Logistic Regression

|  |  |  |
| --- | --- | --- |
| Jan Carlo D. Rabacca  11106344 | Suzette Dela Cruz  11118113 | Marc Janssen Chiu  11114525 |

# Introduction

Logistic Regression is a method for analyzing data, it is similar to linear regression but logistic regression covers the problem where the data can have more than true or false (e.g. binary 1 or 0) answers [1].

# . Procedures

1. First, load the training data and create a 2D representation plot (refer to Fig.1).



Fig. 1

1. Initialize variables with the size of the x data then concatenate a column with a value of 1 to the x array for convenience to the formula to be used (refer to Fig.2).



Fig. 2

1. Separate the values 1 and 0 from the y array (refer to Fig.3).

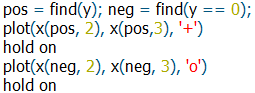
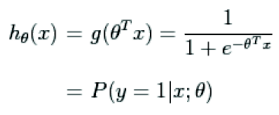


Fig. 3

1. Code the formula for the hypothesis function (refer to Eqn.1).

 (Eqn. 1)

The code should look like this: (refer to Fig.4)



Fig. 4

1. Define the sigmoid to be used in the gradient and hessian formulas (refer to Fig.5).

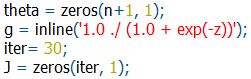


Fig. 5

1. Apply the gradient descent then the hessian function then the cost function (refer to Fig.6).



Fig. 6

1. Plot the result of the Newton’s method then calculate the decision boundary line (refer to Fig.7).

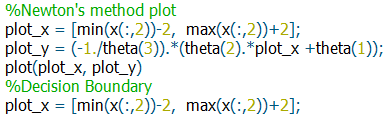


Fig. 7

1. Plot the learning curve (refer to Fig.8).  
   

Fig. 8

The code used for this experiment is similar to the code used on the linear regression experiment as well as with the multivariate experiment in which the cost function was introduced, however, in this experiment, new terms were introduced which are the sigmoid function, Hessian function, etc.

The implementation is the same with the previous experiments but differs in the approach in determining the output or the application of how it the machine would project or predict output data.

# Data and Results

## Procedure 3.1

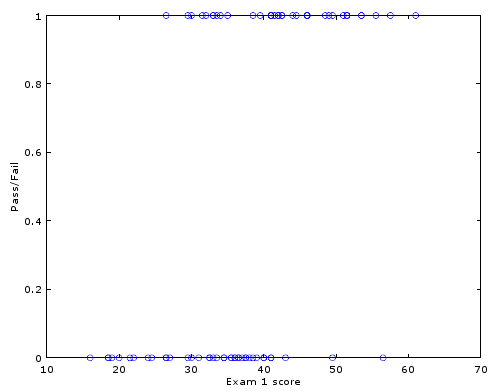


Fig. 9 Exam 1 Scores vs. Y-data

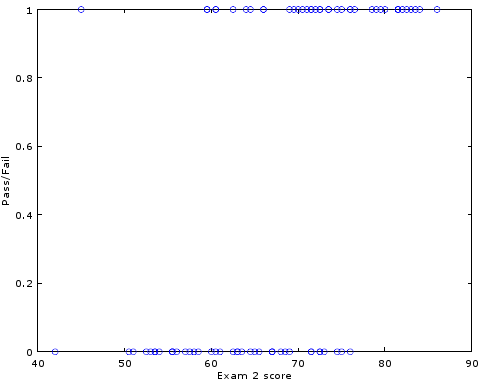


Fig. 10 Exam 2 Scores vs. Y-data

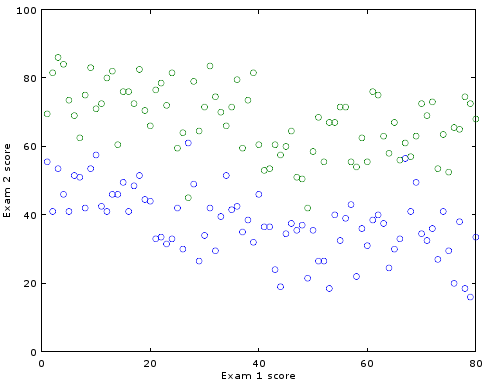


Fig. 11 Exam 1 Score (Blue) vs. Exam 2 Scores (Green)

## Procedure 3.2

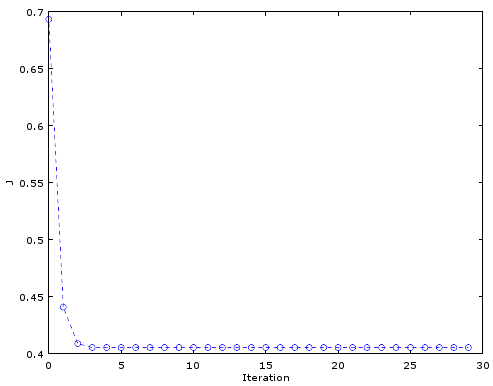


Fig. 12 Iteration vs. J

## Procedure 3.3

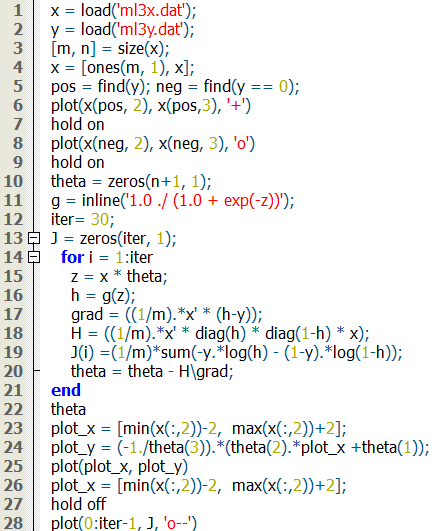
1. ɵ = -16.37874

0.14834

0.15891

1. Iterations = 5
2. Probability = 0.66802 or 67%

Code Used:



## Exercise

The x.dat file contains the data wherein the first column represents the income while the second column represents the residence length, the y.dat file contains the information whether professional or not [3].

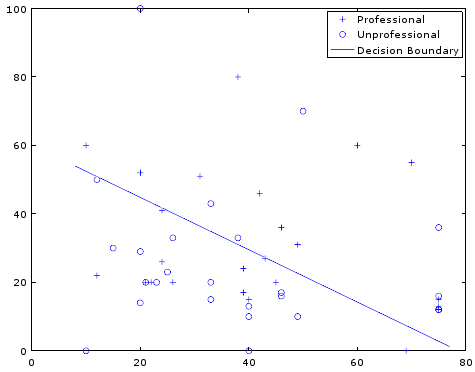


Fig. 13 Decision Boundary Line

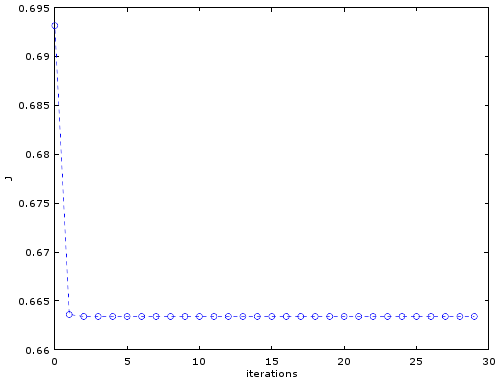


Fig. 14 Learning Curve

# Analysis and conclusion

Logistic Regression is a classification method used to determine or project an outcome based on variables that does not have a corresponding pattern, these variables are often considered as multiclass variables [2] or in some references the dependent variable are called dichotomous [1]. The logistic regression has a property called the decision boundary line, basically it is a straight line that decides the parameter value.

In conclusion, logistic regression is more accurate than linear regression and is more dynamic in terms of predicting or projecting data outcomes based in training data because logistic regression has a capability of considering more parameters rather than making an average of the data values as with the case of linear regression. Overall, logistic regression is a more intelligent method compared to the linear regression approach.

# REFERENCES

1. F. Schoonjans, 'Logistic regression', MedCalc, 2015. [Online]. Available: https://www.medcalc.org/manual/logistic\_regression.php. [Accessed: 22- Sep- 2015].
2. Holehouse.org, '06\_Logistic\_Regression', 2015. [Online]. Available: http://www.holehouse.org/mlclass/06\_Logistic\_Regression.html. [Accessed: 22- Sep- 2015].
3. Logisticregressionanalysis.com, 2015. [Online]. Available: http://logisticregressionanalysis.com/MiscPages/KidCreative.html. [Accessed: 22- Sep- 2015].