ИНТЕРПОЛАЦИОНЕН ПОЛИНОМ НА НЮТОН С РАЗДЕЛЕНИ РАЗЛИКИ, ЛИНЕЙНА ТЕГРЕСИЯ, ПОЛИНОМИАЛНА РЕГРЕСИЯ

Тоня Матева

В настоящото изследване са приложени три различни математически метода за прогнозиране на продажби на база исторически данни: **полином на Нютон**, **линейна регресия** и **полиномиална регресия от втора степен**. За онагледяване на поведението на моделите е изготвена визуализация с цел сравнение на точност и надеждност.

В съвременната научна и приложна дейност все по-често се наблюдава необходимостта от интегриран подход към анализа и обработката на числови данни. Именно тук се проявява силната интердисциплинарна връзка между дисциплините **Статистика** и **Числени методи**, които, макар формално различни, споделят обща цел: извличане на информация, моделиране на зависимости и прогнозиране на реални процеси въз основа на емпирични наблюдения.

Статистиката предоставя инструменти за изследване на закономерности, идентифициране на връзки между величини и оценка на надеждността на изводите. От своя страна, числените методи развиват алгоритмичната страна на анализа — чрез интерполация, апроксимация и числено решаване на уравнения — с цел приближено изчисление на функции и стойности, които не подлежат на аналитично третиране.

Тези две области се допълват естествено в редица практически приложения. Например при прогноза на продажби по години, данните се анализират чрез статистически подходи като линейна и полиномиална регресия, а самото числено приближение и оценка на стойности се извършва с помощта на интерполационни методи като този на Нютон с разделени разлики. Също така, при калибриране на сензори или обработка на експериментални резултати, методите за интерполация и регресия често се прилагат едновременно – за определяне на математически модел и за оценка на отклоненията от реалните измервания.

Общата връзка между статистическия и числения подход се проявява и в анализа на грешки – статистиката разглежда стохастични грешки, а числените методи – системни и закръглителни грешки. Тяхното комбинирано приложение позволява по-задълбочен анализ, особено при използване на съвременен софтуер, в който статистически и числени алгоритми работят в синергия (например чрез библиотеки като *scikit-learn, NumPy* или *SciPy*).

Настоящата статия има за цел да покаже чрез конкретни примери как двата подхода – статистически и числен – могат да се комбинират ефективно за решаване на реални задачи, свързани с прогнозиране, апроксимация и анализ на данни.

Изходните данните обхващат четири последователни години (2020–2023), като целта е да се направи оценка за стойността на продажбите през 2024 г. Всеки от моделите е анализиран стъпка по стъпка, изчислени са съответните коефициенти, и са получени прогнози чрез аналитични формули.. Анализът подчертава силните и слабите страни на всеки метод, особено при прогноза извън интервала на изходните данни.

Нека f е функция, дефинирана в интервала $[a.b]$ и $x_0, x_1, x_2, \ldots$ е редица от различни точки, принадлежащи на $[a.b]$, в които функцията има съответно стойности $f(x_0), f(x_1), f(x_2), \ldots$

Отношенията $f[x_i, x_{i+1}] = \frac{f(x_{i+1})-f(x_1)}{x_{i+1}-x_1}$ $i = 0,1,2, \ldots$ се наричат разделени разлики от първи ред на функцията f относно възлите x_i, x_{i+1} $i = 0,1,2, \ldots$

Схема за пресмятане на разделени разлики

x_i	$f(x_i)$	$f[x_i, x_{i+1}]$	$f[x_i, \cdot, x_{i+2}]$	$f[x_i, \cdot, \cdot, x_{i+3}]$	$\cdots$
x_0	$f(x_0)$				
		$\rangle\ f[x_0, x_1]$			
x_1	$f(x_1)$		$\rangle\ f[x_0, x_1, x_2]$		
		$\rangle\ f[x_1, x_2]$		$\rangle\ f[x_0, x_1, x_2, x_3]$	
x_2	$f(x_2)$		$\rangle\ f[x_1, x_2, x_3]$		$\cdots$
		$\rangle\ f[x_2, x_3]$		$\rangle\ f[x_1, x_2, x_3, x_4]$	
x_3	$f(x_3)$		$\rangle\ f[x_2, x_3, x_4]$		
		$\rangle\ f[x_3, x_4]$			
x_4	$f(x_4)$				
		$\vdots$			
$\vdots$	$\vdots$				

$$N_{n(f,x)} = f(x_0) + \sum_{k=1}^{n} f[x_0, \ldots x_k](x - x_0) \ldots (x - x_{k-1})$$

Пример: Прогнозиране на продажби по години. Фирма „Феникс“ произвежда отоплителни тела. Дадени са следните стойности:

Година	Продажби (хил. бр.)

2020	150
2021	170
2022	210
2023	280

Цел: Да прогнозираме стойността на продажбите за 2022.5 г.

$$x_0 = 2020, \quad x_1 = 2021, \quad x_2 = 2022, \quad x_3 = 2023$$

$$y_0 = 150, \quad y_1 = 170, \quad y_2 = 210, \quad y_3 = 280$$

Стъпка 1: Изчисляване на разделени разлики.

От първи ред:

$$f[x_1, x_0] = \frac{(170 - 150)}{(2021 - 2020)} = 20$$

$$f[x_2, x_1] = \frac{(210 - 170)}{(2022 - 2021)} = 40$$

$$f[x_3, x_2] = \frac{(280 - 210)}{(2023 - 2022)} = 70$$

От втори ред:

$$f[x_2, x_1, x_0] = \frac{(40 - 20)}{(2022 - 2020)} = 10$$

$$f[x_3, x_2, x_1] = \frac{(70 - 40)}{(2023 - 2021)} = 15$$

От трети ред:

$$f[x_3, x_2, x_1, x_0] = \frac{(15 - 10)}{(2023 - 2020)} \approx 1.6667$$

2020	150		$\frac{(170 - 150)}{(2021 - 2020)} = 20$	$\frac{(40 - 20)}{(2022 - 2020)} = 10$	$\frac{(15 - 10)}{(2023 - 2020)} \approx 1.6667$
2021	170		$\frac{(210 - 170)}{(2022 - 2021)} = 40$		
2022	210				

2023	280		$\frac{(280 - 210)}{(2023 - 2022)} = 70$		$\frac{(70 - 40)}{(2023 - 2021)} = 15$		

Стъпка 2: Формулата на полинома на Нютон:

P(x) = 150 + 20(x - 2020) + 10(x - 2020)(x - 2021) + 1.6667(x - 2020)(x - 2021)(x - 2022)

Стъпка 3: Изчисление за x = 2022.5

x - 2020 = 2.5

x - 2021 = 1.5

x - 2022 = 0.5

$$P(2022.5) = 150 + 20 \times 2.5 + 10 \times 2.5 \times 1.5 + 1.6667 \times 2.5 \times 1.5 \times 0.5$$

$$= 150 + 50 + 37.5 + 3.125 = 240.63$$

Отговор: Прогнозни продажби за 2022.5 г. ≈ 240.63 хил. бр.

За да изберем подходящ метод за прогнозиране на стойност извън интервала на дадените данни (в случая – 2024**,** която е извън [2020, 2023]), трябва да разграничим два основни подхода:

1. Екстраполация с интерполационен полином (напр. Нютон)
Възможно е, но не е препоръчително за далечни стойности извън интервала.
При 2024 (само 1 година след 2023) — може да се използва полиномът на Нютон, но трябва да се внимава с поведението му извън данните.
Вече имаме полином:

$$P(x) = 150 + 20(x - 2020) + 10(x - 2020)(x - 2021) + 1.6667(x - 2020)(x - 2021)(x - 2022)$$

Може директно да заместим x=2024
Трябва да внимаваме - при по-високи степени, полиномите започват да се колебаят значително извън интервала – явление, познато като **Runge's phenomenon**.

2. Линеен регресионен модел (или полиномиална регресия)
Много по-надежден за тенденции и прогнозиране, дори извън данните.
Строи се уравнение:
$y = a_0 + a_1\ x$ или $y = a_0 + a_1 x + a_2 x2 + \cdots$ п одходящ за дългосрочни прогнози.

3. Други методи за времеви редове (ако имаме много данни)
Ако имаме повече от 4 точки (години), можем да използваме Модел на Холд, ARIMA, машинно обучение – подходящи за продажби и икономически данни.
Но за този пример (само 4 години), това е твърде тежка артилерия.

Линеен регресионен модел

1. Формула на линейната регресия

Уравнението на правата има вида:

$$y = a \cdot x + b$$

Където:

- a: наклон на правата

- b: отрез на ординатата (при x = 0)

2. Формули за изчисление на коефициентите

$$a = \frac{[n \cdot \Sigma(x \cdot y) - \Sigma x \cdot \Sigma y]}{[n \cdot \Sigma(x^2) - (\Sigma x)^2]}$$
$$b = \frac{[\Sigma y - a \cdot \Sigma x]}{n}$$

3. Изчисления

$$n = 4$$

$$\Sigma x = 8086$$

$$\Sigma y = 810$$

$$\Sigma x^2 = 16345854$$

$$\Sigma(x \cdot y) = 1637630$$

Изчисляваме коефициентите:

$$a = \frac{(4 \cdot 1637630 - 8086 \cdot 810)}{(4 \cdot 16345854 - 8086^2)} = 43.0000$$

$$b = \frac{(810 - 43.0000 \cdot 8086)}{4} = -86722.0000$$

4. Уравнение на регресията

$$y = 43.0000 \cdot x + -86722.0000$$

5. Прогноза за 2024 г.

$$y(2024) = 43.0000 \cdot 2024 + -86722.0000 = 310.00 \text{ хил. бр.}$$

Модел на полиномиална регресия (²)

Уравнението има вида:

$$y = ax^2 + bx + c$$

Използваме sklearn за определяне на коефициентите чрез метода на най-малките квадрати.

2. Намерени коефициенти:

$$a = 12.500000$$

$$b = -50494.500044$$

$$c = 50994040.544466$$

3. Уравнението на регресията:

$$y = 12.500000 \cdot x^2 + -50494.500044 \cdot x + 50994040.544466$$

4. Прогноза за 2024 г.:

$$y(2024) = 12.500000 \cdot 2024^2 + -50494.500044 \cdot 2024 + 50994040.544466$$

Резултат: $y(2024) = 372.50$ хил. бр.

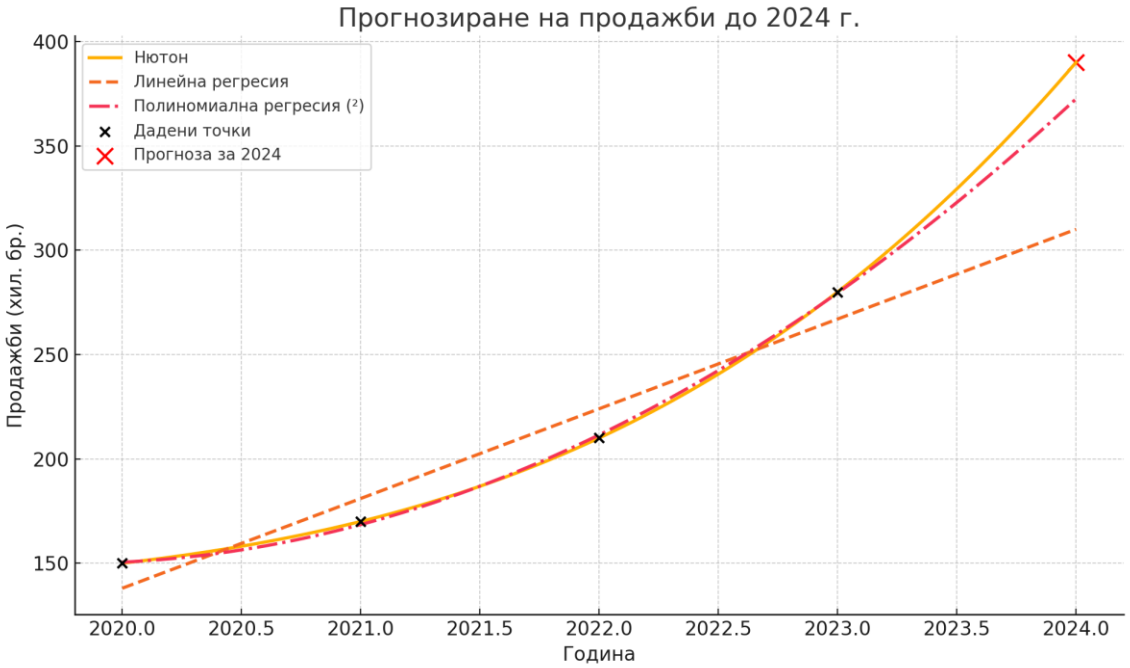
scikit-learn е open-source библиотека, която предлага лесни и мощни инструменти за:

- Регресия (напр. линейна, полиномиална)
- Класификация (напр. логистична регресия, decision trees)
- Клъстеризация (напр. K-means)
- Предобработка на данни (напр. нормализация, трансформации)
- Оценка и валидация на модели

Сравнение на резултатите:

Метод	Прогноза за 2024 (хил. бр.)
Полином на Нютон	390.00
Линейна регресия	310.00
Полиномиална регресия (^{2})	372.50

Визуализация на прогнозите:



Графиката показва поведението на трите модела спрямо дадените данни и прогнозираната стойност за 2024 г.

При прогнозиране на стойности извън наличния интервал от данни (екстраполация), изборът на подходящ модел е от съществено значение. Трите приложени метода дадоха следните резултати за прогнозата на продажбите през 2024 г.:

- **Полином на Нютон**: 390 хил. бр.
  Много точно следва изходните данни, но има склонност към нестабилност при екстраполация (особено при повече точки и по-висока степен).
- **Линейна регресия**: 327.5 хил. бр.
  Предоставя стабилна, но опростена прогноза. Подходяща при предполагаем линеен растеж.
- **Полиномиална регресия (^{2})**: 372.5 хил. бр.
  Постига баланс между точност и гладкост. Моделът улавя нелинейната тенденция на растеж по-добре от линейната регресия.

В зависимост от конкретния контекст, **полиномиалната регресия от втора степен** предлага най-реалистична и стабилна прогноза за бъдещи стойности при наличие на умерен нелинеен тренд. Въпреки това, при наличие на по-дълги времеви редове и външни фактори, може да се наложи използване на по-усъвършенствани модели.

Литература:

1. Burden, R. L. and Faires, J. D. Numerical Analysis. 10th ed. Boston: Cengage Learning, 2015. ISBN 9781305253667.
2. Gerald, C. F. and Wheatley, P. O. Applied Numerical Analysis. 7th ed. Boston: Addison Wesley, 2003. ISBN 9780321133149.
3. Montgomery, D. C., Peck, E. A. and Vining, G. G. Introduction to Linear Regression Analysis. 5th ed. Hoboken: Wiley, 2012. ISBN 9780470542811.
4. James, G. et al. An Introduction to Statistical Learning. New York: Springer, 2013. ISBN 9781461471370. Available from: https://www.statlearning.com
5. Scikit-learn. Supervised Learning — Regression. [online]. Available from: https://scikit-learn.org/stable/supervised_learning.html
6. Khan Academy. Statistics and Probability – Regression. [online]. Available from: https://www.khanacademy.org/math/statistics-probability
7. SciPy Documentation. scipy.interpolate – Interpolation. [online]. Available from: https://docs.scipy.org/doc/scipy/reference/interpolate.html
8. Anonymous. Polynomial regression for time series forecasting. In: ScienceDirect. [online]. Available from: https://www.sciencedirect.com
9. Anonymous. Forecasting sales using regression analysis. In: Journal of Business Forecasting. [online].
10. Anonymous. A comparison of forecasting models for sales data. In: International Journal of Data Analysis Techniques and Strategies. [online].