Kun-Chou Lee, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 10, Issue 7, (Series-II) July 2020, pp. 32-34

RESEARCH ARTICLE OPEN ACCESS

Prediction of Complicated Mathematical Problems by Machine Learning of KNN Regression

Kun-Chou Lee[1]*,Chieh-Hsin Lee[2]
[1]*Department of Systems and Naval Machatronic Engineering, National Cheng-Kung University, Tainan, Taiwan.*
[2]*Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei, Taiwan*
** Corresponding author: Kun-Chou Lee*

ABSTRACT
Under some circumstances, an experiment may be difficult to implement, or a mathematical formula may be difficult to compute. This then motivated us to develop a predicting model for dealing with such difficult problems. In this paper, the prediction of complicated mathematical problems is given by machine learning of KNN (K nearest neighbors) regression. The KNN algorithm belongs to machine learning, which plays an important role in artificial intelligence. It can be applied to both classification and regression. This study utilizes it for regression. The procedures are divided into two stages, which are learning and predicting. As the KNN regression is well trained, it can predict results of complicated mathematical problems accurately.
***Keywords*–** machine learning, K nearest neighbors, classification, regression, mathematics.

Date of Submission: 18-06-2020 Date of Acceptance: 06-07-2020

I. INTRODUCTIONS

Recently, the artificial intelligence has been widespread all over the world. Machine learning [1] belongs to artificial intelligence. It is usually utilized to model the relationship between input and output of a nonlinear system. The nonlinear system may be a difficult experiment, a complicated theoretical computation, …, etc. Under some circumstances, an experiment may be difficult to implement, or a mathematical formula may be difficult to compute. This then motivated us to develop a predicting model for dealing with such difficult problems.

In this paper, the machine learning of KNN (K nearest neighbors) [2-7] is applied to the prediction of a complicated mathematical function and its definite integral. The KNN algorithm is a non-parametric and supervised method used for both classification [2-5] and regression [6-7].It utilizes the *K* closest training samples to provide information for predicting the property value of an unknown sample.Since the KNN regression is basically a supervised machine learning algorithm, it is divided into two stages, which are learning and predicting. Numerical simulation shows that our application of KNN regression to complicated mathematical problems is not only accurate, but also efficient. Note that the complicated mathematical problem in this study is only an illustration example. There exists no limitation on the types of problems.

II. KNN REGRESSIONS

The KNN algorithm is a very simple machine learning algorithm, which is initially proposed by Cover & Hart [2] for classification. It is a supervised algorithm. To illustrate the classification, a two dimensional KNN example with two categories is given in Figure 1. In Figure 1, there are samples from two categories, which are blue triangles (label #1) and red rectangles (label #2). As an unknown object"green circle" appears, we want to know which category the new object might fall into. In KNN algorithm, the judging principle is to count which category has more members among the considered neighbors. The "K" means the total number of considered neighbors. For example, as the consideredsamplesare bounded by the solid circle line in Figure 1, the category #2 has more members than category #1 (2:1, K=3). Thus the unknown object is judged to belong to category #2. As the considered samples are bounded by the dash circle line in Figure 1, the category #1 has more members than category #2 (3:2, K=5). Thus the unknown object is judged to belong to category #1.

The above KNN classification can be modified and extended to deal with regression. As the KNN is for regression, in Figure 1, each sample has a continuous real number to represent its property value. As an unknown object "green circle" appears, we want to predict the property value of this new object. Assume there are totally N existing samples in Figure 1. Each sample has a location vector $\bar{r}_i$ and property value β_i, where i = 1, 2, ..., N. The unknown object has a location vector $\bar{r}$ and we want to predict its property value β. In K-NN regression, we first choose the value of K, i.e., the number of nearest neighbors. Next, the Euclidean distance between the unknown object and existing samplesis computed and denoted as $d(\bar{r}, \bar{r}_k)$, k = 1, 2, ..., K. The property value for the unknown object is predicted as

$$\beta = \sum_{k=1}^{K} w_k \beta_k \qquad (1)$$

where

$$w_k = \frac{1/d(\bar{r}, \bar{r}_k)}{\sum_{k=1}^{K} [1/d(\bar{r}, \bar{r}_k)]} \qquad (2)$$

The above formulas hint that a near sample has great influence on the unknown object prediction, whereas a far sample has only little impact.

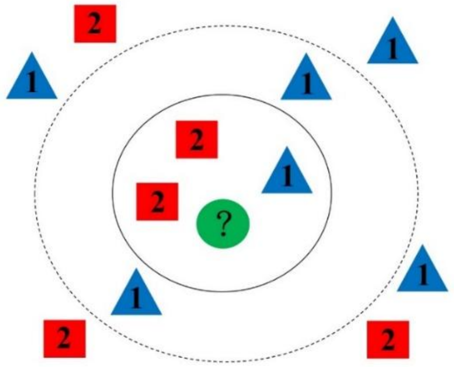

Figure 1. Two dimensional illustration of the KNN algorithm with two categories.

III. NUMERICAL EXAMPLE

In this section, two numerical examples are given to illustrate the above KNN regression. The input and output may be any theoretical or experimental data. For simplicity, a complicated mathematical function and its integration are utilized to generate the training and testing data.

In the first example, the function

$$f(x) = \left[J_0\left(\frac{x}{2}\right)\right]^2 \cdot \exp\left(\frac{x^2}{50}\right) \cdot \ln(x+1) \qquad (3)$$

is considered, where $J_0(\cdot)$is the Bessel function of the first kind with order zero and $\ln(\cdot)$ is the natural logarithm. Obviously, equation (3) is verycomplicated to compute. Sowe are motivated to construct a machine learning model for predicting the function values. Initially, the function values of equation (3) for x = 0, 0.1, 0.2, …, 10, are numerically computed. The number of the whole data sets is 101. Among the 101 data sets, we randomly select 80of them for training the KNN regression. The remainder 21 data sets are for testing.In KNN algorithm, the number of nearest neighbors, i.e., K, is chosen as 5. Figure 2 shows the function value of equation (3) for the 21 testing data sets by theoretical computation (i.e., answer) and KNN prediction, respectively.It shows that they are in good agreement.

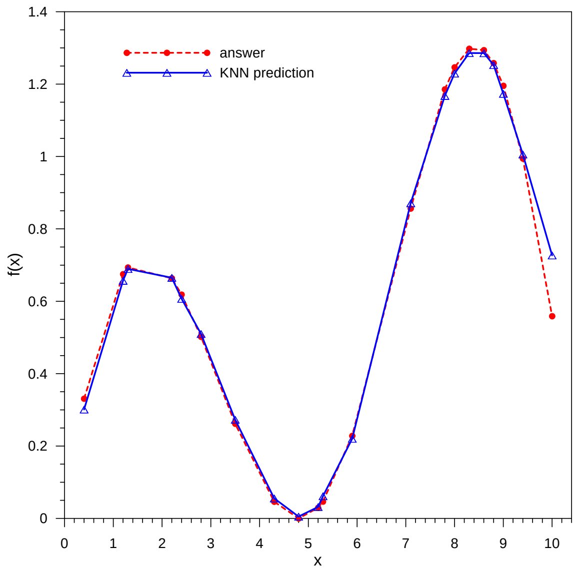

Figure 2. The function value of equation (3) for the 21 testing data setsby theoretical computation (i.e., answer) and KNN prediction, respectively.

In the second example, the definite integral

$$\int_0^d f(x)dx \qquad (4)$$

is considered. The $f(x)$ of equation (4) is defined in equation (3). Obviously, the above definite integral is very complicated to compute. Since the integral has no analytical solution, we are motivated to construct a machine learning model for predicting the integral results. Initially, the integration ofequation (4) are numerically computedfor d = 0, 0.1, 0.2, …, 10, respectively. The number of the whole data sets is 101. Among the 101 data sets, we randomly select 80 of them for training the KNN regression. The remainder 21 data sets are for testing. Figure 3 shows the definite integral of equation (4) for the 21 testing data sets (i.e., different values of integral bound d) by numerical computation (i.e., answer) and KNN prediction, respectively. It shows that they are in good agreement.

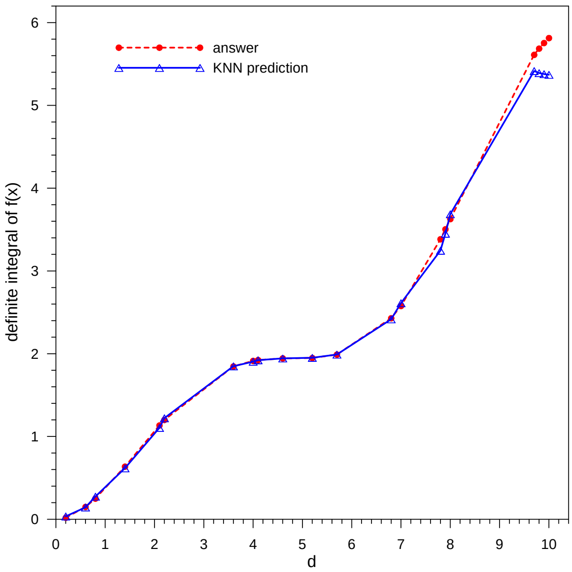

Figure 3. The definite integral of equation (4) for the 21 testing data sets (i.e., different values of integral bound *d*) by numerical computation (i.e., answer) and KNN prediction, respectively.

From Figure 2 and Figure 3, compared with most parts of the curves, the KNN prediction near the right endpoint has a slight mismatch. This may be because the KNN near the right endpoint has no sufficient neighbors to help prediction. The above numerical simulation is performed by using the Python programming language. The computer CPU is Intel(R) Core™ i7-4790 with 3.60 GHz.

IV. Conclusion

This paper has successfully utilized the KNN regression to model a complicated mathematical function and its definite integration. Note that there exists no limitation on the types of data or problems. The KNN is a supervised algorithm, it requires some known data in advance. The data may be obtained from difficult experiments or complicated theoretical computations.The KNN regression is inherently a black box. It can model a strongly nonlinear system, even though the system characteristic does not exist. The research flow chart of this study can be extended to deal with many complicated engineering and scientific problems.

ACKNOWLEDGEMENTS

The authors would like to acknowledge (1) the financial support of the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2221-E-006-091, (2) the National Center for High Performance Computing, Taiwan, for computer time and facilities.

REFERENCES

[1] S. Raschka, *Python machine learning*(Birmingham, UK: Packt Publishing Ltd., 2016).

[2] T. M. Cover and P. E. Hart, Nearest neighborpattern classification,*IEEE Transactions on Information Theory, 13(1)*,1967, 21-27.

[3] P. E. Hart, The condensed nearest neighbor rule, *IEEE Transactions on Information Theory*, *14(3)*, 1968, 515-516.

[4] P. Hall, B. U. Park and R. J. Samworth, Choice ofneighbor order in nearest neighbor classification,*Annals of Statistics*, *36(5)*,2008, 2135-2152.

[5] I. Mesecan and I. Bucak, Searching the effects ofimage scaling for underground object detectionusing K-means and K-NN,*Proc. The 8th EuropeanModelling Symposium (EMS)*, Pisa, Italy, 2014, 180-184.

[6] N. S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression,*Annals of Statistics*,46(3), 1992, 175-185.

[7] K. C. Lee, Radar target recognition by machine learning of k-nearest neighbors regression on angular diversity RCS,*ACES Journal*, 34(1), 2019, 75-81.