ISTANBUL SEHIR UNIVERSITY, ECE 592B

# Logistic Regression

Oğuzhan Sağoğlu

May 23, 2016

## 1 INTRODUCTION

Logistic Regression is used to estimate the probability of a thing being one of the two categories by feeding its features to a hypothesis.
As for any Machine Learning algorithm we should have a hypothesis function. For our purposes, this function should return values between 0 and 1, since Logistic Regression should return a probability of an event. So:

$$0 < h_Q(x) < 1 \tag{1.1}$$

Since our famous polynomial function$(\theta_0 + \theta_1 X_1 + \theta_2 X_2 ...)$ can return values between $-\infty$ and $+\infty$ we should map it with the sigmoid function in order to get values between 0 and 1. Lets see:

$$-\infty < \theta_0 + \theta_1 X_1 + \theta_2 X_2 < +\infty \tag{1.2}$$

$$sigmoid(z) = \frac{1}{1 + e^{-z}} \tag{1.3}$$

$$sigmoid(-\infty) < sigmoid(\theta_0 + \theta_1 X_1 + \theta_2 X_2 ...) < sigmoid(+\infty) \tag{1.4}$$

$$0 < sigmoid(\theta_0 + \theta_1 X_1 + \theta_2 X_2 ...) < 1 \tag{1.5}$$

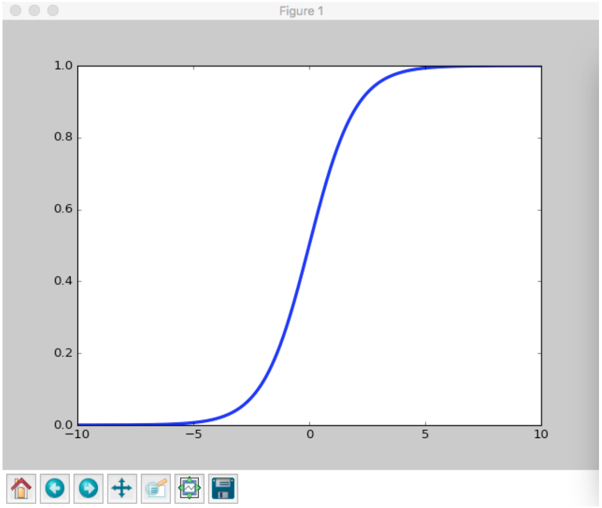
So:

$$h(x) = sigmoid(\theta_0 + \theta_1 X_1 + \theta_2 X_2 ...) \tag{1.6}$$

So we can say that:

$$h(x) = P(y = 1|x; Q) \tag{1.7}$$

$$1 - h(x) = P(y = 0|y;Q) \tag{1.8}$$



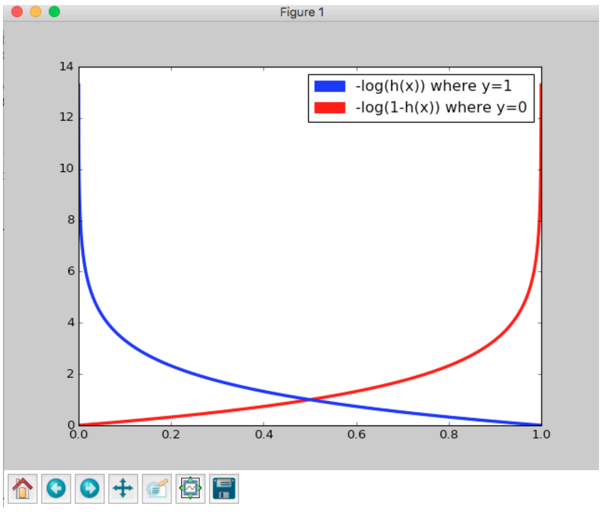

(1.9)

This is the graph of the sigmoid function

To calculate optimum values for the Q, we should minimize cost function:

$$\underset{J}{\text{minimize}} J(\theta) = -ylog(h(x)) - (1-y)log(1-h(x)) \tag{1.10}$$

Note: This cost function is different than the linear regression's cost function($[h(x)-y]^2$) since hypothesis function of Logistic Regression prevent the cost function to be convex. Cost function above is ensure the convexity for both y=0 and y=1 as shown in the graph 1.11.




(1.11)

In order to minimize cost function, we use gradient descent method like this:

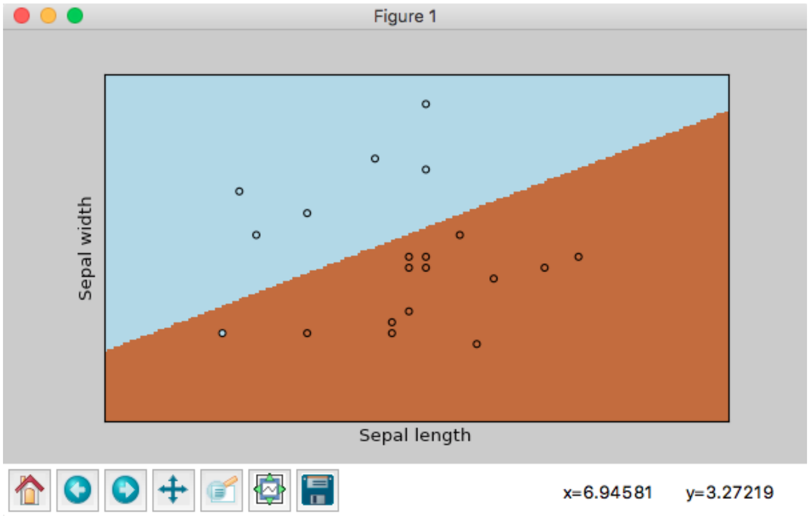
$$\theta_i := \theta_i - \alpha \frac{\partial}{\partial Q_i} J(Q) \tag{1.12}$$

$$\frac{\partial}{\partial Q_i} J(Q) = [h(X) - Y] * X_i \tag{1.13}$$

This is the same gradient function with the linear regression's gradient function, only difference is the hypothesis function which is mapped to a sigmoid function.

# 2 Methodology

To prove that equations above are correct, we implemented logistic regression in Python programming language with it's library called NumPy. We used iris dataset, we choose sepal length and sepal width as the features and two target values setosa(0) and versicolor(1). We used 80% of the data as training set and 20% as test set. When we draw the graph:




(2.1)

All of the test examples are estimated correctly except one.
Codes and examples are in my personal Github.