Logistic Regression using Gradient Descent

Author: Rodriguez Noh Santiago Miguel
Professor: Ph.D. Anabel Martin Gonzalez
Link to code: https://github.com/Santiagomrn/Logistic_Regression

I. Introduction

Implement logistic regression using gradient descent to classify gender based on the weight and height of the subject. We will use height and weight as input variables x1 and x2 respectively; and gender will be the binary output variable y, where y=0 stands for female gender, and y=1 for male gender.

A. Theoretical framework

Hypothesis:

$$g(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$h(x_1, x_2) = g(b_0 + b_1 x_1 + b_2 x_2) \tag{2}$$

Loss Function:

$$C(h(x_1,x_2),y) = \begin{cases} -log(h(x_1,x_2)) & \text{if y = 1} \\ -log(1-h(x_1,x_2)) & \text{if y = 0} \end{cases} \tag{3}$$

$$L(b_0,b_1,b_2) = \frac{1}{n}\sum_{i=1}^{n} C(h(x_1^i,x_2^i),y^i) \tag{4}$$

Update Rule:
for b0:

$$\begin{aligned} b_0 &= b_0 - \alpha\frac{dL(b_0,b_1,b_2)}{db_0} \\ b_0 &= b_0 - \alpha\frac{1}{n}\sum_{i=1}^{n}(h(x_1^i,x_2^i)-y^i) \end{aligned} \tag{5}$$

for b1:

$$\begin{aligned} b_1 &= b_1 - \alpha\frac{dL(b_0,b_1,b_2)}{db_1} \\ b_1 &= b_1 - \alpha\frac{1}{n}\sum_{i=1}^{n}(h(x_1^i,x_2^i)-y^i)x_1^i \end{aligned} \tag{6}$$

for b2:

$$\begin{aligned} b_2 &= b_2 - \alpha\frac{dL(b_0,b_1,b_2)}{db_2} \\ b_2 &= b_2 - \alpha\frac{1}{n}\sum_{i=1}^{n}(h(x_1^i,x_2^i)-y^i)x_2^i \end{aligned} \tag{7}$$

II. Development of practice

A. Plot your data set and label the axes (“Heights”, “Weights”). Plot height-weight data for males in red and females in blue.

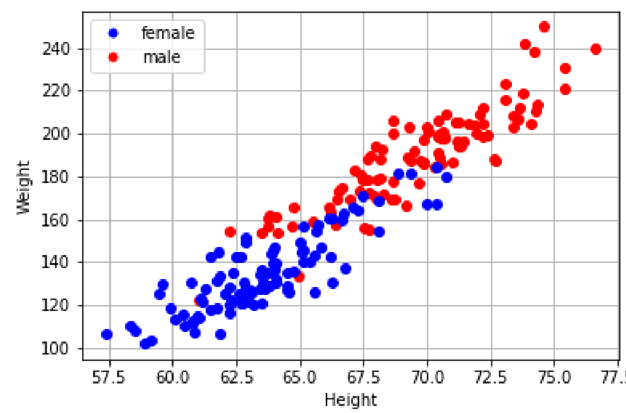

Fig. 1. Weight versus height graph.

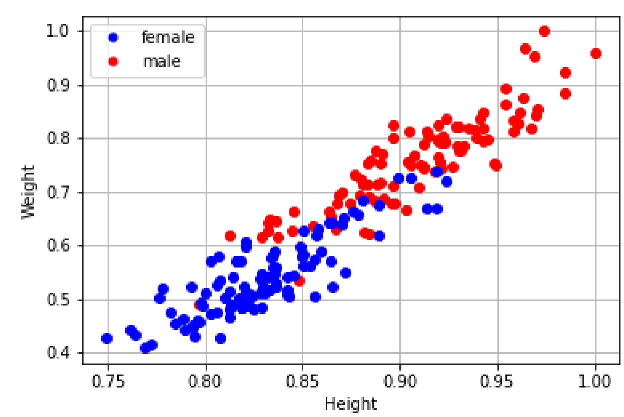

Fig. 2. Normalized weight versus height graph.

B. Plot your loss values after each iteration and record your initial and final loss values.

```
step: 0
learningRate 0.5
loss:  0.6931471805599465
gradient b: 0.0   gradient b0: -0.019407971011520146 gradient b1: -0.0535917855218883
b: 0.0 ,b0: 0.009703985505760073 b1: 0.02679589276094415


step: 0
learningRate 1
loss:  0.6931471805599465
gradient b: 0.0   gradient b0: -0.019407971011520146 gradient b1: -0.0535917855218883
b: 0.0 ,b0: 0.019407971011520146 b1: 0.0535917855218883


step: 0
learningRate 1.5
loss:  0.6931471805599465
gradient b: 0.0   gradient b0: -0.019407971011520146 gradient b1: -0.0535917855218883
b: 0.0 ,b0: 0.029111956517280217 b1: 0.0803876782828283244


step: 0
learningRate 2
loss:  0.6931471805599465
gradient b: 0.0   gradient b0: -0.019407971011520146 gradient b1: -0.0535917855218883
b: 0.0 ,b0: 0.03881594202304029 b1: 0.1071835710437766


step: 0
learningRate 2.5
loss:  0.6931471805599465
gradient b: 0.0   gradient b0: -0.019407971011520146 gradient b1: -0.0535917855218883
b: 0.0 ,b0: 0.048519927528800355 b1: 0.13397946380472076
```

Fig. 3. Results of the first iterations with various learning rates

```
step: 188836
learningRate 0.5
loss:  0.2393126480774959
gradient b: -6.280190293286363e-05   gradient b0: 0.00011393618448440357 gradient b1: -5.544353607379263e-05
b: -0.3345703138780685 ,b0: -29.23005787625595 b1: 39.79282801473424


step: 114326
learningRate 1
loss:  0.23902573909193114
gradient b: -4.4060703697587094e-05   gradient b0: 8.052261950349172e-05 gradient b1: -3.967879793001466e-05
b: 0.7173934016839891 ,b0: -31.145544306939 b1: 40.73084091181499


step: 84085
learningRate 1.5
loss:  0.23892874564125816
gradient b: -3.584881207107241e-05   gradient b0: 6.572966370067914e-05 gradient b1: -3.25689096853594e-05
b: 1.1871979269118493 ,b0: -32.00553543302925 b1: 41.15579129210056


step: 67280
learningRate 2
loss:  0.23887989828589973
gradient b: -3.097987059628129e-05   gradient b0: 5.6913677988847466e-05 gradient b1: -2.8293462685920625e-05
b: 1.4684750395331179 ,b0: -32.52176916436986 b1: 41.41200548381746


step: 56452
learningRate 2.5
loss:  0.2388504574527841
gradient b: -2.766879892768798e-05   gradient b0: 5.0899007913363624e-05 gradient b1: -2.5360123745839098e-05
b: 1.6609535926207188 ,b0: -32.87561206808029 b1: 41.58810853037462
```

Fig. 4. Results of the final iterations with various learning rates

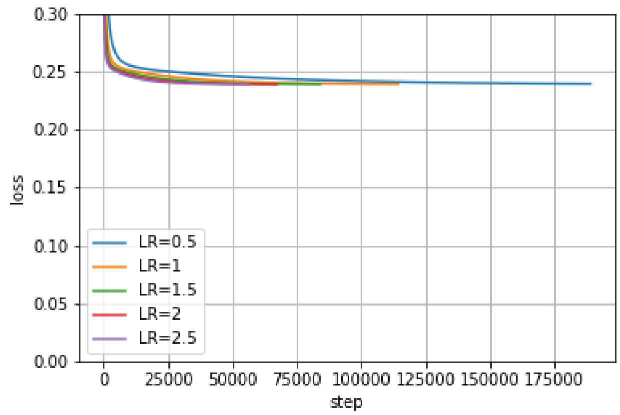

Fig. 5. Loss with various learning rates

C. What learning rate α value did you use?

Based on the results of the cost function, the best model is the one that was trained with a learning rate of 2.5. The predictions will be made with this model.

D. After convergence, plot the decision boundary from your algorithm on the same graph as your data and show your optimized parameters values.

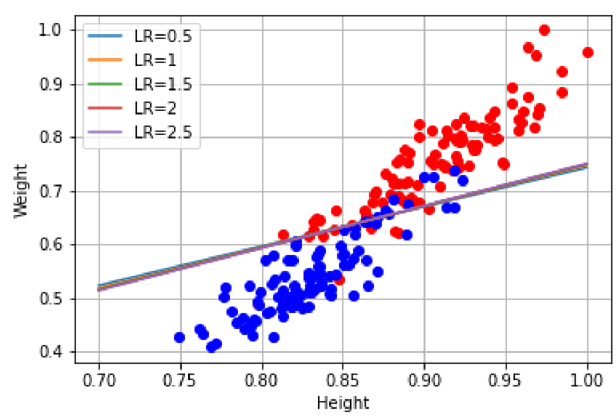

Fig. 6. Decision boundaries of various models

The final values of the trained weights are:

$$
\begin{aligned}
b_0 &= 1.6609535926207188 \\
b_1 &= -32.87561206808029 \\
b_2 &= 41.58810853037462
\end{aligned} \tag{8}
$$

Taking into account that the training was with normalized data the weights are:

$$b_0 = 1.6609535926207188$$
$$b_1 = \frac{-32.87561206808029}{76.6001829545762} = -0.42918451105496025 \quad (9)$$
$$b_2 = \frac{41.58810853037462}{249.946283195065} = 0.16638818548831194$$

The purposed model is:

$$h(x) = g(b_0 + b_1 x_1 + b_2 x_2) \quad (10)$$

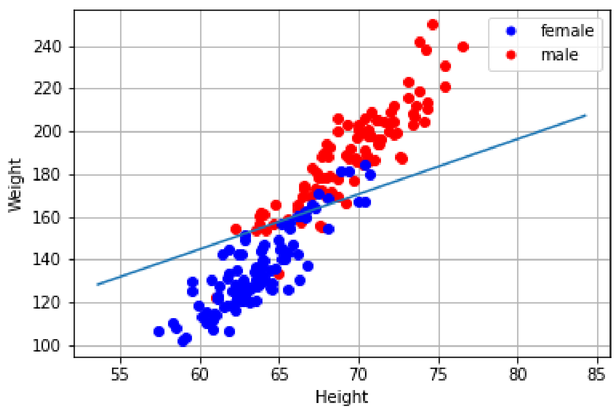

Fig. 7. Final Model

E. After you have completed the training, report the training accuracy of your classifier by computing the percentage of examples it got correct.

The purposed model classified correctly 91% of training examples.

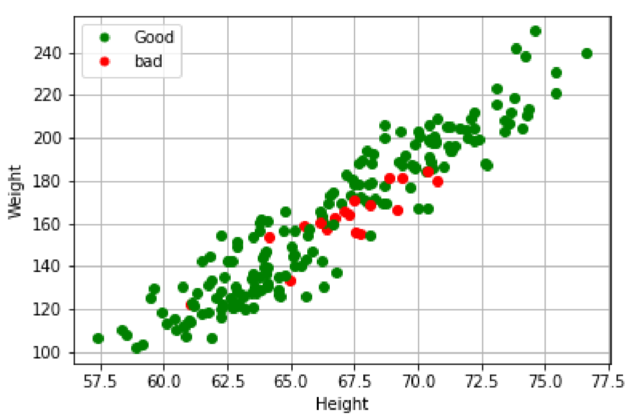

Fig. 8. Examples classified correctly

F. Finally, use your trained model to classify gender for a person of height 71.731 and weight 220.042. Show your results.

The value obtained was 0.9994402769337613 and this tells us that the person who has these measurements is a man.

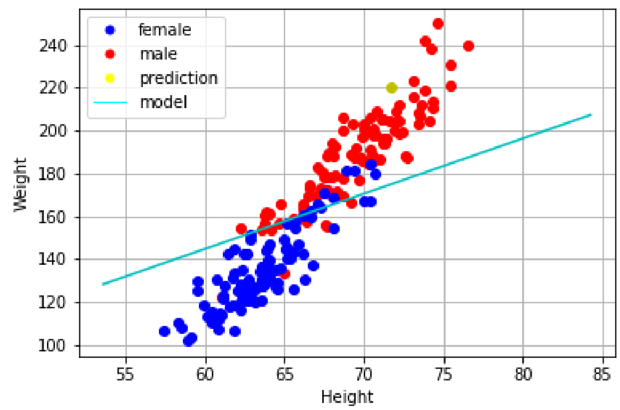

Fig. 9. Prediction

REFERENCES

[1] Andrew Ng.*CS229 Lecture Notes*, http://cs229.stanford.edu/notes2020spring/cs229-notes1.pdf.