Optimal Solution of Classification (Prediction)Problem

Mohammad S. Khrisat

Computer Engineering Department, Balqa Applied University, Jordan

Summary
Classification or prediction problem is how to solve it using a specific feature to obtain the predicted class. A wheat seeds specifications 4 3 classes of seeds will be used in a prediction process. A multi linear regression will be built, and a prediction error ratio will be calculated. To enhance the prediction ratio an ANN model will be built and trained. The obtained results will be examined to show how to make a prediction tool capable to compute a predicted class number very close to the target class number.

Keywords:
Classification, prediction, ANN, FFANN, training, MLR, prediction error ratio, features, regression coefficients, regression equation .

1. Introduction

As you Classification is the study of methods that are used to categorize data based on distinct classes. We might call classes targets, labels, or categories. Categorization can be done by three or more methods. These methods can be listed as follows. Distinct labeling of data or what is known as supervised learning. Data division into classes as in unsupervised learning. Distinguished features selection, and combinations of these stated methods

Based on the input of features or patterns [4], [5], [6]. The classifier can be used to distinguish the inputted data by performing needed actions to output a predicted class. Thereafter it's possible to use the calculated classifier to implement an action [7], [8], [9]. In classification without the label, the data is inputted to the model, the model should return a class in a specific place [10], [11]. There are two forms of data mining that can be used to extract models. And these forms are classification descriptions and predication. To predict future data direction, the following is of concern:-

1- Find the missing elements in the datasheet.
2- Predicting the outcome by the classification model.
3- No dependence on the label of the class
4- Predication is based on both the label and the class model.

1.1 Classification by Regression

In general, classification involves the prediction of labeling, while regression involves the prediction of quantity. Multiple Linear Regression is an estimator for the relationship between two independent variables or several variables, used as an input, and a single dependent variable used as an output. Input variable can be categorical, which contain a finite number of categories or distinct groups. Also, the input variables could be continuous variables.

To describe relationship between variables, we usually fit a line in the observed data. This will allow us to estimate how the output variable changes in relation to the inputted variables. To perform multiple line regression, we must use a formula as shown in figure (1) where y is the observed dependent variable [4], [5].

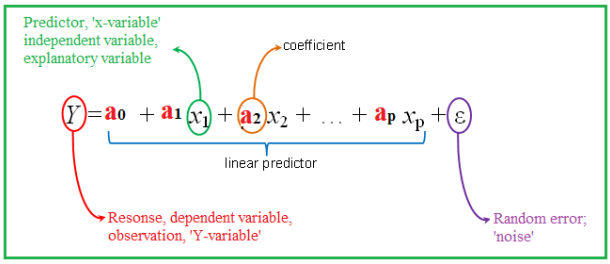

Figure 1: MLR model

To find a class value using MLR we have to follow up the following phases (see figure 2)

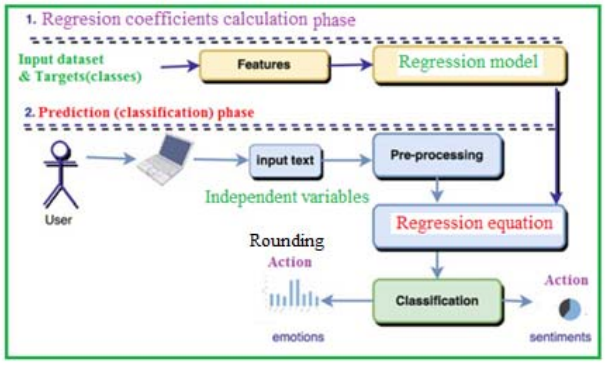

Figure 2: Classification using MLR model

Phase 1:

Use the input dataset(features) to build MLR model, then apply this model to obtain the regression coeficients.

Phase 2:

The following steps were used to implement this phase:

- Obtain the features values(independent variables values).

Manuscript received September 5, 2023
Manuscript revised September 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.9.16

- Use the regression coefficients to build the regression equation.
- Apply the regression equation to find the class value.
- Round the result to get the class value.

1.2 Classification using artificial neural networks

Artificial neural network (ANN) [16], [17]is a powerfull computational model used in various fetal applications such as curve fitting(regresion analysiss), pattern recognition(classification) , clustering and and time series. One variant of ANN is a feed forword ANN(FFANN)[18], [19]. FFANN is a set of fully connected neurons, arranged in layers as shown in figure 3, each neuron acts as a computational cell and performs two main functions as shown in figure 4. The first fuctions is sumation of products of the inputs and the associated weights, the second function is computing the neuron outputs depnding on the activation function (logsig, tansig or linear) selected for the neuron within a specified layer[20], [21].

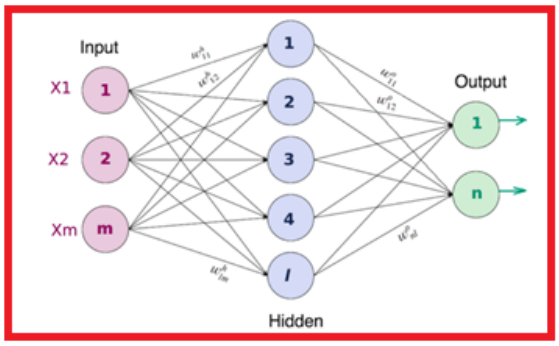

Figure 3: FFANN architecture example

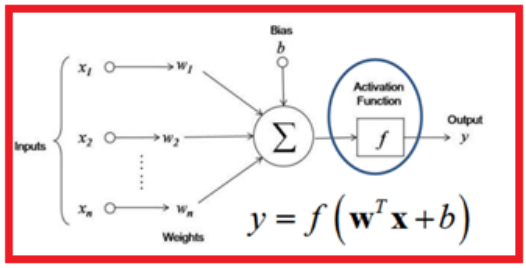

Figure 4: Neuron operations

To use FFANN as a classification tool we have to follow up the following procedures (see figure 5)[22], [23], [24]:

Phase 1: Training

In this phase we have to apply the following steps:

1. Select the input dataset(features) and targets (classes values), if required normalize the data(preproccesing).
2. Create FFANN (select FFANN architecture) by the definition of
   a) number of layers,
   b) Number of neurons in each indivisual layer
   c) The function of activation for all layers.
3. Initialize FFANN.
4. Define some parametors for the net, such as the gaol (the error between the target and the calculated output must equal or closed to zero), the number of training cycles(epochs).
5. Train the net using the features and classes.
6. Chech the error value, if the error is acceptable save the net to be used later as a a recognition tool, else increase the number of training cycle or adjust the net architecture and train it again.

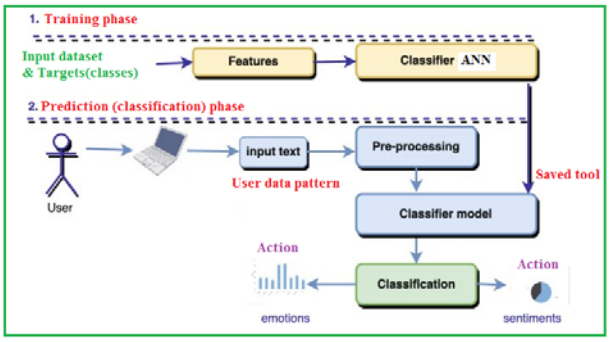

Figure 5: Using ANN as a classifier

Each training cycle contains two phases as shown if figures 6 and 7. The feedforward phase by calculating the neurons outputs starting from the input layer, and backword phase staring from the ouput layer to adjust the neurons weight.

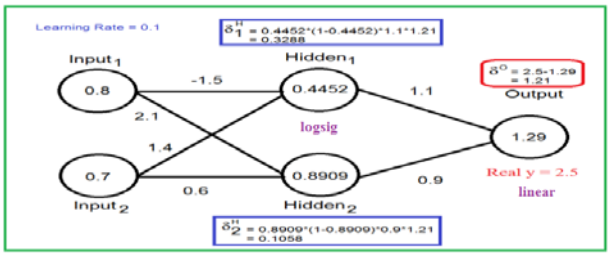

Figure 6: Calculating the outputs (Forward phase)

2. Implementation and experimental results

The wheat seeds dataset was used as an input to the created models to generate classification. When the dataset for the seeds includes the prediction of species. The measurement of seeds was given from different kinds of wheat. This was a two-class classification binary case. The observation for the individual classes was balanced. There exist 210 observations with seven input variables and one output variable. The variables names are as follows:

1- Area
2- Compactness
3- Perimeter
4- Length of kernel
5- Width of kernel
6- Asymmetric Coefficient
7- Length kernel groove
8- Class (1,2,3).

Nine samples are shown below in figure 7.

Word or equivalent Word Processors and justify to block. The heading of each section should be printed in small, 12pt, left justified, bold, serif. You must use the Arabic numbers 1, 2, 3, for the sections numbering and not the Roman numbers (I, II, III). Please, follow the paragraph indentation that is used in this template.

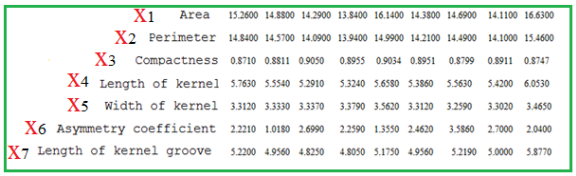

Figure 7: Some samples of the input dataset

First we apply MLR , the coeffecients of the regerssion outputs are shown in table 1:

Table 1: Regression coefficients

Coefficient	Value
a0	53.4436
a1	1.4891
a2	-3.2204
a3	-30.6774
a4	-2.3151
a5	0.2460
a6	0.1149
a7	2.1926

Applying the regression equation using the obtained regression coefficients we found the predicted values of the classes, rounding these values we can obtain the predicted class number for any given values of the independent variables. From the used 210 samples, 36 classes were wrong calculated, with prediction error ratio equal 17.14 %., figure 8 shows the target and the predicted classes, while figure 9 shows the error between them (before rounding).

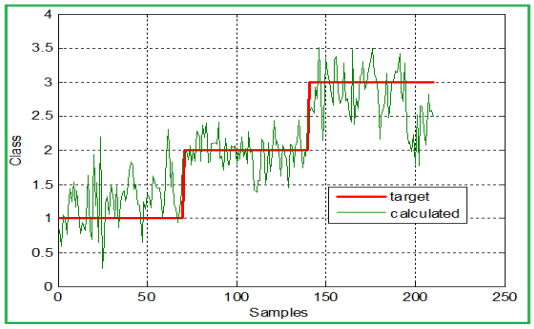

Figure 8: Targets and predicted classes (regression model)

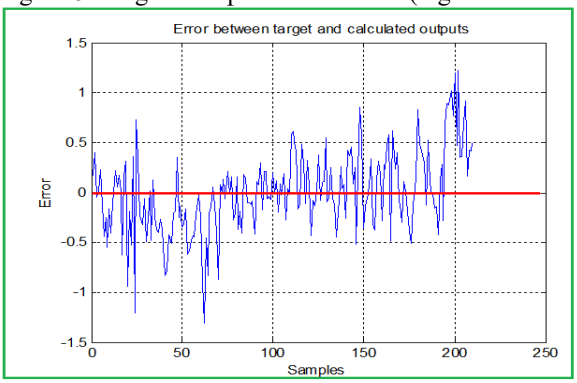

Figure 9: Error between the target and predicted classes(regression model)

Second we apply ANN mode:

ANN was created using the follwong features:

- Number of inputs 7.
- Number of outputs 1.
- ANN contains 2 layers, the first layer with 7 neurons and tansig activation function, the second layer with 1 neuron and linear activation function.
- The goal (error) was set to zero.
- The number of trainging cycles was set to 2000.

ANN was trained using the same data set and targets, the trained ANN was used to predict the class number using a specific features, here the max error obtained was equal 1.7979e-005, which means that the prediction ratio is very closed to 100%, figure 10 shows the target and the predicted classes, while figure 11 shows the error between them(befor rounding).

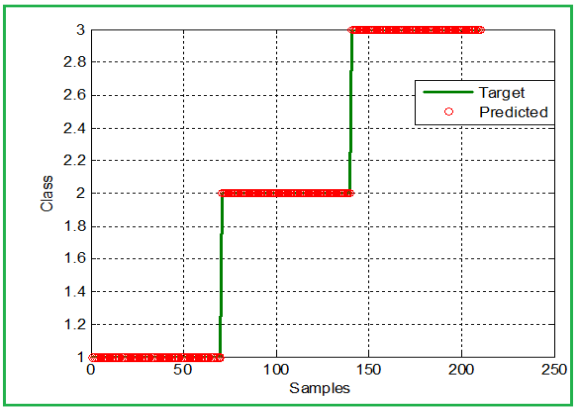

Figure 10: Targets and predicted classes(ANN model)

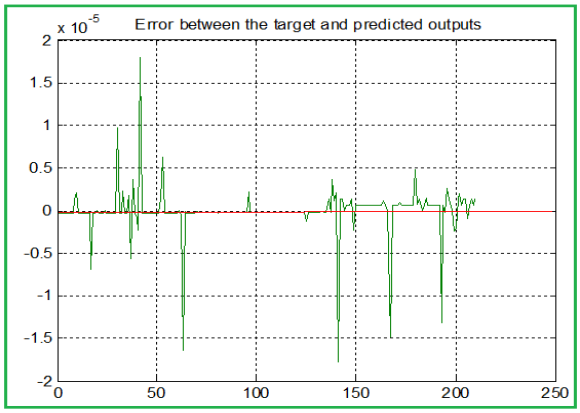

Figure 11: Error between the target and predicted classes(ANN model)

From the obtained experimental results, we can raise the following facts:

- MLR model is restricted to a specific prediction problem.
- Using MLR model for prediction causes a high regression error ratio.
- To rapidly decrease the prediction error ratio, we recommend using ANN model.
- Building ANN model si a very simple process, this model can be easily updated to suit any input data set with any target classes.

3. Conclusion

A wheat seeds dataset with various specifications for 3 classes of wheats were used as an input data set for a classification problem, a MLR model was built and implemented to calculate the wheat class number depending on a selected specification, the obtained results was acceptable but with prediction error equal 17.14 %. To minimize the prediction error ration, and to obtain a predicted class number closed to the target class, we recommend using ANN model. The ANN model decreases rapidly the prediction error, it is very simple to build and update.

References

[1] Jamil Al-Azzeh Naseem Asad, Ziad Alqadi, Ismail Shayeb, Qazem Jaber, Jamil Al-Azzeh, Simple Procedures to Create HSCS, International Journal of Engineering Research And Management (IJERM), vol. 7, issue 5, pp. 6-10, 2020.

[2] Aws Al-Qaisi, A Manasreh, A Sharadqeh, Z Alqadi, Digital color image classification based on modified local binary pattern using neural network, International Journal on Communications Antenna and Propagation (I. Re. CAP), vol. 9, issue 6, pp. 403-408, 2019.

[3] Dr. Mohammad S. Khrisat Prof. Ziad Alqadi, COLOR IMAGES CLASSIFIER OPTIMIZATION, International Journal of Engineering Technology Research & Management, vol. 5, issue 3, pp. 6-14, 2021.

[4] Abdullah Al-Hasanat, Haitham Alasha'ary, Khaled Matrouk, Ziad Al-Qadi, Hasan Al-Shalabi, Experimental Investigation of Training Algorithms used in Back propagation Artificial Neural Networks to Apply Curve Fitting, European Journal of Scientific Research, vol. 121, issue 4, pp. 328-335, 2014.

[5] Belal Ayyoub, Ahmad Sharadqh, Ziad Alqadi, Jamil Al-azzeh, Simulink based RNN models to solve LPM, International Journal of Research in Advanced Engineering and Technology, vol. 5, issue 1, pp. 49-55, 2019.

[6] AlQaisi Aws, AlTarawneh Mokhled, A Alqadi Ziad, A Sharadqah Ahmad, Analysis of Color Image Features Extraction using Texture Methods, TELKOMNIKA, vol. 17, issue 3, 2018.

[7] Mohammed Ashraf Al Zudool, Saleh Khawatreh, Ziad A. Alqadi, Efficient Methods used to Extract Color Image Features, IJCSMC, vol. 6, issue 12, pp. 7-14, 2017.

[8] Bilal Zahran Belal Ayyoub, Jihad Nader, Ziad Al-Qadi, Suggested Method to Create Color Image Features Victor, Journal of Engineering and Applied Sciences, vol. 14, issue 1, pp. 2203-2207, 2019.

[9] Ziad AlQadi, M Elsayyed Hussein, Window Averaging Method to Create a Feature Victor for RGB Color Image, International Journal of Computer Science and Mobile Computing, vol. 6, issue 2, pp. 60-66, 2017.

[10] Dr Rushdi S Abu Zneit, Dr Ziad AlQadi, Dr Mohammad Abu Zalata, A Methodology to Create a Fingerprint for RGB Color Image, IJCSMC, vol. 6, issue 1, pp. 205-212, 2017.

[11] Ahmad Sharadqh Naseem Asad, Ismail Shayeb, Qazem Jaber, Belal Ayyoub, Ziad Alqadi, Creating a Stable and Fixed Features Array for Digital Color Image, IJCSMC, vol. 8, issue 8, pp. 50-56, 2019.

[12] Ahmad Sharadqh Jamil Al-Azzeh, Rashad Rasras, Ziad Alqadi, Belal Ayyoub, Adaptation of matlab K-means clustering function to create Color Image Features, International Journal of Research in Advanced Engineering and Technology, vol. 5, issue 2, pp. 10-18, 2019.

[13] ZIAD ALQADI, A MODIFIED LBP METHOD TO EXTRACT FEATURES FROM COLOR IMAGES, Journal of Theoretical and Applied Information Technology, vol. 96, issue 10, pp. 3014-3024, 2018.

[14] Aws Al-Qaisi, Saleh A Khawatreh, Ahmad A Sharadqah, Ziad A Alqadi, Wave file features extraction using reduced LBP, International Journal of Electrical and Computer Engineering, vol. 8, issue 5, 2018.

[15] Ahmad Sharadqh Naseem Asad, Ismail Shayeb, Qazem Jaber, Belal Ayyoub, Ziad Alqadi, Creating a Stable and Fixed Features Array for Digital Color Image, IJCSMC, vol. 8, issue 8, pp. 50-56, 2019.

[16] Ziad AlQadi, Yehya Abded Allatif, Musbah J Aqel, A Proposed methodology for image objects recognition using Artificial neural networks, IJCSS, Vol.3, No.1, pp. 49-56, 2011

[17] Akram A Moustafa, Ziad A Alqadi, Eyad A Shahroury, Performance evaluation of artificial neural networks for spatial data analysis, WSEAS Transactions on Computers, vol. 10, issue 4, pp. 115-124, 2011.

[18] Jamil Al-Azzeh, Ziad Alqadi, Mohammed Abuzalata, Performance Analysis of Artificial Neural

[19] Networks used for Color Image Recognition and Retrieving, international Journal of Computer Science and Mobile computing, vol. 8, issue 2, pp. 20-33, 2019.

[20] Khaled M Matrouk, Haitham A Alasha'ary, Abdullah I Al-Hasanat, Ziad A Al-Qadi, Hasan M Al-Shalabi, Investigation and Analysis of ANN Parameters, European Journal of Scientific Research, vol. 121, issue 2, pp. 217-225, 2014.

[21] Ziad A AlQadi Amjad Y Hindi, O Dwairi Majed, PROCEDURES FOR SPEECH RECOGNITION USING LPC AND ANN, International Journal of Engineering Technology Research & Management, vol. 4, issue 2, pp. 48-55, 2020.

[22] Dr. Amjad Hindi, Dr. Majed Omar Dwairi, Prof. Ziad Alqadi, Analysis of Procedures used to build an Optimal Fingerprint Recognition System, International Journal of Computer Science and Mobile Computing, vol. 9, issue 2, pp. 21 – 37, 2020.

[23] Prof. Mohammed Abu Zalata, Dr. Ghazi, M. Qaryouti, Dr.Saleh Khawatreh, Prof. Ziad A.A. Alqadi, Optimal Color Image Recognition System (OCIRS), International Journal of Advanced Computer Science and Technology, vol. 7, issue 1, pp. 91-99, 2017.

[24] Hatim Ghazi Zaini, Ziad AlQadi, Analysis of FFANN Used for Pattern Recognition, International Journal of Computer Science and Mobile Computing, vol. 10, issue 3, pp. 55 – 65, 2021.

[25] Prof. Mohammed K, Abu Zalata Dr. Ghazi M. Qaryouti, Prof. Ziad A.A. Alqadi, A Novel Method for Color Image Recognition, International Journal of Computer Science and Mobile Computing, vol. 5, issue 11, pp. 57 – 64, 2016.

Khrisat Mohammad Dr. Eng. degrees from People Friendship of Russia 2017. working as a professor assistant (from 2018 in the Dept. of Computer Engineering, the Balqa Applied University Faculty of Engineering Technology