \sum_{j=0}^{\infty} s_j \cos(\omega_j (x_1 - x_2)) d\omega
\end{aligned} \tag{16}$$

Т.е. мы получили автокорреляционную функцию (что и следовало ожидать для стационарной случайной функции):

$$k_f(\tau) = \sum_{j=0}^{\infty} s_j \cos(\omega_j \tau) d\omega \tag{17}$$

Если теперь в (17) обратно вернуться к интегралу, то получим:

$$k_f(\tau) = \int_{R_1^n} S(\omega) \cos(\omega \tau) d\omega \tag{18}$$

Что является известным каноническим представлением корреляционной функции, где $S(\omega)$ – спектральная плотность.

Таким образом мы показали, что представление белого шума из интегрального канонического разложения стационарной случайной функции через $V(\omega)$ (описываемого (11)) и спектральной плотности $S(\omega)$ возможно через выражение (9), а саму случайную функцию будет корректно представить через (10) в комплексной форме или через (13) в действительной.

Теперь рассмотрим, как можно выразить наиболее вероятную реализацию случайной функции (10) (или ее же в представлении (13) или (14)), которая лучше всего соответствует обучающей выборке (исходная задача).

Поскольку мера μ является гауссовой и порождена независимыми нормальными коэффициентами v_j^R и v_j^I в каноническом разложении (15), естественно рассмотреть её конечномерные проекции. Для любого конечного набора из N коэффициентов v_j^R и v_j^I соответствующая плотность вероятности пропорциональна:

$$\exp \left(-\frac{1}{2} \sum_{j=0}^N ((v_j^R)^2 + (v_j^I)^2) \right) \tag{19}$$

Хотя в бесконечном пределе $N \rightarrow \infty$ эта плотность как функция от всех коэффициентов формально не существует, выражение (20) сохраняет смысл как функционал, определяющий относительный “вес” реализаций в рамках МАР-оценки.

$$\exp\left(-\frac{1}{2}\sum_{j=0}^{\infty}\left((v_j^R)^2 + (v_j^I)^2\right)\right) \quad (20)$$

Именно максимизация (20) позволяет (с учетом данных) найти наиболее вероятную реализацию. Этот подход стандартен при работе с гауссовыми мерами в функциональных пространствах (Rasmussen & Williams [14]).

Также, из п.1 и п.2 об инвариантности меры μ относительно преобразований (3) и (4) следует, что для любых пар функций, для которых выполняется (3) или (4) предел отношения таких их функционалов между собой (при переходе от (19) к (20) или далее к интегралу) будет равен единице.

Но (19) – (20) учитывают только априорную плотность вероятности без учета u_i . Допустим, что все случайные величины u_i в (1) имеют некоторую дисперсию σ^2 . Умножим (20) на совместный нормальный закон распределения u_i , тогда получившаяся экспоненциальная часть будет:

$$\exp\left(-\frac{1}{2}\sum_{j=0}^{\infty}\left((v_j^R)^2 + (v_j^I)^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^k u_i^2\right) \quad (21)$$

Тогда (аналогично рассуждениям с решением задачи с использованием выражения (2)) поиск наилучшей реализации в (14) или (15) будет поиском таких последовательностей v_j^R и v_j^I которые максимизируют значение (21).

Но поиск максимума (21) равносильно поиску минимума:

$$\frac{1}{2}\sum_{j=0}^{\infty}\left((v_j^R)^2 + (v_j^I)^2\right) + \frac{1}{2\sigma^2}\sum_{i=1}^k u_i^2 \rightarrow \min \quad (22)$$

Обучающая выборка на основании (1) и (15) породит систему уравнений:

$$\begin{cases} \sum_{j=0}^{\infty} v_j^R \sqrt{s_j d\omega} \cos(\omega_j x_1) - \sum_{j=0}^{\infty} v_j^I \sqrt{s_j d\omega} \sin(\omega_j x_1) + u_1 = y_1 \\ \sum_{j=0}^{\infty} v_j^R \sqrt{s_j d\omega} \cos(\omega_j x_2) - \sum_{j=0}^{\infty} v_j^I \sqrt{s_j d\omega} \sin(\omega_j x_2) + u_2 = y_2 \\ \dots \\ \sum_{j=0}^{\infty} v_j^R \sqrt{s_j d\omega} \cos(\omega_j x_k) - \sum_{j=0}^{\infty} v_j^I \sqrt{s_j d\omega} \sin(\omega_j x_k) + u_k = y_k \end{cases} \quad (23)$$

Получается задача минимизации (22) при системе ограничений в виде равенств (23). Это вариационная задача в гильбертовом пространстве, которая

имеет единственное решение (при $\sigma > 0$). Решить ее можно методом множителей Лагранжа.

Функция Лагранжа получится следующая:

$$L(v_1^R, v_2^R, \dots, v_1^I, v_2^I, \dots, u_1, \dots, u_k, \lambda_1, \dots, \lambda_k) = \frac{1}{2} \sum_{j=0}^{\infty} ((v_j^R)^2 + (v_j^I)^2) + \frac{1}{2\sigma^2} \sum_{i=1}^k u_i^2 + \sum_{i=1}^k \lambda_i \left(y_i - u_i - \sum_{j=0}^{\infty} v_j^R \sqrt{s_j d\omega} \cos(\omega_j x_i) + \sum_{j=0}^{\infty} v_j^I \sqrt{s_j d\omega} \sin(\omega_j x_i) \right), \quad (24)$$

где λ_i – множители Лагранжа, количество которых равно размеру обучающей выборки.

Из условий $\frac{dL}{d\lambda_i} = 0$ получим снова систему уравнений (23).

Из условий $\frac{dL}{dv_j^R} = 0$ получим:

$$v_j^R = \sqrt{s_j d\omega} \sum_{i=1}^k \lambda_i \cos(\omega_j x_i) \quad (25)$$

Из условий $\frac{dL}{dv_j^I} = 0$ получим:

$$v_j^I = -\sqrt{s_j d\omega} \sum_{i=1}^k \lambda_i \sin(\omega_j x_i) \quad (26)$$

Из условий $\frac{dL}{du_i} = 0$ получим:

$$u_i = \sigma^2 \lambda_i \quad (27)$$

Т.е. разница, которая обозначена через случайные величины u_i , между наилучшим вариантом $f(x)$ и обучающей выборкой в (1) должна быть пропорциональна множителям Лагранжа (при решении задачи (22) и (23)). И связаны они будут через коэффициент σ^2 , представляющий собой дисперсию величин u_i .

Выразим наиболее вероятную реализацию $f(x)$, подставив (25) и (26) в (15):

$$f(x) = \sum_{j=0}^{\infty} s_j \sum_{i=1}^k \lambda_i \cos(\omega_j x_i) \cos(\omega_j x) d\omega + \sum_{j=0}^{\infty} s_j \sum_{i=1}^k \lambda_i \sin(\omega_j x_i) \sin(\omega_j x) d\omega =$$

$$= \sum_{i=1}^k \lambda_i \sum_{j=0}^{\infty} s_j \cos(\omega_j(x_i - x)) d\omega \quad (28)$$

В (28) получилось не что иное как сумма канонических разложений корреляционной функции (которую уже получили в (16)) с множителями λ_i .

В итоге:

$$f(x) = \sum_{i=1}^k \lambda_i k_f(x_i - x) \quad (29)$$

Таким образом получаем, что если допустить выполнение п.1 и п.3, то решением (1) должна быть линейная комбинация корреляционных функций (29).

Чтобы найти коэффициенты Лагранжа в (29) подставим (25), (26) и (27) в систему (23).

$$\begin{cases} \sum_{i=1}^k \lambda_i k_f(x_i - x_1) + \sigma^2 \lambda_1 = y_1 \\ \sum_{i=1}^k \lambda_i k_f(x_i - x_2) + \sigma^2 \lambda_2 = y_2 \\ \dots \\ \sum_{i=1}^k \lambda_i k_f(x_i - x_k) + \sigma^2 \lambda_k = y_k \end{cases} \quad (30)$$

Запишем (30) в матричной форме. В этом случае σ^2 прибавляется к главной диагонали:

$$(K + \sigma^2 E) \lambda = Y, \quad (31)$$

где K – квадратная матрица элементов $k_{ij} = k_f(x_i - x_j)$

E – единичная матрица

λ – вектор столбец $(\lambda_1, \lambda_2, \dots, \lambda_k)$

Y – вектор столбец $(y_1, y_2, \dots, y_k)$

Соответственно λ из (31) выражается как:

$$\lambda = (K + \sigma^2 E)^{-1} Y \quad (32)$$

Если в (32) взять $\sigma^2 = 0$, то очевидно, что задача аппроксимации (1) превращается в задачу интерполяции. Регулируя значение σ^2 можно определять степень точности соответствия $f(x)$ и обучающей выборки в (1), чтобы избежать проблему “переобучения”.

В итоге, через выражения (29) и (32) было получено решение задачи (1), однако в преобразованиях была использована спектральная плотность $S(\omega)$ (через которую выражается корреляционная функция в (18)), которую пока никак не определили.

Обозначим как $S_{f_1}(\omega)$ спектральное представление функции $f_1(x)$, которая является некоторой конкретной реализацией случайной функции (10), в виде:

$$S_{f_1}(\omega) = \frac{1}{2} V_1(\omega) \sqrt{\frac{S(\omega)}{d\omega}}, \quad (33)$$

где $V_1(\omega)$ – некоторая конкретная реализация $V(\omega)$, т.е. совокупность значений, которые $V(\omega)$ приняла, в результате чего (10) обратилась в одну из своих реализаций $f_1(x)$.

$$f_1(x) = \int_{R^n} S_{f_1}(\omega) e^{i\omega x} d\omega \quad (34)$$

Аналогично обозначим как $S_{f_2}(\omega)$ спектральное представление некоторой другой функции $f_2(x)$.

Предположим, что $f_1(x)$ и $f_2(x)$ такие, что для них выполняется (4).

$$\begin{aligned} f_2(x) &= \int_{R^n} S_{f_2}(\omega) e^{i\omega x} d\omega = k f_1\left(\frac{x}{k}\right) = k \int_{R^n} S_{f_1}(\omega) e^{i\omega \frac{x}{k}} d\omega = \\ &= \int_{R^n} k^{n+1} S_{f_1}(\omega) e^{i\omega \frac{x}{k}} d\frac{\omega}{k} = \int_{R^n} k^{n+1} S_{f_1}(\omega k) e^{i\omega x} d\omega \end{aligned} \quad (35)$$

Сравнив начало и конец в (35) получим соотношение между спектральными представлениями для $f_1(x)$ и $f_2(x)$ для которых выполняется (4):

$$S_{f_2}(\omega) = k^{n+1} S_{f_1}(\omega k) \quad (36)$$

Вернемся теперь к (33), но запишем его иначе, выразив $V_1(\omega)$:

$$V_1(\omega) = 2S_{f_1}(\omega) \sqrt{\frac{d\omega}{S(\omega)}} \quad (37)$$

Поскольку для $f_1(x)$ и $f_2(x)$ выполняется (4), то как уже было сказано выше, предел отношения их функционалов (20) должен быть равен единице.

Но выражение под суммой в (20) мы можем заменить как:

$$(v_j^R)^2 + (v_j^I)^2 = |V(\omega_j)|^2 \quad (38)$$

Квадрат модуля (37) для частоты ω_j будет:

$$|V_1(\omega_j)|^2 = 4 \frac{|S_{f_1}(\omega_j)|^2}{S(\omega_j)} d\omega \quad (39)$$

Подставив (39) в (20), получим:

$$\exp\left(-2 \sum_{j=0}^{\infty} \frac{|S_{f_1}(\omega_j)|^2}{S(\omega_j)} d\omega\right) \quad (40)$$

Но эта сумма в (40) в пределе даст интеграл:

$$\exp\left(-2 \int_{R_1^n} \frac{|S_{f_1}(\omega)|^2}{S(\omega)} d\omega\right) = \exp\left(- \int_{R^n} \frac{|S_{f_1}(\omega)|^2}{S(\omega)} d\omega\right) \quad (41)$$

В итоге, для $f_1(x)$ и $f_2(x)$, факт их одинакового “веса” как реализаций, выражающийся в единичном пределе соотношения их функционалов, которые теперь представили в виде (41), мы можем выразить приравняв для них интегральные части:

$$\int_{R^n} \frac{|S_{f_1}(\omega)|^2}{S(\omega)} d\omega = \int_{R^n} \frac{|S_{f_2}(\omega)|^2}{S(\omega)} d\omega \quad (42)$$

Подставим в (42) найденное ранее соотношение между спектральными представлениями (36):

$$\begin{aligned} \int_{R^n} \frac{|S_{f_1}(\omega)|^2}{S(\omega)} d\omega &= \int_{R^n} \frac{|S_{f_2}(\omega)|^2}{S(\omega)} d\omega = \int_{R^n} \frac{k^{2n+2} |S_{f_1}(\omega k)|^2}{S(\omega)} d\omega = \\ &= \int_{R^n} \frac{k^{n+2} |S_{f_1}(\omega k)|^2}{S(\omega)} d(\omega k) = \int_{R^n} \frac{k^{n+2} |S_{f_1}(\omega)|^2}{S\left(\frac{\omega}{k}\right)} d\omega \end{aligned} \quad (43)$$

Сравнив начало и конец в (43) получим соотношение для спектральной плотности:

$$\frac{S(\omega/k)}{S(\omega)} = k^{n+2} \quad (44)$$

Поскольку рассматриваемая случайная функция стационарна, то автокорреляционная функция (18) должна обладать радиальной симметрией. Тогда радиальной симметрией должна обладать и ее спектральная плотность.

Такой функцией, удовлетворяющей (44) и радиальной симметрии будет:

$$S(\omega) = a \|\omega\|^{-(n+2)}, \quad (45)$$

где a – некоторый коэффициент

Так как $\omega \in R^n$ и является многомерной величиной $(\omega_1, \omega_2, \dots, \omega_n)$, то (45) можно записать и так:

$$S(\omega_1, \omega_2, \dots, \omega_n) = a(\omega_1^2 + \omega_2^2 + \dots + \omega_n^2)^{-\frac{n+2}{2}}, \omega_1, \omega_2, \dots, \omega_n \in R \quad (46)$$

Как видно, выражение (44) будет справедливо при любом коэффициенте a в (45) или (46), который можно выбрать исходя из представления, что $k_f(0)$ будет являться дисперсией случайной функции. Соотношение $k_f(0)$ и σ^2 будет определять, насколько точно $f(x)$ должна соответствовать обучающей выборке. Как видно по (32) и (29), если одновременно умножить k_f и σ^2 на некоторый множитель, то функция (29) останется неизменной.

Рассмотрим более подробно результат, получившийся в (45) и (46).

Обозначим τ в $k_f(\tau)$ в (18) как вектор $(\tau_1, \tau_2, \dots, \tau_n)$ и используя спектральную плотность (46) запишем (заодно перейдя из области интегрирования R_1^n в R^n):

$$k_f(\tau_1, \tau_2, \dots, \tau_n) = \frac{a}{2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\cos(\omega_1 \tau_1 + \omega_2 \tau_2 + \dots + \omega_n \tau_n)}{(\omega_1^2 + \omega_2^2 + \dots + \omega_n^2)^{\frac{n+2}{2}}} d\omega_n d\omega_{n-1} \dots d\omega_1 \quad (47)$$

Чем будет являться сечение корреляционной функции (47), например, при условии $\tau_n=0$? Рассмотрим самый внутренний интеграл в (47) при $\tau_n=0$.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\cos(\omega_1 \tau_1 + \omega_2 \tau_2 + \dots + \omega_{n-1} \tau_{n-1})}{(\omega_1^2 + \omega_2^2 + \dots + \omega_n^2)^{\frac{n+2}{2}}} d\omega_n = \\ = 2\cos(\omega_1 \tau_1 + \omega_2 \tau_2 + \dots + \omega_{n-1} \tau_{n-1}) \int_0^{\infty} \frac{d\omega_n}{(\omega_1^2 + \omega_2^2 + \dots + \omega_n^2)^{\frac{n+2}{2}}} \end{aligned} \quad (48)$$

Сделаем замену:

$$\frac{\omega_1^2 + \omega_2^2 + \dots + \omega_{n-1}^2}{\omega_1^2 + \omega_2^2 + \dots + \omega_{n-1}^2 + \omega_n^2} = 1 - t \quad (49)$$

Если выполнить преобразования, тогда (48) преобразуется в:

$$\begin{aligned} & \frac{\cos(\omega_1 \tau_1 + \omega_2 \tau_2 + \dots + \omega_{n-1} \tau_{n-1})}{(\omega_1^2 + \omega_2^2 + \dots + \omega_{n-1}^2)^{\frac{(n-1)+2}{2}}} \int_0^1 t^{-\frac{1}{2}} (1-t)^{\frac{n-1}{2}} dt = \\ & = \frac{\cos(\omega_1 \tau_1 + \omega_2 \tau_2 + \dots + \omega_{n-1} \tau_{n-1})}{(\omega_1^2 + \omega_2^2 + \dots + \omega_{n-1}^2)^{\frac{(n-1)+2}{2}}} B\left(\frac{1}{2}, \frac{n+1}{2}\right), \end{aligned} \quad (50)$$

где B – бета-функция Эйлера

Тогда сечение (47) при $\tau_n=0$ будет:

$$\begin{aligned} k_f(\tau_1, \dots, \tau_{n-1}) &= \\ &= \frac{a}{2} B\left(\frac{1}{2}, \frac{n+1}{2}\right) \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\cos(\omega_1 \tau_1 + \dots + \omega_n \tau_{n-1})}{(\omega_1^2 + \omega_2^2 + \dots + \omega_{n-1}^2)^{\frac{(n-1)+2}{2}}} d\omega_{n-1} \dots d\omega_1 \end{aligned} \quad (51)$$

Поскольку $\frac{a}{2} B\left(\frac{1}{2}, \frac{n+1}{2}\right)$ играет роль некоторого коэффициента, сравнивая (51) и (47) можно сделать вывод, что если для корреляционной функции в пространстве R^n провести через начало координат сечение размерностью R^{n-1} , то оно будет эквивалентно корреляционной функции в пространстве R^{n-1} выраженное снова через (47) (отличаясь лишь коэффициентом a в (45)-(46)).

Отсюда можно сделать вывод, чтобы найти корреляционную функцию, достаточно будет определить её для одномерного случая на промежутке $\tau \in [0, +\infty)$, а затем пользуясь свойством ее радиальной симметрии, определить ее для пространства любой размерности.

Спектральная плотность (45) (если взять $a = 1$) для одномерного случая будет:

$$S(\omega) = |\omega|^{-3}, \omega \in R \quad (52)$$

Случайная функция, выраженная в виде (13) в одномерном случае примет вид:

$$F(x) = \int_0^{\infty} \frac{V_R(\omega) \cos(\omega x) - V_I(\omega) \sin(\omega x)}{\omega \sqrt{\omega} d\omega} d\omega, \quad x, \omega \in R \quad (53)$$

Корреляционная функция (18) в одномерном случае:

$$k_f(\tau) = \int_0^{\infty} \frac{\cos(\omega\tau)}{\omega^3} d\omega \quad (54)$$

Модуль от частоты в (53) и (54) можно убрать, поскольку интегрирование в будет только в положительной области.

Однако в (53) и (54) видно, что при приближении частоты ω к нулю выражение под интегралом стремится к бесконечности.

Данная сингулярность в спектральной плотности при $\omega \rightarrow 0$ (и как следствие расходимость при интегрировании (54)) указывает на то, что рассматриваемая случайная функция принадлежит к классу собственных случайных функций (Intrinsic Random Functions, IRF) порядка 1 (Matheron G. [12]). Это означает, что дисперсия случайной функции бесконечна, но ее первые производные (или, эквивалентно, приращения вдоль произвольных направлений) образуют стационарное случайное поле. Функция $k_f(\tau)$, полученная далее в (64) после регуляризации, соответствует обобщенной ковариации (Generalized Covariance) для данного класса IRF порядка 1. Введение малого параметра ω_0 является одним из способов регуляризации, эквивалентным ограничению пространства функций условиями на поведение в низкочастотной области, и позволяет перейти к хорошо определенной ковариационной функции для практических вычислений.

Возьмем некоторое очень малое значение $\omega_0 > 0$ и рассмотрим случайную функцию, аналогичную (53), но у которой (у всех ее реализаций) частоты меньше ω_0 отсутствуют.

Ее выражение будет отличаться от (53) только нижним пределом интегрирования:

$$F(x) = \int_{\omega_0}^{\infty} \frac{V_R(\omega) \cos(\omega x) - V_I(\omega) \sin(\omega x)}{\omega \sqrt{\omega} d\omega} d\omega \quad (55)$$

Корреляционная функция для нее будет:

$$k_f(\tau) = \int_{\omega_0}^{\infty} \frac{\cos(\omega\tau)}{\omega^3} d\omega \quad (56)$$

Гармоническому колебанию с частотой ω_0 будет соответствовать период:

$$T_0 = \frac{2\pi}{\omega_0} \quad (57)$$

Пусть ω_0 взято такое, что период T_0 многократно больше, чем диапазон значений x из обучающей выборки. Тогда все колебания частот, меньших чем ω_0 , внутри этой области будут вырождаться в линейные слагаемые и мало отличаться друг от друга и от частот несколько больших, чем ω_0 .

Поэтому можно ожидать, что если значения обучающей выборки x_i ограничены некоторым конечным диапазоном, то всегда можно взять такое малое ω_0 , что (55) и (56) будет допустимой с практической точки зрения заменой (53) и (54), а при устремлении ω_0 к нулю они становятся эквивалентными.

Выполним дважды интегрирование (56) по частям:

$$\begin{aligned} k_f(\tau) &= \int_{\omega_0}^{\infty} \frac{\cos(\omega\tau)}{\omega^3} d\omega = -\frac{\cos(\omega\tau)}{2\omega^2} \Big|_{\omega_0}^{\infty} - \tau \int_{\omega_0}^{\infty} \frac{\sin(\omega\tau)}{2\omega^2} d\omega = \\ &= \left(-\frac{\cos(\omega\tau)}{2\omega^2} + \tau \frac{\sin(\omega\tau)}{2\omega} \right) \Big|_{\omega_0}^{\infty} - \frac{\tau^2}{2} \int_{\omega_0}^{\infty} \frac{\cos(\omega\tau)}{\omega} d\omega \end{aligned} \quad (58)$$

Одним из множителей в последнем слагаемом в (58) получился интегральный косинус.

Рассмотрим его более подробно (принимая во внимание, что рассматриваем лишь $\tau \geq 0$).

$$\begin{aligned} - \int_{\omega_0}^{\infty} \frac{\cos(\omega\tau)}{\omega} d\omega &= - \int_{\omega_0\tau}^{\infty} \frac{\cos(\omega\tau)}{\omega\tau} d(\omega\tau) = \\ &= \gamma + \ln(\omega_0\tau) + \int_0^{\omega_0\tau} \frac{\cos(\omega) - 1}{\omega} d\omega, \end{aligned} \quad (59)$$

где γ – постоянная Эйлера-Маскерони

Тогда (58) преобразуется в:

$$\begin{aligned} &\frac{\cos(\omega_0\tau)}{2\omega_0^2} - \tau^2 \frac{\sin(\omega_0\tau)}{2\omega_0\tau} + \frac{1}{2}\tau^2 \left(\gamma + \ln(\tau) + \ln(\omega_0) + \int_0^{\omega_0\tau} \frac{\cos(\omega) - 1}{\omega} d\omega \right) = \\ &= \frac{1}{2}\tau^2 \left(\ln(\tau) + \left(\ln(\omega_0) + \gamma + \int_0^{\omega_0\tau} \frac{\cos(\omega) - 1}{\omega} d\omega - \frac{\sin(\omega_0\tau)}{2\omega_0\tau} \right) \right) + \end{aligned}$$

$$+ \frac{\cos(\omega_0 \tau)}{2\omega_0^2} \quad (60)$$

Поскольку, как уже рассмотрели выше, $k_f(\tau)$ мы можем умножить на произвольный коэффициент, а важно лишь соотношение $k_f(0)$ и σ^2 , то в (60) мы можем убрать $\frac{1}{2}$, а (60) записать как:

$$k_f(\tau) = \tau^2(\ln(\tau) - b(\tau)) + c(\tau), \quad (61)$$

где

$$b(\tau) = \frac{\sin(\omega_0 \tau)}{\omega_0 \tau} - \ln(\omega_0) - \gamma - \int_0^{\omega_0 \tau} \frac{\cos(\omega) - 1}{\omega} d\omega \quad (62)$$

$$c(\tau) = \frac{\cos(\omega_0 \tau)}{\omega_0^2} \quad (63)$$

Если устремить ω_0 к нулю, то $b(\tau)$ и $c(\tau)$ устремятся к бесконечности.

$$\lim_{\omega_0 \rightarrow +0} b(\tau) \rightarrow +\infty,$$

$$\lim_{\omega_0 \rightarrow +0} c(\tau) \rightarrow +\infty$$

А случайная функция, описываемая (55) и (56) будет превращаться в случайную функцию, описываемую (53) и (54). Хотя видно, что $c(\tau)$ будет устремляться к бесконечности гораздо быстрее, чем $b(\tau)$, значение которого будет в основном определяться логарифмом $(-\ln(\omega_0))$.

Таким образом мы получили регуляризованную ковариационную функцию (61) для одномерного случая и $\tau \in [0, +\infty)$. Ее легко обобщить на многомерный случай, поскольку она должна обладать радиальной симметрией.

Но если взять ω_0 достаточно малым, чтобы период (57) был значительно больше, чем диапазон изменения τ , то $b(\tau)$ и $c(\tau)$ фактически превращаются в константы. Если сделать такую замену в (61), мы получим обобщенную ковариацию, которую удобно использовать для вычислений.

Получаем итоговый результат:

Обобщенная ковариация:

$$k_f(\tau) = \|\tau\|^2(\ln(\|\tau\|) - b) + c, \tau \in R^n \quad (64)$$

где b и c – константы, которые можно оценить по формулам (62) и (63)

Полученная таким образом функция (64) является обобщенной ковариацией (Generalized Covariance) для рассматриваемого класса собственных случайных функций порядка 1 [12]. В дальнейшем изложении, мы будем использовать этот термин. Константы b и c обеспечивают устойчивость решения, поглощая полиномиальные компоненты.

Выпишем еще раз полученные формулы (29) и (32).

Решением задачи аппроксимации (1) будет функция:

$$f(x) = \sum_{i=1}^k \lambda_i k_f(x_i - x) \quad (65)$$

Коэффициенты λ_i определяются системой уравнений:

$$\lambda = (K + \sigma^2 E)^{-1} Y \quad (66)$$

где K – квадратная матрица элементов $k_{ij} = k_f(x_i - x_j)$

E – единичная матрица

λ – вектор столбец $(\lambda_1, \lambda_2, \dots, \lambda_k)$

Y – вектор столбец $(y_1, y_2, \dots, y_k)$

σ^2 – дисперсия случайных величин u_i из (1)

Полученная линейная комбинация (65) функций (64) известна также как полигармонический сплайн (или его разновидность, сплайн тонкой пластины (thin plate spline) R.L. Harder, R.N. Desmarais, Buhmann M.D. [10] и [4]). Но в данном случае есть нюансы использования, такие как использование коэффициентов b и c , которые оцениваются через (62) и (63), прямое введение σ^2 случайных величин u_i в систему уравнений (66), чтобы решить задачу аппроксимации.

Дополнительно стоит сказать, что полученная функция (64) соответствует обобщенному полигармоническому сплайну (polyharmonic spline) в многомерном пространстве R^n (Buhmann, M.D., Wendland H. [5] и [18]). В частности, слагаемое $\tau^2 \ln \|\tau\|$ является каноническим ядром для сплайна минимального штрафа, инвариантного относительно сдвига, вращения и масштабирования в R^n . Важно подчеркнуть, что данный вид ядра выведен здесь не из эвристических соображений или аналогии с 2D случаем, а является прямым следствием постулированных симметрий п.1 – п.3 вероятностной меры в бесконечномерном функциональном пространстве и справедлив для произвольной размерности n входного пространства.

Также можно сказать, полученная обобщённая ковариация (64) тесно связана с обобщёнными сплайнами Дюшона (Duchon splines) (Duchon, 1977 [8]) для порядка гладкости $m = 2$. Её ведущее слагаемое $\tau^2 \ln \|\tau\|$ в точности соответствует каноническому ядру сплайна Дюшона в размерности $n=2$. Для других размерностей структура ядра (64), выведенная из принципов симметрии и регуляризации, обеспечивает устойчивость решения и включает компоненты $-b\|\tau\|^2 + c$, соответствующие полиномиальному ядру оператора, что является стандартным подходом для работы с собственными случайными функциями (Matheron, 1973 [12]). Таким образом, данный результат даёт вероятностную интерпретацию и обоснование для класса методов сглаживания на основе сплайнов Дюшона.

Что касается полученной системы уравнений (66) или (31), то она математически идентична задаче Гауссовского процессного регрессионного анализа “Gaussian Process Regression, GPR” (Rasmussen & Williams, 2006 [14]) с ковариационной функцией k_f и параметром дисперсии шума σ^2 . В терминах более широкого контекста машинного обучения, данный метод также эквивалентен Ridge-регрессии с ядром (Kernel Ridge Regression, KRR) (Saunders et al., 1998; Bishop, 2006 [15] и [2]) в Reproducing Kernel Hilbert Space (RKHS), порожденном ядром k_f , где параметр σ^2 играет роль коэффициента L2-регуляризации.

Ключевой результат и новизна данной работы заключаются не в форме решающих уравнений, а в строгом выводе явного вида ядра (64) из первых принципов — постулатов о симметрии вероятностной меры п.1 – п.3 в бесконечномерном функциональном пространстве

Коснемся теперь свойств матрицы K в (66). Она построена на основе обобщенной ковариации (64) и может быть вырожденной из-за наличия в ее ядре полиномиальных компонент (констант и линейных членов) (Chilès & Delfiner, 2012, гл. 4 [6]). На практике это проявляется, если расстояния между точками выборки удовлетворяют определенным соотношениям (например, для двух точек при $\|x_1 - x_2\| = e^b$). Однако, в силу следующих ниже факторов проблема обычно не возникает.

Параметр $\sigma^2 > 0$ в (66) гарантирует обратимость матрицы $K + \sigma^2 E$.

Типичные значения b (определяемые через ω_0 в (62)) делают критическое расстояние e^b экстремально большим (например, при $\omega_0 = 0.001$, $b \approx 7.33$, $e^b \approx 1500$), что значительно превышает характерные расстояния между точками в ограниченной области аппроксимации.

Для задач точной интерполяции ($\sigma^2 = 0$) даже при самом неудачном расположении точек можно использовать функцию (61), с вычислением b и c

через (62) и (63). А также можно использовать стандартную для полигармонических сплайнов процедуру с явным выделением полиномиальной части и условиями ортогональности (Wahba G. 1990 [17]).

Пример.

В качестве наглядного примера рассмотрим одномерный случай аппроксимации.

Возьмем обучающую выборку, у которой большая часть входных значений лежит в интервале от 0 до 10, т.е. для оценки значений b и c примем, что τ изменяется от 0 до 10.

Возьмем $\omega_0 = 0.001$. Таким образом соответствующий этой частоте период (57) будет равен $T_0 = 6283.1853$, что примерно в 600 раз больше, чем изменения между значениями x_i . Вполне можно допустить, что наличие частот меньших ω_0 (с еще большим периодом) для решения задачи в нашей области аппроксимации не имеет существенного значения.

Сравним значения $b(\tau)$ и $c(\tau)$ из (62) и (63) при выбранной $\omega_0 = 0.001$ и значениях $\tau = 0$ и $\tau = 10$.

$$b(0) = 1 + 6.907755 - 0.577216 + 0 = 7.33054$$

$$b(10) = 0.999983 + 6.907755 - 0.577216 + 0.000025 = 7.330548$$

$$c(0) = 1000000$$

$$c(10) = 999950.000416$$

Значения $b(\tau)$ и $c(\tau)$ при $\tau = 10$ относительно их же при $\tau = 0$ изменились на 0.0001% и 0.005% соответственно. Поэтому возьмем в качестве них просто константы $b = 7.33054$ и $c = 1000000$.

Попробуем сначала взять $\sigma^2 = 0$ в (66). В этом случае все u_i в (1) становятся равными нулю, $f(x)$ должна точно без ошибок пройти через все точки из обучающей выборки. Задача аппроксимации превращается в задачу интерполяции. Что и наблюдается на рисунке 1.

Замечание. При практическом вычислении элементов матрицы K по формуле (64) необходимо учитывать, что выражение $\|\tau\|^2 \ln(\|\tau\|)$ не определено при $\|\tau\| = 0$. Однако, поскольку $\lim_{\|\tau\| \rightarrow 0} \|\tau\|^2 \ln(\|\tau\|) = 0$, для диагональных элементов матрицы K (где $x_i = x_j$ и $\|\tau\| = 0$) следует явно задавать значение:

$$k_f(0) = c$$

Для недиагональных элементов с очень малыми расстояниями $\|\tau\|$ (например, $\ll 10^{-10}$) также рекомендуется использовать аппроксимацию $k_f(\tau) \approx c$ во избежание численных нестабильностей.

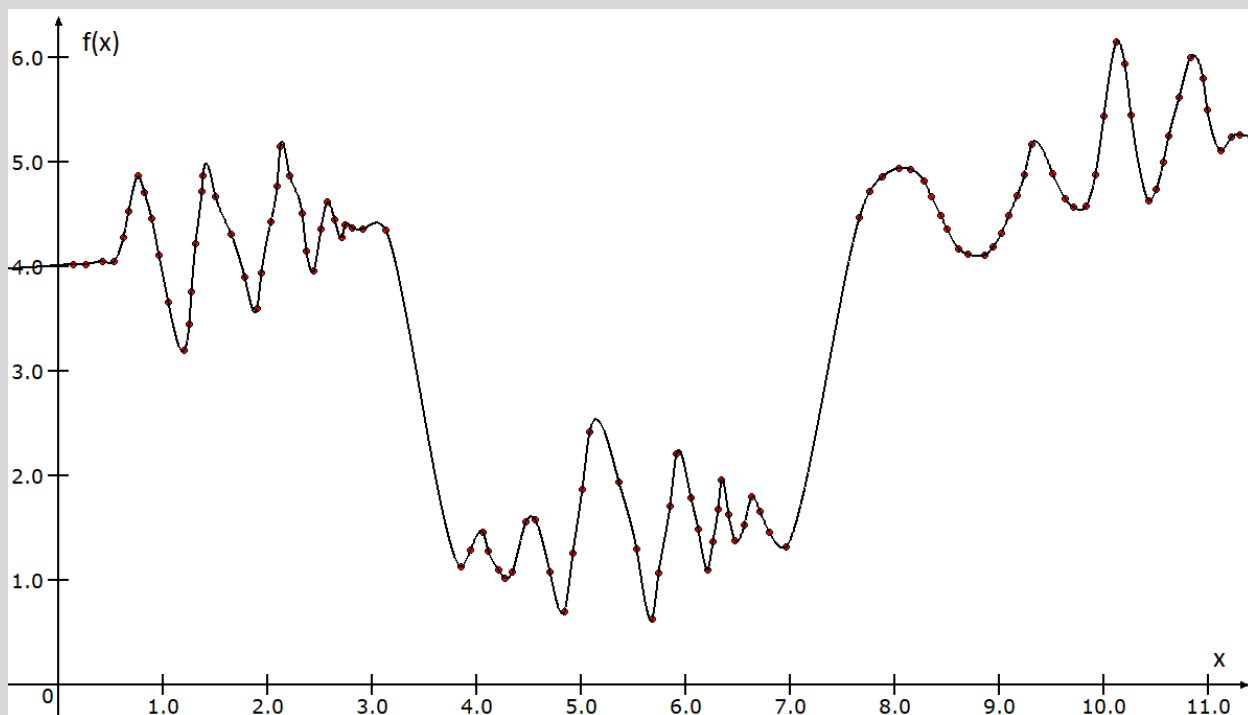


Рисунок 1.

Функция (65) легко воспроизводит любую сложную нелинейность без каких-либо скачков или осцилляций между соседними точками.

Если взять $\sigma^2 > 0$, то переходим к исходной задаче аппроксимации (1).

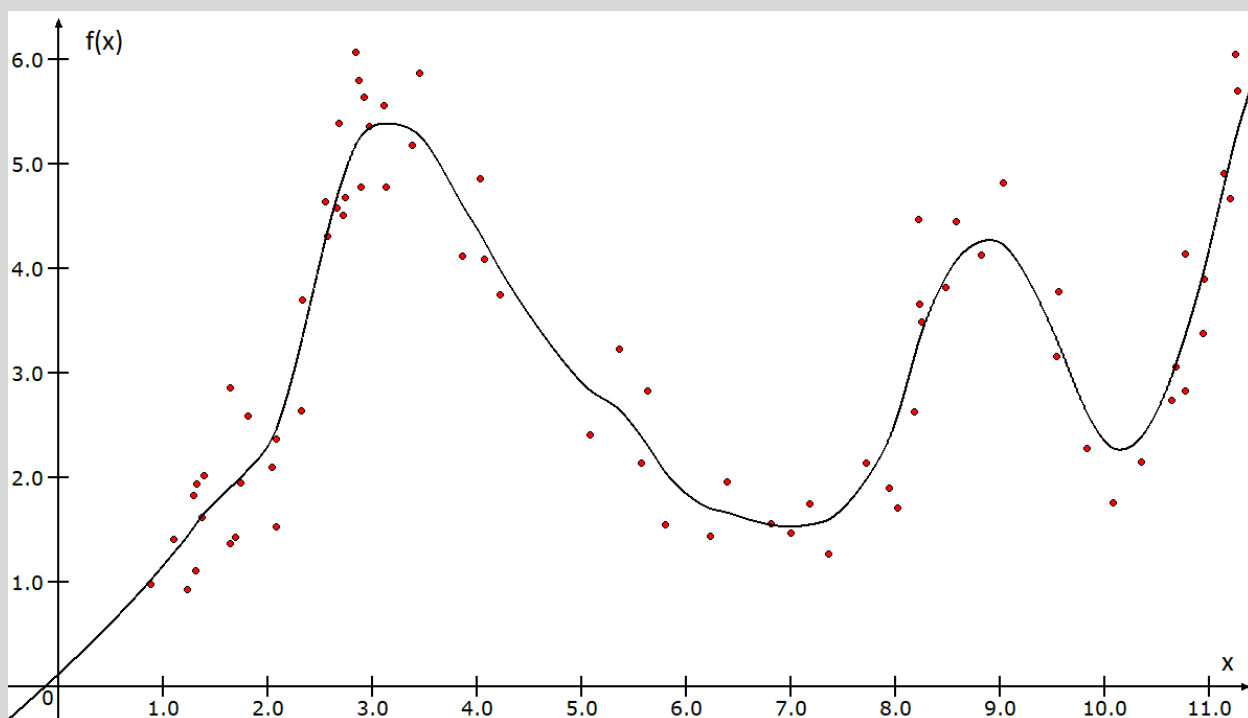


Рисунок 2.

В данном случае, изображенном на рисунке 2, $\sigma^2 = 0.625$.

Обсуждение.

Полученное решение (65), (66) с ядром (64) обладает рядом фундаментальных преимуществ, вытекающих из его спектральных свойств и теоретического вывода:

1. Устойчивость и отсутствие осцилляций. В отличие от многих других ядер (например, гауссова), спектральная плотность (45), соответствующая выведенному ядру, имеет степенной характер $S(\omega) = a\|\omega\|^{-(n+2)}$. Такой спектр не имеет "провалов" или экспоненциально быстрого затухания на высоких частотах. Это означает, что базис, порожденный данным ядром, является "полным" в широком классе функций и не склонен к порождению осцилляций между узлами интерполяции, что является частой проблемой при использовании, например, гауссовых радиальных базисных функций, которые могут приводить к переобучению и неустойчивости (явление Гиббса). Решение всегда дает плавные и предсказуемые поверхности.
2. Естественная регуляризация через параметр σ^2 . Параметр σ^2 обеспечивает плавный переход от точной интерполяции ($\sigma^2 = 0$) к аппроксимации ($\sigma^2 > 0$), эффективно борясь с переобучением. Однако, в отличие от многих других методов, даже в режиме интерполяции решение остается устойчивым благодаря свойствам ядра.
3. Обоснованность выбора ядра. Обычно выбор ядра (например, гауссова, полиномиального) в методах вроде Kernel Ridge Regression или Gaussian Process Regression является эмпирическим и часто сводится к перебору. В данной работе ядро не выбирается, а выводится из фундаментальных принципов индифферентности и симметрии. Это означает, что для задач, где априорная информация отсутствует, полигармонический сплайн (64) является не просто одним из вариантов, а естественным априорным выбором. Попытка использовать ядро с иным спектром (например, быстро затухающим, как у гауссова) будет эквивалентна необоснованному априорному предположению о том, что высокочастотные компоненты в искомой функции $f(x)$ маловероятны, что может привести к потере деталей или, наоборот, к осцилляциям при попытке эти детали восстановить.

Таким образом, представленный метод не только предоставляет конкретные формулы для решения, но и дает теоретическую гарантию их

устойчивости и адекватности для широкого класса задач аппроксимации в условиях полного отсутствия априорной информации.

Заключение.

В статье было показано, что регрессионная задача машинного обучения как задача многомерной аппроксимации, выраженная (1), может быть решена на основе теории случайных функций, если предположить существование в функциональном пространстве вероятностной меры, обладающей симметриями, постулированными в пунктах 1–3. Решением будет линейная комбинация корреляционной функции, спектральная плотность которой (45)–(46) также может быть выведена из одной из симметрий (4). Далее показано, что обобщенной ковариация с такой плотностью соответствует полигармоническому сплайну в виде (64). И как итог, выражения (62)–(66) для решения задачи (1).

В отличие от большинства подходов машинного обучения, где выбор ядра или структуры модели является эмпирическим, в данной работе вся архитектура решения — включая вид базисных функций, форму регуляризации и параметризацию шума — выводится аналитически из минимального набора естественных предположений о симметриях пространства функций. Это позволяет рассматривать полигармонический сплайн не просто как удобный аппроксиматор, а как теоретически обоснованное решение в условиях полного отсутствия априорной информации.

Автор благодарит анонимных рецензентов за ценные отзывы о более ранних версиях этой работы.

Список литературы:

1. Бахвалов Ю.Н. Метод многомерной интерполяции и аппроксимации и его приложения. — М.: Спутник+, 2007. - 108 с
2. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. (Chapter 6)
3. Bogachev, V. I. (1998). Gaussian Measures. American Mathematical Society. Mathematical Surveys and Monographs, Vol. 62. Providence, RI. <https://doi.org/10.1090/surv/062>

4. Bookstein, F. L. (June 1989). "Principal warps: thin plate splines and the decomposition of deformations". IEEE Transactions on Pattern Analysis and Machine Intelligence. 11 (6): 567–585. doi:10.1109/34.24792
5. Buhmann, M.D. (2003). Radial Basis Functions: Theory and Implementations. Chapter 8: Polyharmonic Splines. Cambridge University Press
6. Chilès, J.P., Delfiner, P. (2012). Geostatistics: Modeling Spatial Uncertainty (2nd ed.). Wiley. Chapter 4: Intrinsic Random Functions of Order k. Wiley
7. Cramér, H. (1942). Mathematical Methods of Statistics. Princeton University Press. (See Chapter 10, Section 10.6: "The Spectral Representation of a Stationary Process" or similar sections in later editions).
8. Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In Constructive Theory of Functions of Several Variables (pp. 85-100). Springer, Berlin, Heidelberg. DOI: 10.1007/BFb0086566
9. Гихман И. И., Скороход А. В. Теория случайных процессов. М., 1971. Т. 1. 664 с
10. Harder R.L. and Desmarais R.N.: Interpolation using surface splines. Journal of Aircraft, 1972, Issue 2, pp. 189–191
11. Kallenberg, O. (2002/2021). Foundations of Modern Probability. Springer.
12. Matheron, G. (1973). The Intrinsic Random Functions and Their Applications. Advances in Applied Probability, 5(3), 439–468. [DOI: <https://doi.org/10.2307/1425829>]
13. Пугачев В.С., Теория случайных функций и её применение к задачам автоматического управления. Изд. 2-ое, перераб. и допол. — М.: Физматлит, 1960.
14. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press.
15. Saunders, C., Gammerman, A., & Vovk, V. (1998). Ridge Regression Learning Algorithm in Dual Variables. ICML.
16. Vakhania, N. N., Tarieladze, V. I., Chobanyan, S. A. (1987). Probability Distributions on Banach Spaces. Springer.
17. Wahba, G. (1990). Spline Models for Observational Data. SIAM.
18. Wendland, H. (2005). Scattered Data Approximation. Chapter 8: Least Squares on Splines, Cambridge University Press.
19. [http://www.machinelearning.ru/wiki/index.php?title=Многомерная интерполяция и аппроксимация на основе теории случайных функций](http://www.machinelearning.ru/wiki/index.php?title=Многомерная_интерполяция_и_аппроксимация_на_основе_теории_случайных_функций) (2009)