Logistic Regression

Cong Cuong PHAM

November 30, 2016

Abstract

This document introduces some fundamental notions of Logistic Regression.

1 Introduction

1.1 What is Logistic Regression?

Logistic Regression is a method for **Classification**. Examples: Spam versus "Ham" emails; Loan defaults, Disease Diagnosis...

Logistic Regression allows to solve classification problems, where we are trying to predict discrete categories (although we have continuous value). The convention for **binary classification** is to have two classes 0 and 1.

We can't use a normal linear regression model on binary groupes. It won't lead to a good fit:

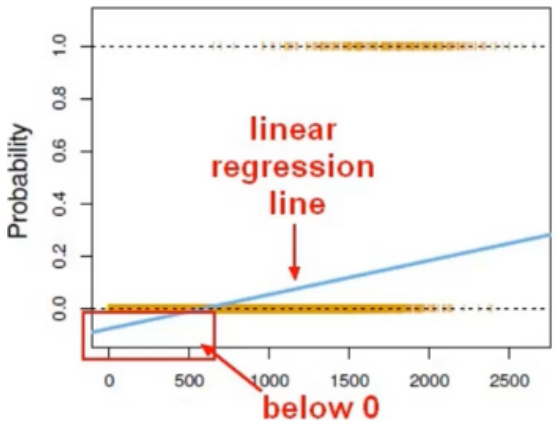

Figure 1: The linear regression is not suitable for the binary classification.

We can transform our linear regression to a logistic regression curve. The Sigmoid (aka Logistic) Function takes in any value and ouputs it to be between 0 and 1 (Figure 3).

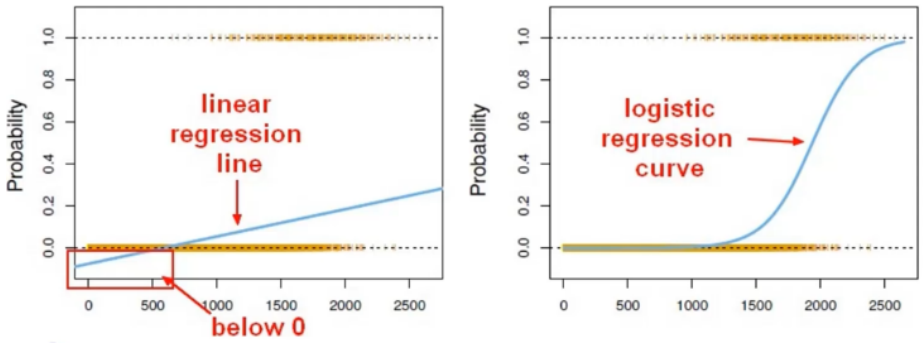

Figure 2: Linear and logistic regression.

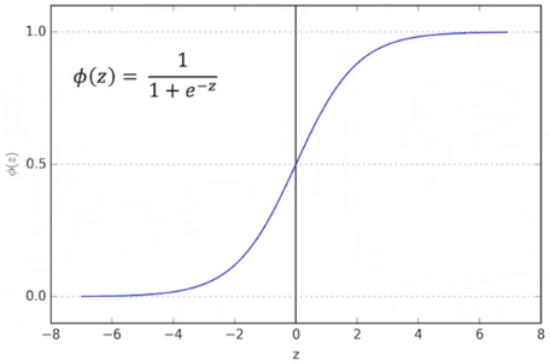

Figure 3: Sigmoid (aka Logistic) function.

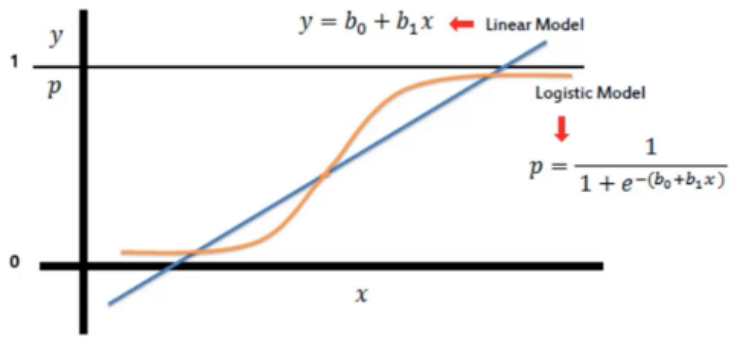

Figure 4: Linear Regression Solution used Sigmoid Function.

After training a logistic regression model on some training data, you will evaluate your model's performance on some test data. A **confusion matrix** (Figure 5) could be used to evaluate classification models.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Example: Test for presence of disease
NO = negative test = False = 0
YES = positive test = True = 1

Figure 5: Confusion matrix used to evaluate classification models.

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Basic Terminology:
- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

Figure 6: Confusion matrix - F value.

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Misclassification Rate (Error Rate):

- Overall, how often is it **wrong**?
- (FP + FN) / total = 15/165 = 0.09

Figure 7: Confusion matrix - Misclassification Rate.

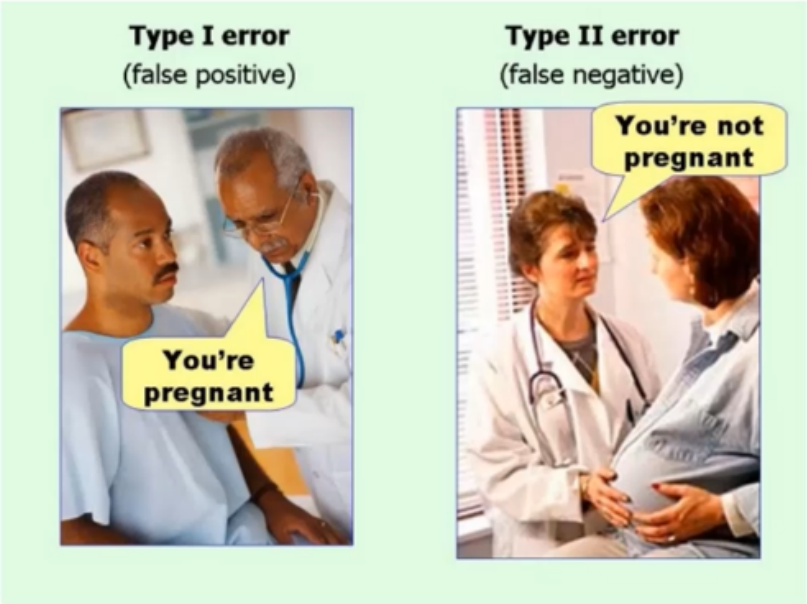

Figure 8: Confusion matrix - Misclassification Rate - Example.