Logistic Regression and Newton‘s Method

Aquino, Patrica Rose
Basallote, Lawrence Andrew
Carag, Stephanie
Jacinto, Dan Emanuel
Lunasco, Jan Osbert

***Abstract*—The goal of linear regression is to see whether the probability of getting a particular value of the nominal variable is associated with the measurement variable whereas for logistic regression is to predict the probability of getting a particular value of the nominal variable, given the measurement variable [1]. Through this experiment, the difference between linear and logistic regression will be observed and as well as achieving proper approximation through the use of Newton's Method**

I. Introduction

The aim of this experiment is to implement logistic regression using Newton's Method. In the previous experiment, Linear regression was used because it deals with continuous data. There are some cases that such method is not suited for a certain problem. Logistic regression is a method that used for dataset with one or more independent variable to determine an outcome [2]. For example obesity and age, a person can be obese regardless of age. Being older doesn't mean he/she is more prone to being obese or vice versa. Newton's Method or also known as Newton's iteration is used to minimize the cost function.

II. Procedure

A. Data

Suppose that a high school has a data set representing 40 students who were admitted to college and 40 students who were not admitted.
Each $(x^{(i)}, y^{(i)})$ training example contains a student's score on two standardized exams and a label of whether the student was admitted. In your training data,

1) The first column of your x array represents all Test 1 scores, and the second column represents all Test 2 scores.
2) The y vector uses '1' to label a student who was admitted and '0' to label a student who was not admitted.

Load the data for the training examples into your program and add the x_0 = 1 intercept term into your x matrix. Before beginning Newton's Method, we will first plot the data using different symbols to represent the two classes. In Octave, you can separate the positive class and the negative class using the find command.

```
% find returns the indices of the
% rows meeting the specified condition
pos = find(y == 1); neg = find(y == 0);
% Assume the features are in the 2nd and
3rd
% columns of x
plot(x(pos, 2), x(pos,3), '+'); hold on
plot(x(neg, 2), x(neg, 3), 'o')
```

B. Newton's Method

Recall that in logistic regression, the hypothesis function is:

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}} = P(y=1|x;\theta) \qquad (1)$$

Octave does not have a library function for the sigmoid, so you will have to define it yourself.The easiest way to do this is through an inline expression:

```
% To find the value of the sigmoid
% evaluated at 2, call g(2)
g = inline('1.0 ./ (1.0 + exp(-z))');
```

Our goal is to use Newton's method to minimize the cost function. The update rule for Newton's method is

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1}\nabla_\theta J \qquad (2)$$

In logistic regression, the gradient and the Hessian are

$$\nabla_\theta J = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)} \qquad (3)$$

$$H = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)})\; x^{(i)}\left(x^{(i)}\right)^T \qquad (4)$$

C. Implementation

Now, implement Newton's Method in your program, starting with the initial value of $\theta = \mathbf{0}$. To determine how many iterations to use, calculate $J(\theta)$ for each iteration and plot your results as you did in Exercise 2. As mentioned in the lecture videos, Newton's method often converges in 5-15 iterations. If you find yourself using far more iterations, you should check for errors in your implementation.

After convergence, use your values of theta to find the decision boundary in the classification problem. The decision boundary is defined as the line where

$$P(y=1|x;\theta) = g(\theta^T x) = 0.5$$

which corresponds to

$$\theta^T x = 0$$

III. Data and Results

A. Procedure 3.1

Plot the data

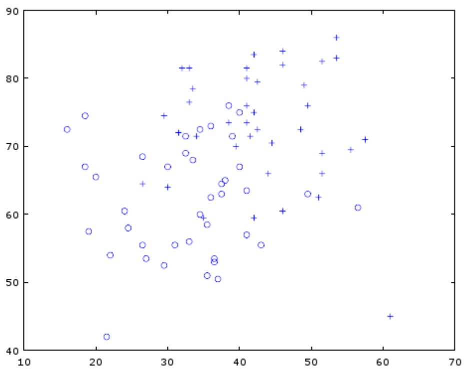

Fig. 1. Data Represents test 1 and test 2 scores

B. Procedure 3.2

Plot the Cost function J()

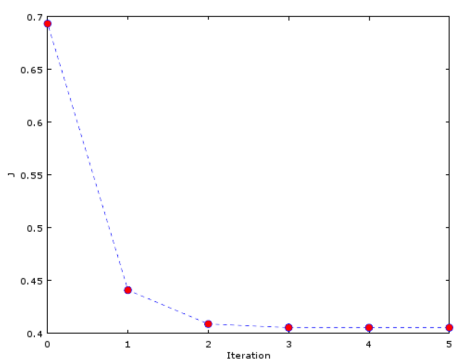

Fig. 2. Plot of the Cost function

C. Procedure 3.3

What values of θ did you get? How many iterations were required for convergence? What is the probability that a student with a score of 20 on Exam 1 and a score of 80 on Exam 2 will not be admitted?

θ_0 = -15.37, θ_1 = 0.14, θ_2 = 0.15

$J(\theta_0)$ = 0.6931, $J(\theta_1)$ = 0.4409, $J(\theta_2)$ = 0.4088, $J(\theta_3)$ = 0.4055, $J(\theta_4)$ = 0.4054, $J(\theta_5)$ = 0.4054,

probability = 66.80

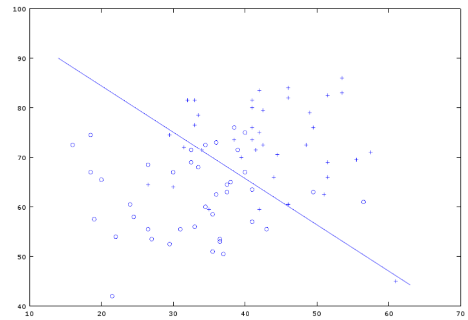

Fig. 3. Plotting the data boundary on the Data where the admitted is the 0 and the + is the admitted according to the boundary line

D. Providing own data

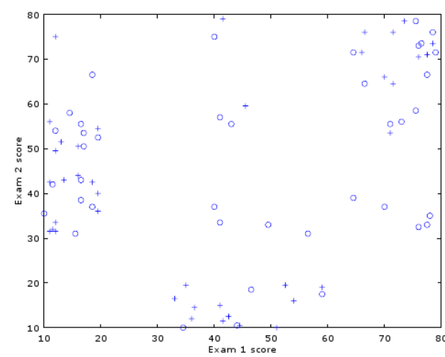

Fig. 4. Plot of the own test scores

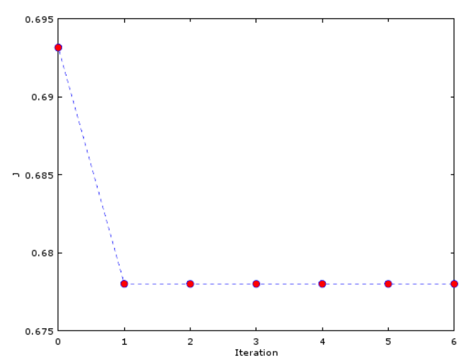

Fig. 5. Cost function

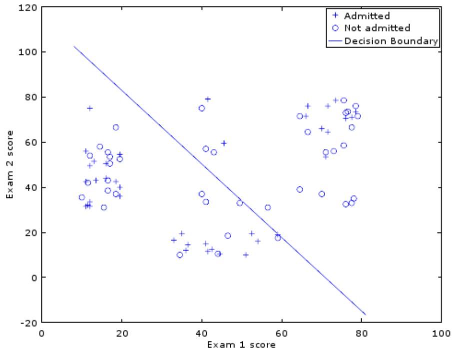

Fig. 6. Plotting the data boundary on the Data

What values of θ did you get? How many iterations were required for convergence? What is the probability that a student with a score of 20 on Exam 1 and a score of 80 on Exam 2 will not be admitted?

θ_0 = 0.8, θ_1 = -0.01, θ_2 = -0.006

probability = 0.49

$J(\theta_0)$ = 0.69315, $J(\theta_1)$ = 0.67801, $J(\theta_2)$ = 0.67801

IV. Conclusion

In conclusion, we used Newton's method to determine the outcome to a data set with multiple independent variables, a task impossible for linear regression. Based on the data gathered, it can be seen that using Newton's method, the graph converges in 5 iterations, which agrees with the theoretical amount of iterations for convergence. Therefore, it can be concluded that the cost function has been successfully minimized.

References

[1] J. McDonald, *Handbook of Biological Statistics*, 3rd ed. Sparky House Publishing, Baltimore, Maryland, 2014.

[2] F. Pampel, *Logistic regression: a primer*, 1st ed. Sage Publications, 2000.