

МАТЕМАТИЧЕСКИЕ МЕТОДЫ АППРОКСИМАЦИИ ДАННЫХ

А.В. Полуэктов¹, А.И. Заревич¹, Ф.В. Макаренко¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. В работе математические методы аппроксимации данных, история их появления, алгоритмы необходимые для выполнения математических расчётов, а также представлены программы на языке программирования *C#* для программной обработки данных. Выполняется сравнительный анализ каждого из методов, оцениваются их индивидуальные особенности, проводится сравнительный анализ.

Ключевые слова: аппроксимация данных, язык программирования *C#*, линейная аппроксимация, полиномиальная аппроксимация, аппроксимацию сплайнами, аппроксимация с использованием нейронных сетей, MathNet, ALGLIB, Accord.NET.

MATHEMATICAL METHODS FOR DATA APPROXIMATION

A.V. Poluektov¹, A.I. Zarevich¹, Ph.V. Makarenko¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. This work examines mathematical methods for data approximation, their history of development, algorithms necessary for performing mathematical calculations, and also presents programs in the *C#* programming language for software data processing. A comparative analysis of each method is performed, their individual characteristics are evaluated, and a comparative analysis is conducted.

Key words: data approximation, *C#* programming language, linear approximation, polynomial approximation, spline approximation, approximation using neural networks, MathNet, ALGLIB, Accord.NET.

Проведение серии опытов заканчивается появлением довольно обширного набора данных, который затем необходимо обработать для определения функции которая с определенной степенью погрешности описывала бы проведенный эксперимент. Для решения данной задачи необходимо провести аппроксимацию

данных, при этом возникает проблема, связанная с тем, что существует довольно большой класс математических методов для аппроксимации данных, так и необходимость компьютерной обработки данных. Необходимость компьютерной обработки данных привязана к тому, что объем обрабатываемых данных может содержать сотни данных, а метод может использовать средства искусственного интеллекта. Поэтому рассмотрим, как сами методы, их историю появления, математический алгоритм преобразования данных, так и пример обработки данных с использованием языка программирования C#.

На сегодняшний день существуют множество методов аппроксимации данных, из них мы рассмотрим: линейную, полиномиальную, аппроксимацию сплайнами, а также аппроксимацию с использованием нейронных сетей.

Первые три метода используются в тех случаях, когда наборы данных визуально «подсказывают» вид графика, последний, когда наборы данных порядка нескольких тысяч, а график визуально не отображается.

Рассмотрим линейную аппроксимацию. Данный метод является одним из старейших и широко используется в математике и прикладных науках. Данный метод использовали ученые и философы при попытке описать природные явления с помощью простых математических моделей. В построение данного метода внесли Пьер-Симон Лаплас, Карл Фридрих Гаусс и Адриен-Мари Лежандр.

Рассмотрим алгоритм выполнения расчетов.

1. Пусть имеется набор из n точек (x_i, y_i) , где $i=1, 2, \dots, n$.
2. Вычисляются средние значения: $\bar{x}$ и $\bar{y}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

3. Вычисляются коэффициент наклона a и свободный член b :

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

$$b = \bar{y} - a\bar{x}$$

4. После нахождения a и b уравнение прямой принимает вид: $y = ax + b$.
5. Проводится оценка качества аппроксимации через коэффициент детерминации R^2 :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где $\hat{y}_i = ax_i + b$ — предсказанное значение.

На основании представленного алгоритма могут быть проведены исследования как в ручном подсчете, так и с использованием программы, ниже представлен код метода на языке C#.

```
using System;
using MathNet.Numerics.LinearRegression;
class Program
{
    static void Main()
    {
        double[] x = { 1, 2, 3, 4, 5 };
        double[] y = { 2, 4, 5, 4, 5 };
        var (a, b) = SimpleRegression.Fit(x, y);
        Console.WriteLine($"Линейная аппроксимация: y = {a}x + {b}");
    }
}
```

Следующий метод аппроксимации – полиномиальная аппроксимация. Она позволяет описывать более сложные зависимости с помощью полиномов. В разработке данного метода участвовали такие ученые, как Исаак Ньютон, Джозеф Луи Лагранж и Карл Фридрих Гаусс.

Полиномиальная аппроксимация заключается в приближении функции или набора данных полиномом степени m :

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m,$$

где $P(x)$ — аппроксимирующий полином, $a_0, a_1, \dots, a_m$ — коэффициенты полинома, m — степень полинома.

Рассмотрим алгоритм для проведения расчетов.

1. Пусть имеется набор из n точек (x_i, y_i) , где $i=1, 2, \dots, n$.
2. Определяется степень полинома m . Обычно $m < n$.
3. Составляется система уравнений. Для нахождения коэффициентов $a_0, a_1, \dots, a_m$ решается система линейных уравнений, которая возникает из условия минимизации суммы квадратов отклонений:

$$\sum_{i=1}^n (y_i - (a_0 + a_1x + a_2x^2 + \dots + a_mx^m))^2 \rightarrow \min$$

Эта система может быть записана в матричной форме: $X^T X a = X^T y$,

где X — матрица значений независимой переменной $X =$

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{pmatrix}, a — \text{вектор коэффициентов } a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, y — \text{вектор значений зависимой переменной } y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

4. Выполняется нахождение корней системы уравнений, т.е. определяются коэффициенты a .

5. Оценка качества аппроксимации строится на основании определения коэффициента детерминации R^2 : $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - P(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$, где $\bar{y}$ — среднее значение y .

Для программной обработки в $\text{c\#}$ можно использовать библиотеку `MathNet.Numerics`.

```
using System;
using MathNet.Numerics;
class Program
{
    static void Main()
    {
        double[] x = { 1, 2, 3, 4, 5 };
        double[] y = { 2, 4, 5, 4, 5 };
        var coefficients = Fit.Polynomial(x, y, 2); // Полином 2-й степени
        Console.WriteLine("Полиномиальная аппроксимация: y = " + coefficients[2] + "x^2 + " + coefficients[1] + "x + " + coefficients[0]);
    }
}
```

Если обработать эксперимент требуется очень точно, то здесь подойдет метод — Аппроксимация сплайнами. Данный термин "сплайн" происходит от гибких линеек, которые использовались в кораблестроении и черчении для создания гладких кривых. Основной вклад в развитие теории сплайнов внесли такие ученые, как Исаак Шёнберг (который ввел понятие сплайна в 1946 году), Карл де

Бур (разработка В-сплайнов) и Пьер Безье (применение сплайнов в компьютерной графике). Наиболее распространены кубические сплайны, которые используют полиномы третьей степени.

Кубический сплайн на отрезке $[x_i, x_{i+1}]$ задается в виде:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3,$$

где a_i, b_i, c_i, d_i — коэффициенты сплайна.

Опишем алгоритм расчетов

1. Пусть имеется набор из n точек (x_i, y_i) , где $i=1, 2, \dots, n$, и $x_1 < x_2 < \dots < x_n$.
2. Определение узлов. Узлы сплайна — это точки x_i , в которых соединяются сегменты сплайна.
3. Задание условий гладкости. Для кубического сплайна требуется выполнение следующих условий:
 - непрерывность функции: $S_i(x_i) = S_{i-1}(x_i)$;
 - непрерывность первой производной: $S'_i(x_i) = S'_{i-1}(x_i)$;
 - непрерывность второй производной: $S''_i(x_i) = S''_{i-1}(x_i)$.
4. Определяем граничные условия. Для однозначного определения сплайна задаются граничные условия на концах интервала. Например:
 - естественные граничные условия: $S''(x_1) = S''(x_n) = 0$;
 - условия на производные: $S'(x_1) = y'_1, S'(x_n) = y'_n$.
5. На основе условий гладкости и граничных условий составляется система линейных уравнений для нахождения коэффициентов a_i, b_i, c_i, d_i .
6. После нахождения коэффициентов сплайн на каждом отрезке $[x_i, x_{i+1}]$ задается в виде $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$.

Для работы со сплайнами в C# можно использовать библиотеку ALGLIB.

```
using System;
using alglib;
class Program
{
    static void Main()
    {
        double[] x = { 1, 2, 3, 4, 5 };
        double[] y = { 2, 4, 5, 4, 5 };
        spline1dinterpolant s;
        spline1dbuildcubic(x, y, out s);
    }
}
```

```

        Console.WriteLine("Аппроксимация сплайнами: значение в точке
x=3.5: " + spline1dcalc(s, 3.5) );
    }
}

```

Последняя аппроксимация предназначена для оценки сложных нелинейных зависимостей для этого используется аппроксимация с использованием нейронных сетей. Данный метод строится на основе использования машинного обучения и искусственного интеллекта. В основе данного метода были работы Франка Розенблатта, Дэвида Румельхарта, Джорджа Хинтона и других ученых.

Идея использования нейронных сетей строится на возможности минимизировать ошибку между предсказанными значениями и истинными значениями функции. А для этого используются различные архитектуры нейронных сетей, такие как:

- многослойный перцептрон (MLP) ;
- радиально-базисные нейронные сети (RBF) ;
- сверточные нейронные сети (CNN) ;
- рекуррентные нейронные сети (RNN).

Процесс обучения нейронной сети для аппроксимации функции можно разделить на несколько основных этапов:

1. Собрать набор данных $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, разделить исходные данные на обучающую, валидационную и тестовую выборки.

- Выбрать тип нейронной сети и задать следующие параметры: количество скрытых слоев; количество нейронов в каждом слое; активационные функции.

2. Инициализировать веса и смещения случайными значениями.

3. Использовать алгоритм обратного распространения ошибки для минимизации функции потерь.

4. Если качество аппроксимации не удовлетворяет требованиям, можно изменить архитектуру сети, размер обучающей выборки или параметры обучения.

Для работы с нейронными сетями в C# используется библиотека Accord.NET:

```

using System;
using Accord.Neuro;

```

```

using Accord.Neuro.Learning;
class Program
{
    static void Main()
    {
        double[][] inputs = { new double[] { 1 }, new double[] { 2 }, new double[]
{ 3 }, new double[] { 4 }, new double[] { 5 } };
        double[] outputs = { 2, 4, 5, 4, 5 };
        var network = new ActivationNetwork(new SigmoidFunction(), 1, 5, 1);
        var teacher = new BackPropagationLearning(network);
        for (int i = 0; i < 1000; i++)
            teacher.RunEpoch(inputs, outputs);
        double result = network.Compute(new double[] { 3.5 }) [0];
        Console.WriteLine("Аппроксимация нейронной сетью: значение в
точке x=3.5: " + result);
    }
}

```

Таким образом, математические методы аппроксимации данных играют важную роль в анализе и моделировании сложных систем и выбор метода зависит от характера данных, требуемой точности и вычислительных ресурсов. Понимание особенностей каждого метода позволяет эффективно применять их в различных областях науки и техники, а использование практических реализаций на языке программирования C# упростит обработку данных.

Список литературы

1. Нейдорф, Р. А. Математическое описание трехмерных существенно нелинейных зависимостей по экспериментальным данным методом "Cut-Glue" аппроксимации / Р. А. Нейдорф, В. В. Полях // Прикаспийский журнал: управление и высокие технологии. – 2018. – № 3(43). – С. 10-19. – EDN YTVYFF.
2. Нейдорф, Р. А. Математическое описание трехмерных существенно нелинейных зависимостей по экспериментальным данным методом "Cut-Glue" аппроксимации / Р. А. Нейдорф, В. В. Полях // Прикаспийский журнал: управление и высокие технологии. – 2018. – № 3(43). – С. 10-19. – EDN YTVYFF.

3. Бирюков, А. Н. Модифицированный метод вложенных математических моделей, основанный на взаимосвязанном управлении информативностью данных и качеством их аппроксимации в байесовском ансамбле нейросетей / А. Н. Бирюков, О. И. Глущенко // Перспективы развития информационных технологий. – 2011. – № 3-2. – С. 182-187. – EDN RRYEWZ.

4. Антибас, И. Р. Сравнение численного метода и компьютерного моделирования линейной регрессии оптимального профиля среза транспортирующего устройства от величины подачи хлебной массы / И. Р. Антибас, А. Г. Дьяченко, Т. П. Савостина // Научная жизнь. – 2019. – Т. 14, № 9(97). – С. 1432-1440. – DOI 10.35679/1991-9476-2019-14-9-1432-1440. – EDN QUGNQJ.

5. Макаров, А. А. Теория минимальных сплайн-всплесков и ее приложения: специальность 05.13.18 "Математическое моделирование, численные методы и комплексы программ": диссертация на соискание ученой степени доктора физико-математических наук / Макаров Антон Александрович. – Санкт-Петербург, 2012. – 345 с. – EDN QFNGZT.

6. Яковлев, К. Н. Аппроксимация экспериментальных данных методом наименьших квадратов аннотация / К. Н. Яковлев // ПЕРСПЕКТИВЫ науки и ОБЩЕСТВА в УСЛОВИЯХ ИННОВАЦИОННОГО РАЗВИТИЯ : сборник статей Международной научно-практической конференции, Пермь, 12 марта 2022 года. – Уфа: Общество с ограниченной ответственностью "ОМЕГА САЙНС", 2022. – С. 35-43. – EDN JGOTYP.

7. Парамонов, А. А. Исследование структуры тропического циклона на основе анализа данных спутникового видеоряда с применением метода почти периодического анализа / А. А. Парамонов, А. В. Калач, А. С. Кравченко // Моделирование систем и процессов. – 2024. – Т. 17, № 4. – С. 67-76. – DOI 10.12737/2219-0767-2024-17-4-67-76. – EDN BOGSTY.

8. Сумин, В. И. Использование ситуационного моделирования в разработке систем принятия решений для сложных организационных систем / В. И. Сумин, А. С. Кравченко, А. В. Толкачев // Моделирование систем и процессов. – 2024. – Т. 17, № 3. – С. 71-79. – DOI 10.12737/2219-0767-2024-69-77. – EDN FBBHJO.

References

1. Neidorf, R. A. Mathematical Description of Three-Dimensional Strongly Nonlinear Dependencies Based on Experimental Data Using the "Cut-Glue" Approximation Method / R. A. Neidorf, V. V. Polyakh // Caspian Journal: Management and High Technologies. – 2018. – No. 3(43). – Pp. 10-19. – EDN YTVYFF.

2. Neidorf, R. A. Mathematical Description of Three-Dimensional Strongly Nonlinear Dependencies Based on Experimental Data Using the "Cut-Glue" Approximation Method / R. A. Neidorf, V. V. Polyakh // Caspian Journal: Management and High Technologies. – 2018. – No. 3(43). – Pp. 10-19. – EDN YTVYFF.

3. Biryukov, A. N. A Modified Method of Nested Mathematical Models Based on Interrelated Control of Data Informativeness and Approximation Quality in a Bayesian Ensemble of Neural Networks / A. N. Biryukov, O. I. Glushchenko // Prospects for the Development of Information Technologies. – 2011. – No. 3-2. – Pp. 182-187. – EDN RRYEWZ.

4. Antibas, I. R. Comparison of a Numerical Method and Computer Simulation of Linear Regression for the Optimal Profile of a Conveying Device's Cut Depending on the Feed Rate of Bread Mass / I. R. Antibas, A. G. Dyachenko, T. P. Savostina // Scientific Life. – 2019. – Vol. 14, No. 9(97). – Pp. 1432-1440. – DOI 10.35679/1991-9476-2019-14-9-1432-1440. – EDN QUGNQJ.

5. Makarov, A. A. Theory of Minimal Spline Wavelets and Its Applications: Specialty 05.13.18 "Mathematical Modeling, Numerical Methods, and Software Complexes": Dissertation for the Degree of Doctor of Physical and Mathematical Sciences / Anton Aleksandrovich Makarov. – St. Petersburg, 2012. – 345 p. – EDN QFNGZT.

6. Yakovlev, K. N. Approximation of Experimental Data Using the Least Squares Method: Abstract / K. N. Yakovlev // Prospects for Science and Society in the Context of Innovative Development: Collection of Articles from the International Scientific and Practical Conference, Perm, March 12, 2022. – Ufa: Omega Science LLC, 2022. – Pp. 35-43. – EDN JGOTYP.

7. Paramonov, A. A. Study of the Structure of a Tropical Cyclone Based on the Analysis of Satellite Video Data Using the Almost Periodic Analysis Method / A. A. Paramonov, A. V. Kalach, A. S. Kravchenko // Modeling of Systems and Processes. – 2024. – Vol. 17, No. 4. – Pp. 67-76. – DOI 10.12737/2219-0767-2024-17-4-67-76. – EDN BOGSTY.

8. Sumin, V. I. Use of Situational Modeling in the Development of Decision-Making Systems for Complex Organizational Systems / V. I. Sumin, A. S. Kravchenko, A. V. Tolkachev // Modeling of Systems and Processes. – 2024. – Vol. 17, No. 3. – Pp. 71-79. – DOI 10.12737/2219-0767-2024-69-77. – EDN FBBHJO.