CS 4980/6980: Data Science © Spring 2018

Lecture 6: Logistic Regression

Instructor: Daniel L. Pimentel-Alarcón Scribed by: Ratanpriya Shrivastava & Emmanuel Obaseki

This is preliminary work and has not been reviewed by instructor. If you have comments about typos, errors, notation inconsistencies, etc., please email Ratanpriya Shrivastava and Emmanuel Obaseki at `rshrivastava1@student.gsu.edu`, `eobaseki2@student.gsu.edu`.

6.1 Introduction

Logistic Regression is a mathematical model which is widely used for analyzing and predicting the probability of data in which there are one or more variables that determine an outcome. It is applied when an event has only two results, such as *True/False* or *Pass/Fail*.
Now, we introduce Logistic Regression as a tool for building models.

For example:

- Generalized Logistic Models.
- Predict variable of interest for rest of data.
- Study log(odds).
- Study the "Likelihood" of the data.
- Difference between probability and likelihood.
- Model the probabilities of variable of interest as a function of some explanatory variables.
- Find argument that finds maximum likelihood.

6.2 Logistic Regression

First, we would like to show a brief introduction of probability

$$y \in \{0, 1\}$$

In the above equation, y is categorical material, which means that output can take only two values; "0" or "1", that represents outcomes such as True/False, Win/Lose or Dead/Alive.

If random variable y has the following equation:

$$P(Y = 1) = p$$

$$P(Y = 0) = 1 - p$$

For $0 < p < 1$, y is called *Bernoulli* random variable and it expressed as

$$Y \sim Bernoulli(p)$$

Probability is used to compare two events. Assume "p" is the event of a success and "1-p" as the event of a failure. Then take probability of success and divide it by probability of a failure to find its **odds**

$$odds := \frac{p}{1-p}$$

Odds by definition, is the extent to which an event is likely to occur. It is measured by the real number of the favorable cases possible. So in the above equation,"p" must be between 0 and 1.

Now, taking the Logarith for probability of odds

$$log(\frac{p}{1-p})$$

Then formally, the model of logistic regression can be represented as:

$$log(\frac{p}{1-p}) = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_D X_D$$

We have taken Log of the odds ratio. The Log of odds are given by the linear combination of the independent variable and the parameters. In a Logistic Model, the parameter of the model beta and X are crucial to determine the odds of a success. The larger the number is, the higher the probability of success.

6.3 Probability

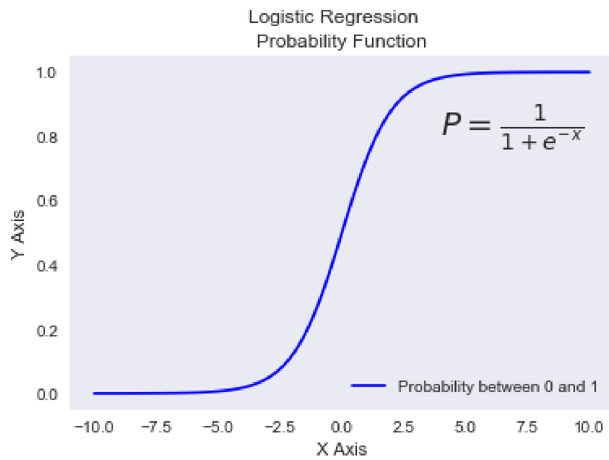
Now, to get the value of probability, Exponential function is applied to solve for p.
Applying Exponential function :

$$\frac{p}{1-p} = e^{(\beta^T X)}$$
$$p = (1-p)e^{(\beta^T X)}$$
$$p = e^{(\beta^T X)} - pe^{(\beta^T X)}$$
$$p + pe^{(\beta^T X)} = e^{(\beta^T X)}$$
$$p(1 + e^{(\beta^T X)}) = e^{(\beta^T X)}$$
$$p = \frac{e^{(\beta^T X)}}{1+e^{(\beta^T X)}}$$

This could be the expression for probability but can be simplified as :

Let's divide Numerator and Denominator and bottom by $e^{(\beta^T X)}$

$$p = \frac{1}{1+e^{(-\beta^T X)}}$$



P represents the probability of 1 and e is the base of the natural logarithm. Finally, beta transpose T and feature vector x are the parameters of the model.

Now, let's take an example of Titanic. In the following equations survival and perished cases are demonstrated. Survival case :- odds of survival

$$(\frac{1}{1+e^{(-\beta^T X)}}) > (1 - \frac{1}{1+e^{(-\beta^T X)}})$$

Here y=1

Perished case :- odds of perished

$$(\frac{1}{1+e^{(-\beta^T X)}}) < (1 - \frac{1}{1+e^{(-\beta^T X)}})$$

Here y=0

Till here, we have completed the setup for Logistic Regression, now our next goal is to find

$$\beta$$

$$\{X_i, Y_i\}^N$$

$$y_i \sim Bernoulli(P)$$

6.4 Likelihood

This section will introduce **Likelihood** function in Logistic Regression. Concept of Probability is similar to the concept of **Likelihood** , in which yi is the random variable and Yi is the fixed sample.

$$L(p|Y_i) = P(y_i|p)$$

$$f(x) = ax + b$$

$$f(a|x,b) = ax + b$$

Example of Likelihood Vs Probability

$$x \sim N(\mu, \sigma^2)$$

$$P(x) = \frac{1}{\sqrt{(2\pi)}\sigma} e^{(\frac{-1}{2})(\frac{x-\mu}{2})^2}$$

$$P(x|\mu, \sigma)$$

Here 'x' is the variable of interest

$$L(x) = \frac{1}{\sqrt{(2\pi)}\sigma} e^{(\frac{-1}{2})(\frac{x-\mu}{2})^2}$$

$$L(\mu|x, \sigma)$$

Here the variable of interest is :-μ

Probability of the i^{th} sample:-

$$P(y_i) = P^{(y_i)}(1-p)^{(1-y_i)}$$

Likelihood of the i^{th} sample:-

$$L(P|Y_i) = P^{(Y_i)}(1-P)^{(1-Y_i)}$$

Now, as we already know

$$P = \frac{1}{1+e^{(-\beta^T X)}}$$

Hence, substituting the values in the above equation.

$$L(\beta|Y_i) = (\frac{1}{1+e^{(-\beta^T X_i)}})^{(Y_i)}(\frac{1}{1+e^{(-\beta^T X_i)}})^{(1-X_i)}$$

And as

$$P(y_1, y_2, ..., y_n)^p = \prod_{i=1}^{N} P(y_i)$$

$$= \prod_{i=1}^{N} p^{(y_i)}(1-p)^{(1-y_i)}$$

Now Similarly for Likelihood

$$L(P|Y_1, Y_2, ..., Y_N) = \prod_{i=1}^{N} L(P|Y_i)$$

$$= \prod_{i=1}^{N} P^{(Yi)}(1-P)^{(1-Yi)}$$

$$= \prod_{i=1}^{N} (\frac{1}{1+e^{(-\beta^T X)}})^{(Yi)}(1-(\frac{1}{1+e^{(-\beta^T X)}}))^{(1-Yi)}$$

We will take derivative of the above equation and solve for β as it is our variable of interest. Now applying log to the above equation

$$log(L(P|Y_1, Y_2, ..., Y_N)) =: l(P|Y)$$

$$log[\prod_{i=1}^{N}(\frac{1}{1+e^{(-\beta^T X_i)}})^{(Y_i)}(1-(\frac{1}{1+e^{(-\beta^T X_i)}}))^{(1-Y_i)}]$$

$$\sum log((\frac{1}{1+e^{(-\beta^T X_i)}})^{(Y_i)} + log((1-(\frac{1}{1+e^{(-\beta^T X_i)}}))^{(1-Y_i)})$$

$$l(P|Y) = \sum_{i=1}^{N} Y_i log(\frac{1}{1+e^{(-\beta^T X_i)}}) + (1-Y_i) log((1-(\frac{1}{1+e^{(-\beta^T X_i)}}))$$

or

$$l(\beta|Y) = \sum_{i=1}^{N} Y_i log(\frac{1}{1+e^{(-\beta^T X_i)}}) + (1-Y_i) log((1-(\frac{1}{1+e^{(-\beta^T X_i)}}))$$

Now, we have to find the best β

$$\hat{\beta} := argmax(l(\beta|Y))$$

Where $\beta \in R^D$

6.5 Conclusion

This lecture shows logic behind Logistic Regression. Like ordinary Regression such as Linear Regression, Logistic Regression provides coefficient "beta" and feature vector "x", which measures each variable as information as well as odds ratio that represents probability of success. The goal is to correctly predict the outcome for each case using the most precise mathematical model.