Method of Machine Learning Based on Discrete Orthogonal Polynomials of Chebyshev

Igor Puleko[1], Svetlana Kravchenko[1], Victor Chumakevych[2] and Vadym Ptashnyk[3]

[1]Zhytomyr Polytechnic State University, Zhytomyr, Ukraine
pulekoigor@gmail.com
[2]Lviv Polytechnic National University, Lviv, Ukraine
chumakevich1970@gmail.com
[3]Lviv National Agrarian University, Lviv, Ukraine
ptashnykproject@gmail.com

Abstract. Among the classical methods of machine learning, regression is often used to solve prediction problems. Nowadays, the linear regression method based on the least squares method has become most widespread. When using higher-order models, the researcher faces the problem of inverting the large-dimensional matrices. To eliminate this drawback, the authors have developed a polynomial regression method based on discrete orthogonal Chebyshev polynomials. Due to the properties of orthogonality and the best approximation, it is possible to construct a recurrent formula for obtaining the values of Chebyshev polynomials of higher degrees, which allows optimizing the process of determining the polynomial coefficients.

Keywords: Machine learning, classical machine learning methods, Regression, Polynomial regression, Chebyshev orthogonal polynomial.

1 Introduction

In recent years, machine learning methods have been widely used to solve the market, product and service prediction problems. Machine learning is a method of artificial intelligence that teaches a computer to solve various applied problems on its own. The purpose of machine learning is to predict the result based on the available input data. The more diverse the input data are, the easier it is for a machine to find some patterns and achieve more accurate results. Machine learning often uses statistical techniques to provide computers with the ability to "learn from data without being explicitly programmed". One of these statistical methods is regression. Regression refers to the classical methods of supervised learning and is based on classical optimization statistical algorithms.

2 Machine Learning Regression Methods

Machine learning regression method is primarily intended for numerical predicting the behavior of numerical data. The regression model is a constructed function of independent variable and of coefficients with random variables included.

The dependent variable is considered to be described by the sum of the values of the model and by independent variables. According to the nature of the distribution of the dependent variable, we make assumptions called the hypothesis of the generation of data. To confirm or reject this hypothesis, statistical tests are conducted (residual analysis is the difference between the values observed and the values predicted by the constructed regression model). It is considered that the dependent variable is free of errors [1–6, 10, 17, 19].

As a model, polynomials are often used. At point-quadratic approximation, as the polynomial deviation

$$Q_m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x_0^m \tag{1}$$

from the given function $y = f(x)$ on the set of points $x_0, x_1, \dots, x_{L-1}$ we take the variable

$$S = \sum_{l=0}^{L-1} [Q_m(x_l) - f(x_l)]^2 , \tag{2}$$

called the quadratic deviation [1–3, 7, 9, 12, 15, 16, 18].

To construct an approximation polynomial, we need to choose the coefficients $a_0, a_1, a_2, \dots, a_m$ so that the value of S would be minimal. Assume $m \leq L - 1$. In the case $m = L - 1$, the coefficients $a_j (j = 0, 1, \dots, L - 1)$ can be determined from the system of equations

$$Q_{L-1}(x_i) = f(x_i) \text{ for } i = 0, 1, \dots, L - 1 , \tag{3}$$

where $S = 0$. The solution of the construction problem follows from the minimization of the functional (2) and leads to the solution

$$A = [X^T X]^{-1} X^T F^T , \tag{4}$$

where

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^m \\ 1 & x_1 & x_1^2 & \cdots & x_1^m \\ . & . & . & \cdots & . \\ 1 & X_{L-1} & X_{L-1}^2 & \cdots & X_{L-1}^m \end{bmatrix}$$

is a matrix-row $m + 1$ of the basic functions $\varphi_0(x) = 1, \varphi_1(x) = x$, $\varphi_2(x) = x^2, \dots, \varphi_m(x) = x^m, \dots$, the values of which are calculated at the points $x_l (l = 0, 1, \dots, L - 1)$, that are expanded in a vertical column;
$F = [f(x_0), f(x_1), \dots, f(x_{L-1})]$ a matrix-row L of the values of the given function on a system of points $x_0, x_1, \dots, x_{L-1}$;

$A^T = [a_0,\ a_1, \ldots, a_m]$ is a transposed matrix-column of unknown coefficients of the polynomial (1).

It follows from (4) that in order to construct a regression model it is necessary to find an inverse matrix. This operation becomes time consuming when the polynomial degree m should be increased. To eliminate this drawback, we suggest a method of construction the polynomial regression based on Chebyshev orthogonal polynomials.

3 Machine Learning Approach Based on Discrete Orthogonal Polynomials of Chebyshev

Let

$$p_0(x), p_1(x), \ldots, p_m(x) \tag{5}$$

be a given system of polynomials orthogonal on a system of points $\{x_0, x_1, \ldots, x_{L-1}\}$. Since the polynomials (5) are linearly independent, an arbitrary polynomial $Q_m(x)$ can be represented as a linear combination of the polynomials from the system (5), i.e.

$$Q_m(x) = b_0 p_0(x)+, \ldots, +b_m p_m(x)\,. \tag{6}$$

This representation is called the expansion of the polynomial $Q_m(x)$ in terms of the system (5).

Perform the orthonormation of the system of basis functions. We call the system of functions defined on the system of points of the interval [0,t] as orthonormal on this interval with the weight ρ if all the functions of this system satisfy the condition

$$(\rho\varphi_h, \varphi_i) = \sum_{l=0}^{L-1} \rho(l)\varphi_h^*(l)\varphi_i(l) = \delta_{h,i}\,. \tag{7}$$

The condition of orthonormality of the polynomials $p_i(x)$ leads to $[X^T X] = E$ and

$$X = P = P(L,m) = \begin{bmatrix} \underset{*}{\hat{p}_0}(L,0) & \underset{*}{\hat{p}_1}(L,0) & \underset{*}{\hat{p}_2}(L,0) & \ldots & \underset{*}{\hat{p}_m}(L,0) \\ \underset{*}{\hat{p}_0}(L,1) & \underset{*}{\hat{p}_1}(L,1) & \underset{*}{\hat{p}_2}(L,1) & \ldots & \underset{*}{\hat{p}_m}(L,1) \\ . & . & . & \ldots & . \\ \underset{*}{\hat{p}_0}(L,L-1) & \underset{*}{\hat{p}_1}(L,L-1) & \underset{*}{\hat{p}_2}(L,L-1) & \ldots & \underset{*}{\hat{p}_m}(L,L-1) \end{bmatrix}.$$

The elements of the matrix $\overline{X}^T$ have the form

$$\underset{*}{\bar{p}_i}(L,l) = \rho(L,l)\underset{*}{\hat{p}_i^*}(L,l)\,.$$

In such a case, the problem of constructing a regression model of the function $y = f(x)$ on the set of points $x_0, x_1, \ldots, x_{L-1}$ orthonormated by a polynomial of the given degree m $(m \leq L-1)$ under the condition of the minimal quadratic deviation will be presented as follows:

$$S = \sum_{l=0}^{L-1} p\,(x_l)[Q_m(x_l) - f(x_l)]^2 = min. \tag{8}$$

To solve it, we need to calculate the coefficients $\mathrm{b_i}(\mathrm{i} = 0, 1, \dots, \mathrm{m})$ according to the formula

$$b_i = \sum_{l=0}^{L-1} f(x_l) p_i^*(x_l)\,. \tag{9}$$

The approximation polynomial has the form

$$Q_m(x) = \sum_{i=0}^{m} b_i p_i(x)\,. \tag{10}$$

Discrete orthonormal Chebyshev polynomials defined on the interval $[0, \mathrm{t}]$ are given by the formula

$$\hat{p}_i(L, l) = \frac{\sqrt{(2i+1)(L-1)^{[i]}}}{(L+1)^{[i+1]}} \sum_{v=0}^{i} l_{i,v} \frac{l^{[v]}}{(L-1)^{[v]}}\,,$$

$$i = 0, 1, \dots, L-1;\ l = 0, 1, \dots, L-1;\ L = 2, 3, 4, \dots,$$

where $s^{[v]} = s(s-1)\dots(s-v+1)$; $l_{i,v} = (-1)^{i-v} C_{i+v}^{i} C_{i}^{i-v}$.

For point approximation, it is convenient to assume that Chebyshev discrete polynomials are given on the system of points $M = \{x_0 = 0,\ x_1,\ x_2, \dots, x_{L-1}\}$ of the interval $[0, t]$ with the constant step $h = x_{i+1} - x_i$. They are orthonormal on this system of points with the weight $\rho(L, l) = 1$. For such discrete orthonormal Chebyshev polynomials, the recurrence formula holds [3, 5–7, 11, 13, 14]

$$\underset{*_{i+1}}{\hat{p}}(L, l) = \frac{1}{i+1} \sqrt{\frac{(2i+1)(2i+3)}{(L-i-1)(L+i+1)}} \times$$

$$\times \left\{ \sqrt{\frac{L(L^2-1)}{3}}\, \underset{*_1}{\hat{p}}(L, l) \underset{*_i}{\hat{p}}(L, l) - i \sqrt{\frac{(L+i)(L-i)}{(2i+1)(2i-1)}}\, \underset{*_{i-1}}{\hat{p}}(L, l) \right\}, \tag{11}$$

$$i = 0, 1, 2, \dots, L-1\,,$$

which allows us to find all Chebyshev polynomials by means of the first two polynomials

$$\underset{*_0}{\hat{p}}(L, l) = \sqrt{\frac{1}{L}}\,;\ \underset{*_1}{\hat{p}}(L, l) = \sqrt{\frac{3(L-1)}{(L+1)L}} \left(\frac{2l}{L-1} - 1 \right). \tag{12}$$

Since $\mathrm{l} = (\mathrm{x} - \mathrm{x_0})/\mathrm{h}$ and given that the Chebyshev polynomials $\mathrm{x_0} = 0$, $\mathrm{h} = \mathrm{t}/\mathrm{L}$, we find the approximation formula for the Chebyshev polynomial

$$Q_m(x) = \sum_{i=0}^{m} b_i \underset{*_i}{\hat{p}} \left(L, \frac{xL}{t} \right) \text{ при } m \le L-1\,. \tag{13}$$

The physical meaning of optimization of the choice of the degree of an approximating polynomial can be explained as follows: as the degree of Chebyshev polynomial increases, the method error of approximation is reduced to the degree that corresponds to the current process of approximation. With a further increase in the degree of the polynomial, it begins to describe random errors rather than the average value of the approximated process, therefore a random error increases, which leads to a decrease in the quality of the approximation. That is, it is necessary to determine the optimal degree of the approximating polynomial.

Based on the formulae obtained, the process of constructing the regression model can be presented in a step by step form (Fig. 1).

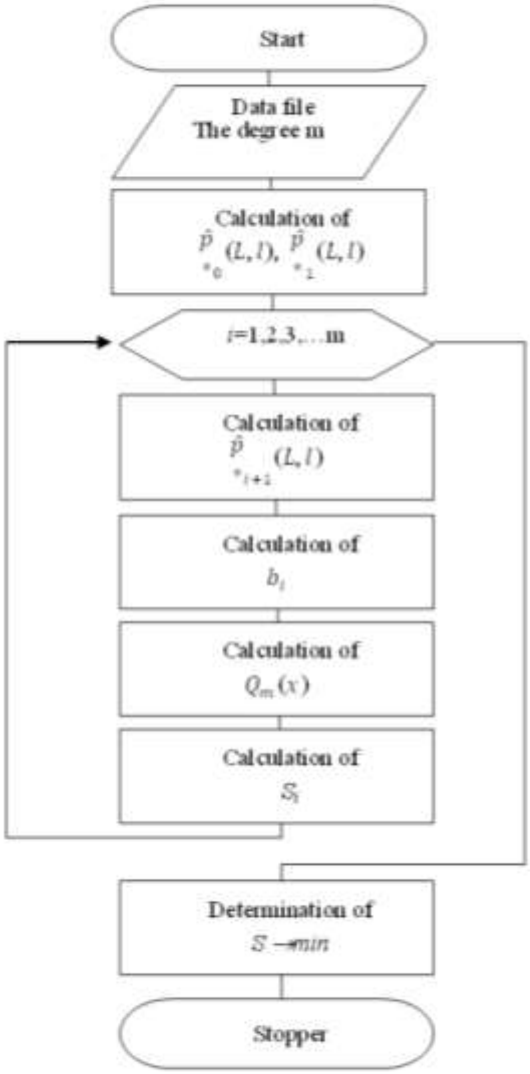

Fig. 1. The process of constructing the regression model

Step 1. The input data and the maximal degree m of Chebyshev polynomial are inputted. The input data should be taken at regular intervals and, for better prediction, it is advisable that the condition $m \ll L - 1$ holds.

Step 2. Chebyshev polynomials of the 0 and 1 orders are calculated according to the formulae (12).

Step 3. With the help of the loop operator, a recurrent procedure is implemented, in which to reach the highest degree m of the polynomial, the following occurs:

- by the formulae (11), the remaining Chebyshev polynomials (till the degree m) are calculated;
- the coefficients b_i of the polynomial according to the formulae (9) are calculated;
- the approximating polynomial (13) is formed.

Step 4. For each degree of the polynomial, the sum of squared differences of the input data and the approximated values (2) is calculated.

Step 5. The minimal value of S, which is the optimal order of the model (8) is determined.

Step 6. The coefficients of the polynomial model with the optimal value of S are chosen as the basis and predictive algorithms are built based on it.

4 Results and Discussions

Verification of the correct functioning of the developed method has been carried out on the problem of prediction of the sales of mineral water (1.5-liter bottles) in a supermarket using Mathcad and Python software (Fig. 2) [2, 6, 8, 11, 14, 18].

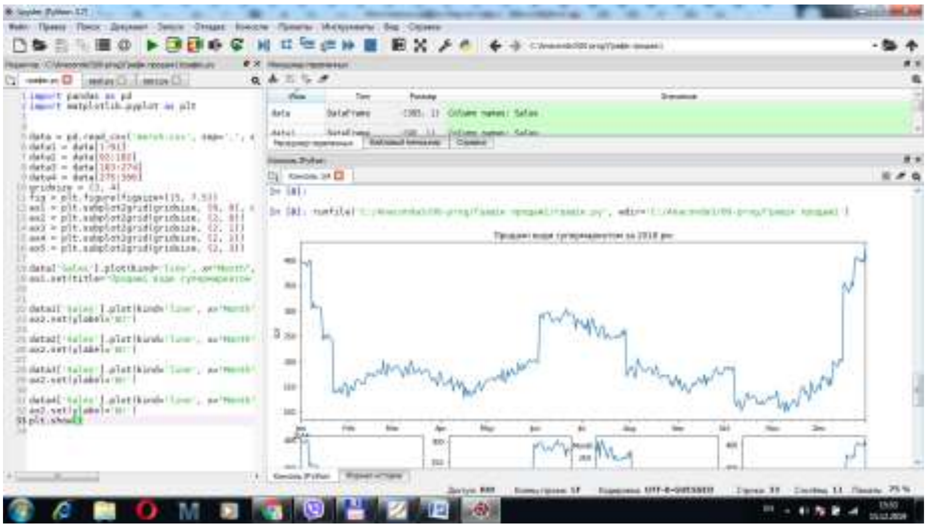

Fig. 2. Software development in Python using IDE Spyder

The machine learning was conducted according to the statistics for 2016, 2017, 2018 as training samples, and the incomplete statistics for 2019 were used as validation data. For example, in Fig. 3, the data of water sales for each day of 2016 and the obtained approximation by means of the developed method (optimal 10-th degree polynomial) are shown.

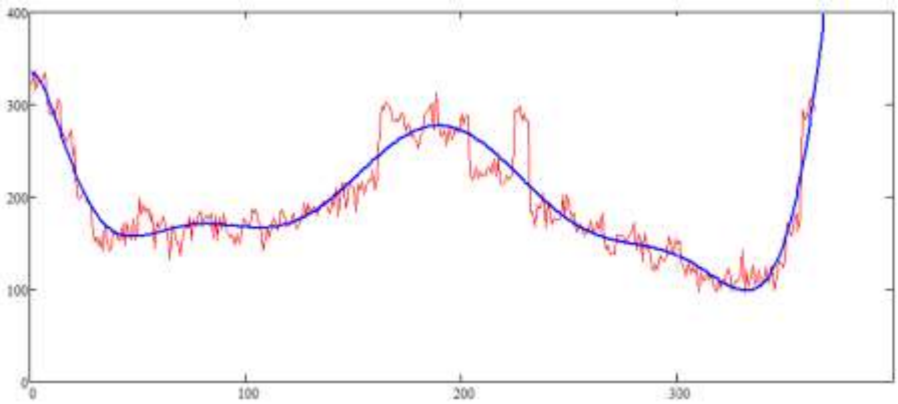

Fig. 3. Approximation of data by Chebyshev 10-th degree polynomial

The refinement of the polynomial coefficients according to the data for 2017 and 2018 has provided us with an opportunity to build an optimal model for the prediction of mineral water sales for 2019 (Fig. 4).

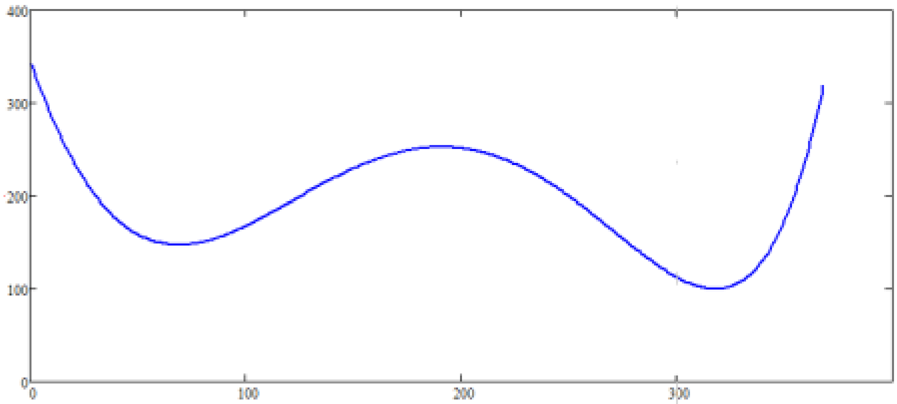

Fig. 4. The optimal model for prediction of mineral water sales for 2019

The prediction of water sales carried out for 2019 with continuous training is compared to the partial data available at that time (Fig. 5).

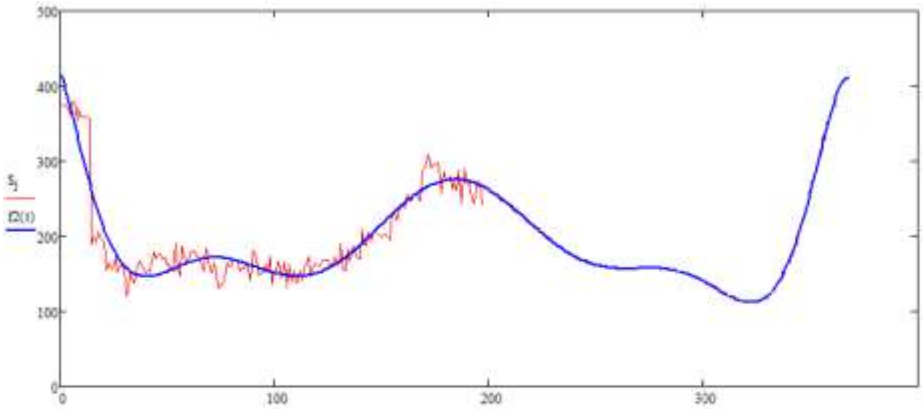

Fig. 5. Refined model for prediction of mineral water sales for 2019 with continuous training on available data

In Fig. 6, a graph of the prediction errors for the available period of data for 2019 is shown. It is seen that over the entire period, except for the first 20 days, the forecast is adequate and the prediction errors are small.

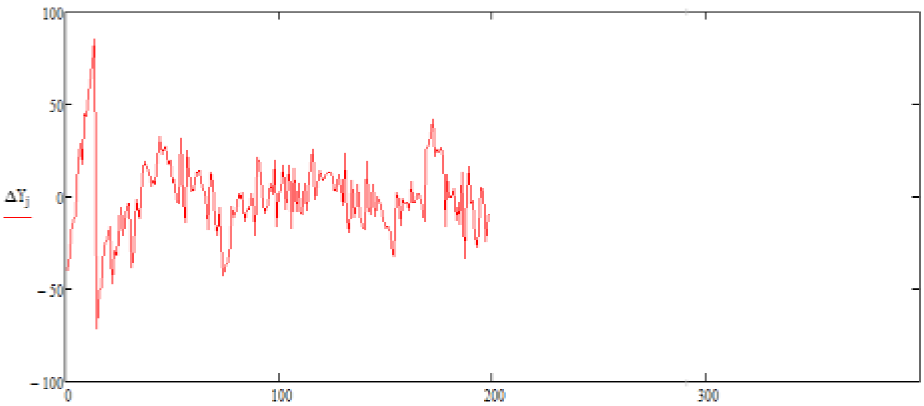

Fig. 6. Error in prediction of mineral water sales for 2019 for the model with continuous training

As can be seen from the graphs presented, the prediction model for the period of available data adequately reflects the state of sales.

As for the first 20 days, here the prediction proved to be not accurate. This is due to the great release of water sales at the beginning of the year, which is, probably, related to the New Year holidays.

5 Conclusions

The use of discrete orthogonal polynomials of Chebyshev for machine learning methods has been theoretically substantiated. The conducted investigations on the application of this method to the prediction of mineral water sales have confirmed the adequacy of the proposed model of machine learning and allow the effective formation of regression models for the predicting complex processes.

In future, it is planned to carry out a study to determine the optimal interval for approximation by means of Chebyshev polynomials and its effect on the terms of prediction. It is also advisable to consider multivariate regression analysis.

References

1. Armstrong J.: Illusions in regression analysis. International Journal of Forecasting, **28**, 689–694 (2012)
2. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. **35**, 1798–1828 (2013)
3. Brandt, S.: Data analysis: statistical and computational methods for scientists and engineers. Springer, Basel (2014)
4. Bzdok, D., Altman, N., Krzywinski, M.: Statistics versus Machine Learning. Nature Methods. 15, 233–234 (2018)
5. Cumming, G.: Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. Routledge, New York (2012)
6. Draper, N., Smith, H.: Applied Regression Analysis (3rd edition). John Wiley & Sons, New York (1998)
7. Duda, R., Hart, P., Stork, D.: Pattern Classification (2nd edition). Wiley, New York (2001)
8. Flach, P.: Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, Cambridge (2012)
9. Hastie, T., Tibshirani R., Friedman J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition). Springer-Verlag, New York (2009)
10. Long, Y.: Human age estimation by metric learning for regression problems. In International Conference on Computer Graphics, Imaging and Visualization, pp. 74–82. (2009)
11. Muller, A., Guido, S.: Introduction to machine learning with python: a guide for data scientists. O’Reilly Media, Sebastopol (2017)
12. Ord, K.: The illusion of predictability: A call to action. International Journal of Forecasting, **28**, 717–718 (2012)
13. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach (3rd edition). Prentice Hall, New Jersey (2010)
14. Rybin, V.: Priblizheniye funktsiy. Komp'yuternaya praktika v sisteme komp'yuternoy matematiki MATHCAD. MAI, Moscow (2004)

15. Sen, A., Srivastava, M.: Regression Analysis — Theory, Methods, and Applications (4th edition). Springer-Verlag, Berlin (2011)
16. Soyer, E., Hogarth, R.: The illusion of predictability: How regression statistics mislead experts. International Journal of Forecasting. **28**, 695–711 (2012)
17. Suetin, P.: Classical orthogonal polynomials (3rd edition). FIZMATLIT, Moscow (2005)
18. VanderPlas, J.: Python Data Science Handbook. O'Reilly Media, Sebastopol (2016)
19. Vapnik, V.: Statistical Learning Theory. John Wiley & Sons, New York (1998)