

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: Аппроксимация временных рядов с помощью В-сплайнов

Студент гр. 3304

Осипов В.Г.

Руководители

Середа В. И.

Кринкин К.В.

Санкт-Петербург

2018

ЗАДАНИЕ
НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Студент Осипов В.Г.

Группа 3304

Тема НИР: Аппроксимация временных рядов с помощью В-сплайнов

Задание на НИР:

Работа над темой НИР.

Сроки выполнения НИР: 13.09.2018 – 27.12.2018

Дата сдачи отчета: 25.12.2018

Дата защиты отчета: 25.12.2018

Студент

Осипов В.Г.

Руководители

Середа В. И.

Кринкин К.В.

АННОТАЦИЯ

В работе были рассмотрены методы аппроксимации на нерегулярной сетке с помощью сплайнов в виде линейной комбинации кубических В-сплайнов. Была разработана соответствующая программа.

SUMMARY

In this paper, approximation methods on an irregular grid using splines in the form of a linear combination of cubic B-splines were considered. There was a corresponding program.

Содержание

Введение	5
1. Обзор предметной области и аналогов	6
1.1 Интерполяция	6
1.2 Полиномиальная аппроксимация	6
1.3 Аппроксимация В-сплайнами	7
1.4 Сравнение	7
2. Аппроксимация	9
2.1 Задача аппроксимации	9
2.2 Решение задачи аппроксимации В-сплайнами	10
2.3 Кубический сплайн	11
3. Программная реализация	13
3.1 Используемые технологии	13
3.2 Тестовые данные	14
3.3 Описание разработанной программы	14
4. Заключение	15
4.1 Выводы	15
Список литературы	16

Введение

Анализ временных рядов включает в себя методы анализа данных с целью получения значимых статистических и других характеристик анализируемых данных.

Для построения модели явления, являющегося источником анализируемого временного ряда, необходимо выявление структуры временного ряда. Сюда входят методы приближения с помощью кривых, интерполяции, аппроксимации и др. Модели временных рядов могут иметь различные формы и представлять различные стохастические процессы.

Анализ результатов любого эксперимента основывается на обработке полученных данных. Во многих случаях эти данные представляют собой временные ряды – расположенные в хронологическом порядке последовательности значений одной или нескольких измеряемых величин [1].

Актуальность работы заключается в том, что интерполяция является частным случаем аппроксимации и может быть использована для обработки данных в форме временного ряда. Аппроксимация используется во многих областях, в частности, в задачах регрессии.

В работе исследуются вопросы аппроксимации функций одной переменной с помощью кубических сплайнов, выраженных в различных формах.

Цель работы состоит в описании методов аппроксимации, в том числе и с помощью В-сплайнов.

Таким образом, имеются следующие задачи:

- 1) формирование списка существующих методов;
- 2) описание методов аппроксимации;
- 3) сравнение рассмотренных методов.

1. Обзор предметной области и аналогов

1.1 Интерполяция

Пусть отрезок $[a, b]$ разбит на m равных неперекрывающихся отрезков $\omega = [x_i, x_{i+1}]$ – сетка интерполяции, где $x_0 = a, x_m = b, x_i = a + ih, i = 0, 1, \dots, m$ – узлы интерполяции, $h = \frac{b-a}{m}$ – шаг интерполяции. Известны значения функции f_i в точках x_i :

$$f_i = f(x_i).$$

Задача приближенного восстановления значений функции f в произвольной точке x внутри отрезка $[a, b]$ называется задачей интерполяции, а приближенное восстановление функции f по формуле $f(x) \approx S(x)$ называется интерполяцией функции [2, 3].

1.2 Полиномиальная аппроксимация

Задана сетка $\omega = [t_i, t_{i+1}], i = 0, 1, \dots, m$ и набор данных

$$y_0, y_2, \dots, y_m.$$

При предположении, что данные представляют собой некоторую лежащую в их основе функцию $b(t)$, для которой $y_i = b(t_i)$ [4], т.е. задана модель данных в виде

$$\text{наблюдение} = \text{модель} + \text{ошибка}.$$

Иными словами, можно записать функцию $b(x)$ следующим образом

$$y_i \approx x_0 \phi_0(t_i) + x_1 \phi_1(t_i) + \dots + x_n \phi_n(t_i),$$

где $\phi_j(t)$ – модельные функции.

Согласно методу наименьших квадратов, для минимизации значений $b - Ax$ необходимо решить задачу

$$\min_x \sum_{j=0}^m [(b - Ax)_j]^2$$

В случаях, когда не все данные одинаково важны или известно, что данные имеют разную точность, можно ввести веса $w_i > 0, \sum_{i=0}^n w_i = 1$ и минимизировать сумму

$$\sum_{j=0}^m w_j [(b - Ax)_j]^2.$$

Если данные аппроксимируются полиномиальной функцией регрессии одной переменной $f(x) = b_0 + \sum_{i=0}^k b_i x^i$, то, матричные уравнения в данном случае примут вид

$$\begin{pmatrix} n & \sum_n x_t & \dots & \sum_n x_t^k \\ \sum_n x_t & \sum_n x_t^2 & \dots & \sum_n x_t^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_n x_t^k & \sum_n x_t^{k+1} & \dots & \sum_n x_t^{2k} \end{pmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_n y_t \\ \sum_n x_t y_t \\ \vdots \\ \sum_n x_t^k y_t \end{bmatrix}.$$

1.3 Аппроксимация В-сплайнами

Сплайн $S(x)$ k -й степени на интервале $[a, b]$ на сетке ω из $m + 1$ узла может быть записан в виде линейной комбинации В-сплайнов:

$$S(x) = \sum_{i=-k}^{m-1} \alpha_i N_i^k(x)$$

где коэффициенты α_i условиями задачи определяются однозначно [5].

Для решения задачи аппроксимации, необходимо подобрать коэффициенты c_i , чтобы минимизировать функцию [4]

$$\delta(c) = \sum_{i=0}^{n-1} \left[w_i \left(y_i - \sum_{j=0}^{g+k} c_j N_{j,k+1}(x_i) \right) \right]^2,$$

где w_i – весовые коэффициенты, g – число узлов.

1.4 Сравнение

Критериями для сравнения были выбраны: влияние шума на построение аппроксимирующей функции, и сложность алгоритма построения аппроксимирующей функции.

При интерполяции в узлах значения функции-интерполянта в точности равны данным, в то время как при аппроксимации допустимы отклонения от заданных точек. Так как при измерении возникают погрешности, имеет смысл их не учитывать [4].

Решение задач интерполяции и аппроксимации сводится к решению СЛАУ. При построении сплайнов, матрица СЛАУ является трехдиагональной, т.е. система может быть решена за линейное время методом прогонки [3]. При полиномиальной аппроксимации в общем виде, решение полной системы методом Гаусса происходит за кубическое время [2, 3].

В табл. 1 приведена сравнительная характеристика методов, рассмотренных в данном разделе.

Таблица 1. Сравнение методов аппроксимации

Критерий	Интерполяция	Полиномиальная аппроксимация	Аппроксимация В-сплайнами
Шум	Шум влияет на результат	Происходит сглаживание	Происходит сглаживание
Сложность алгоритма	Вычисление коэффициентов для функции- интерполянта – решение СЛАУ с трехдиагональной матрицей.	Вычисление оптимальных коэффициентов – решение СЛАУ.	Вычисление оптимальных коэффициентов – решение СЛАУ с трехдиагональной матрицей.

2. Аппроксимация

2.1 Задача аппроксимации

Задана сетка $\omega = [t_i, t_{i+1}]$, $i = 0, 1, \dots, m$ и набор данных

$$y_0, y_2, \dots, y_m.$$

При предположении, что данные представляют собой некоторую лежащую в их основе функцию $y(t)$, для которой $y_i = y(t_i)$ [4, 5], т.е. задана модель данных в виде

$$\text{наблюдение} = \text{модель} + \text{ошибка}.$$

Иными словами, можно записать функцию $y(t)$ следующим образом

$$y_i \approx x_0 \phi_0(t_i) + x_1 \phi_1(t_i) + \dots + x_n \phi_n(t_i),$$

где $\phi_j(t)$ – модельные функции.

Поставим задачу приближения данных, используя матрично-векторные обозначения. Матрица $A^{m \times n}$ определяется как

$$a_{ij} = (A)_{ij} = \phi_j(t_i),$$

а b и x – векторы наблюдений (данных) и параметров соответственно. Тогда модель примет вид

$$\begin{pmatrix} \phi_0(t_0) & \dots & \phi_n(t_0) \\ \vdots & & \vdots \\ \phi_0(t_m) & \dots & \phi_n(t_m) \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ x_n \end{pmatrix} \approx \begin{pmatrix} y_0 \\ \vdots \\ y_m \end{pmatrix}$$

или более кратко

$$Ax \approx y \text{ или } Ax - y \approx 0.$$

Согласно методу наименьших квадратов, для минимизации значений $y - Ax$ необходимо решить задачу

$$\min_x \sum_{j=0}^m [(b - Ax)_j]^2 \quad (1.1)$$

или в эквивалентной форме с использованием евклидовой 2-нормы вектора

$$\min_x \|y - Ax\|_2^2 = (y - Ax)^T (y - Ax).$$

В случаях, когда не все данные одинаково важны или известно, что данные имеют разную точность, можно ввести веса $w_i > 0, \sum_{i=0}^n w_i = 1$ и минимизировать сумму

$$\sum_{j=0}^m w_j [(y - Ax)_j]^2.$$

Для получения минимума решение должно удовлетворять системе уравнений [4]:

$$(A^T A)x = (A^T y). \quad (1.2)$$

С учетом весовых коэффициентов:

$$(A^T W^2 A)x = (A^T W^2 y)$$

где $W = \begin{pmatrix} w_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_m \end{pmatrix}$ – диагональная матрица, составленная из весовых коэффициентов.

2.2 Решение задачи аппроксимации В-сплайнами

Как было указано ранее, на сетке

$$\omega^* = [t_i, t_{i+1}], i = 0, \dots, n + 2k + 1$$

$$t_0 = t_1 = \dots = t_k$$

$$t_{n+k} = t_{n+k+1} = \dots = t_{n+2k+1}$$

сплайн может быть представлен в базисной форме

$$S(t) = \sum_{i=0}^{n+k} x_i N_i^k(t). \quad (2.1)$$

Заданы данные $(t_i, y_i), i = 0, \dots, m$, которые необходимо аппроксимировать. Используем В-сплайны в качестве модельных функций. Подставим (2.1) в (1.1), получим выражение для ошибки

$$r = \sum_{i=0}^m \left[y_i - \sum_{j=0}^{n+k} x_j N_j^k(t) \right]^2$$

или в матричном виде с использованием нормы вектора

$$r = \|y - Ax\|_2^2$$

где

$$A = \begin{pmatrix} N_0^k(t_0) & \cdots & N_{n+k}^k(t_0) \\ \vdots & \ddots & \vdots \\ N_0^k(t_m) & \cdots & N_{n+k}^k(t_m) \end{pmatrix}, x = \begin{pmatrix} x_0 \\ \vdots \\ x_{n+k} \end{pmatrix}, y = \begin{pmatrix} y_0 \\ \vdots \\ y_m \end{pmatrix}$$

Задача аппроксимации сведется к поиску коэффициентов x , минимизирующих ошибку:

$$\min_x r(x).$$

Обозначим

$$\langle N_i^k, N_j^k \rangle = \sum_{l=0}^m N_i^k(x_l) N_j^k(x_l),$$

$$\langle N_i^k, b \rangle = \sum_{l=0}^m y_l N_i^k(x_l)$$

и, подставляя выражения в (1.2), получим

$$(A^T A) = \begin{pmatrix} \langle N_0^k, N_0^k \rangle & \cdots & \langle N_0^k, N_{n+k}^k \rangle \\ \vdots & \ddots & \vdots \\ \langle N_{n+k}^k, N_0^k \rangle & \cdots & \langle N_{n+k}^k, N_{n+k}^k \rangle \end{pmatrix}, (A^T y) = \begin{pmatrix} \langle N_0^k, b \rangle \\ \vdots \\ \langle N_{n+k}^k, b \rangle \end{pmatrix}.$$

В результате решения системы (1.2) будет получен оптимальный вектор коэффициентов x .

Таким образом, получена система уравнений, решив которую, можно получить аппроксимирующий сплайн.

2.3 Кубический сплайн

Пример аппроксимации кубическим сплайном представлен на рис. 1.

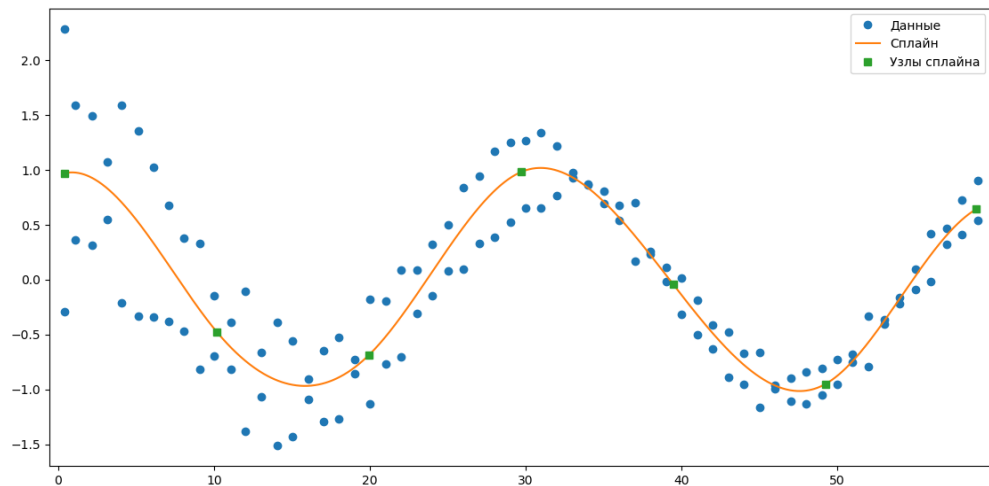


Рисунок 1. Аппроксимация данных кубическим сплайном.

При аппроксимации в одной точке может быть задано больше одного значения на сетке с непостоянным шагом. Как видно по рис. 1, 7 узлов для сплайна достаточно для аппроксимации данных из 120 значений.

При небольшом количестве данных относительно степени сплайна и количеству узлов, возникает проблема «переподгонки» [4], показанное на рис. 2.

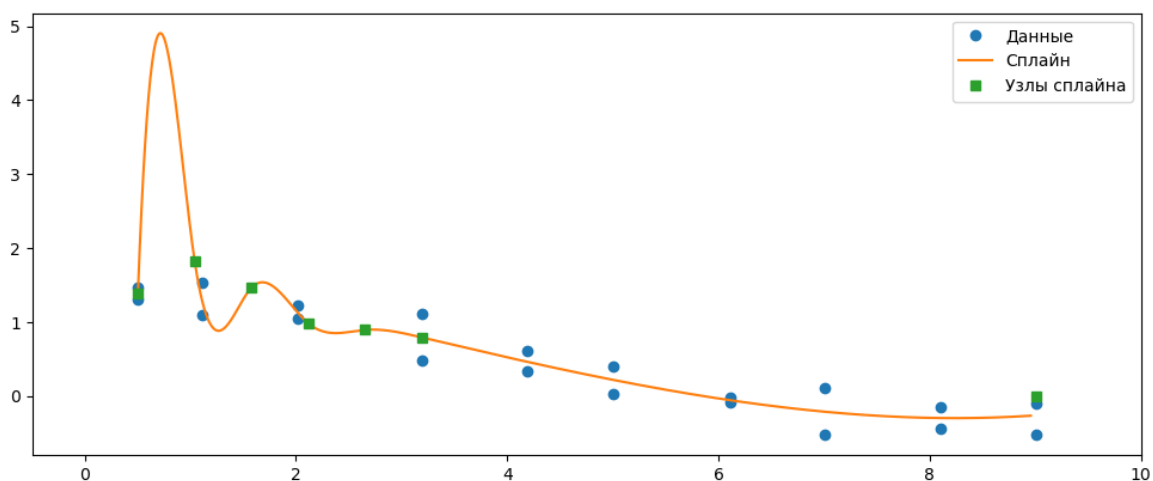


Рисунок 2. «Переподгонка»

3. Программная реализация

3.1 Используемые технологии

Было разработано приложение с графическим интерфейсом с использованием языка программирования Python. Для построения графиков была использована библиотека matplotlib 3, а для построения графического интерфейса библиотека – tkinter. Matplotlib 3 является наиболее популярной библиотекой для визуализации данных и проста в использовании, при этом обладая большим функционалом. Пример построения графика, изображенного на рис. 3, с фиксированными точками с настройками по умолчанию:

```
plt.plot([1, 2, 3, 4, 5, 6], [10, 5, 9, 1, 3, 8,])  
plt.show()
```

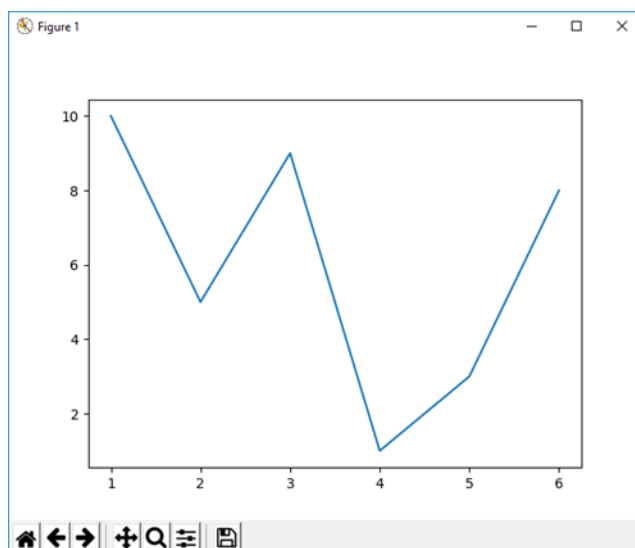


Рисунок .3 Пример работы matplotlib

В качестве средств визуализации matplotlib 3 использует tkinter [9], поэтому совместное использование двух библиотек наиболее простое при разработке приложения с графическим интерфейсом.

Источником данных является загружаемый файл в формате CSV, который должен содержать непосредственно наблюдаемые данные и точки, через которые будет проведён сплайн.

3.2 Тестовые данные

В качестве наблюдаемых данных может быть использован набор значений, распределенный по времени (например, погодные данные за какой-то период [10]). Критерий соответствия построенного сплайна фактической закономерности может быть только при наличии исходной функции, значения которой аппроксимируются. На практике часто оказывается, что исходная закономерность неизвестна, и правильность построения сплайна определяется исследователем.

3.3 Описание разработанной программы

Программная реализация предоставляет класс, содержащий методы для построения аппроксимирующей функции. Доступна по ссылке <https://github.com/Veelz/interp>.

4. Заключение

4.1 Выводы

В работе были рассмотрены метод аппроксимации на нерегулярной сетке с помощью сплайнов в виде линейной комбинации В-сплайнов. Приведены примеры построенных кубических сплайнов.

Список литературы

1. Истомин И. А., Котляров О. Л., Лоскутов А. Ю. К проблеме обработки временных рядов: расширение возможностей метода локальной аппроксимации посредством сингулярного спектрального анализа //Теоретическая и математическая физика. – 2005. – Т. 142. – №. 1. – С. 148-159.
2. Березин И. С., Жидков Н. П. Методы вычислений. – 1962.
3. Волков Е. Численные методы. – 1987.
4. Каханер Д., Моулер К., Нэш С. Численные методы и программное обеспечение. – М. : Мир, 1998.
5. Шикин Е. В., Плис А. И. Кривые и поверхности на экране компьютера. – М. : АО" Диалог-МИФИ", 1996.
6. Роджерс Д. Математические основы машинной графики. – 1980.
7. Бур К. Практическое руководство по сплайнам. – 1985.
8. Переобучение. URL: <https://ru.wikipedia.org/wiki/Переобучение> (дата обращения: 15.12.2018).
9. Matplotlib URL: <https://matplotlib.org/> (дата обращения: 15.12.2018).
10. ВНИИГМИ-МЦД URL: <http://meteo.ru/data> (дата обращения: 15.12.2018).