Daffodil International University

Department of Computer Science and Engineering (CSE)
Faculty of Science and Information Technology (FSIT)

Logistic Regression Mathematical Examples Lecture Sheet

Semester: Spring 2024
Course Code and Title: CSE315 – Introduction to Data Science
Course Teacher and Initial: Fahim Faisal (FF)
Section: 61_L

Problem: We have a dataset of students where we want to predict whether a student will pass (1) or fail (0) an exam based on the number of hours they studied. We have the following data:

Hours Studied	Exam Result
2	0
3	0
4	0
5	1
6	1
7	1
8	1

Based on the above data, build a logistic regression model to predict the probability of a student passing the exam if he/she has studied for 5 hours.

Solution:

We know, the formula for logistic regression model,

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Here,

$P(Y = 1 \mid X)$ is the probability of passing the exam given the number of hours studied
e is the base of the natural logarithm.
β_0 is the intercept.
β_1 is the coefficient for the number of hours studied.

We want to estimate the coefficients β_0 and β_1 that maximize the likelihood of observing the given data. Let's assume we find the following coefficients after fitting the logistic regression model:

$\beta_0 = -4$
$\beta_1 = 0.8$

Now, we can use these coefficients to predict the probability of passing the exam for a student who studied for, let's say, 5 hours:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(-4+0.8\times 5)}}$$

$$Or, P(Y = 1 \mid X) = \frac{1}{1 + e^{-(-4+4)}}$$

$$Or, P(Y = 1 \mid X) = \frac{1}{1 + e^{-(0)}}$$

$$Or, P(Y = 1 \mid X) = \frac{1}{1 + 1}$$

$$Or, P(Y = 1 \mid X) = \frac{1}{2}$$

So, the probability of passing the exam for a student who studied for 5 hours is $\frac{1}{2}$ or 50%.

Problem: Estimate β_0 and β_1 for the following dataset using gradient descent:

X	Y
1	0
2	0
3	0
4	1
5	1

Solution:

Step-1: First, we'll initialize β_0 and β_1 to some arbitrary values, say $\beta_0 = 0$ and $\beta_1 = 0$ and then we'll use gradient descent to update these coefficients.

Step-2: For each observation, we need to calculate the predicted probability using the logistic function:

$$P(Y = 1 \mid X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_i)}}$$

Step-3: Next, we need to compute the gradient of the log-likelihood function with respect to β_0 and β_1. The gradient gives us the direction in which we need to update the coefficients to increase the likelihood of the data:

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{N} \left(Y_i - P(Y = 1 \mid X_i)\right)$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{N} \left(Y_i - P(Y = 1 \mid X_i)\right) \times X$$

Step-4: Next, we'll update the coefficients using the gradient and a learning rate (α):

$$\beta_0 \coloneqq \beta_0 + \alpha \times \frac{\partial L}{\partial \beta_0}$$

$$\beta_1 \coloneqq \beta_1 + \alpha \times \frac{\partial L}{\partial \beta_1}$$

We need to repeat steps 2-4 until β_0 and β_1 converges or a maximum number of iterations is reached. Let us assume $\alpha = 0.01$ and we want to iterate for 10 iterations. In that case, predicted probabilities are as follows:

[0.5 0.5 0.5 0.5 0.5]
[0.50249998 0.50624967 0.50999867 0.51374653 0.51749286]
[0.5044063 0.51143581 0.51846079 0.52547849 0.53248613]
[0.50580797 0.51571693 0.52561354 0.53549007 0.54533885]

[0.50678106 0.51922902 0.53165315 0.54403816 0.55636895]
[0.50739027 0.52208796 0.53674751 0.55134381 0.5658522]
[0.50769054 0.52439219 0.54103946 0.55759563 0.57402479]
[0.50772842 0.52622513 0.54465014 0.56295371 0.58108743]
[0.5075434 0.52765744 0.54768208 0.56755355 0.58721]
[0.50716894 0.52874899 0.55022211 0.57150971 0.59253589]

Gradient of $\beta_0 \left(\frac{\partial L}{\partial \beta_0}\right)$ values are as follows:

-0.5
-0.5499877131150956
-0.5922675180702318
-0.6279673643299398
-0.6580703313896464
-0.6834217620306211
-0.7047426154264421
-0.7226448275204158
-0.7376464679251347
-0.7501856521507916

Gradient of $\beta_1 \left(\frac{\partial L}{\partial \beta_1}\right)$ values are as follows:

1.5
1.3125542432528388
1.1529951106647682
1.0172630177823172
0.9018040862245722
0.8035550026842251
0.7199002052243437
0.6486189049070079
0.5878312865871211
0.5359484241300958

β_0 values are as follows:

-0.005
-0.010499877131150956
-0.016422552311853272
-0.02270222595515267
-0.029282929269049136
-0.03611714688935535
-0.04316457304361977
-0.050391021318823934
-0.05776748599807528
-0.0652693425195832

β_1 values are as follows:

0.015

0.028125542432528388
0.03965549353917607
0.04982812371699925
0.05884616457924497
0.06688171460608722
0.07408071665833066
0.08056690570740074
0.08644521857327195
0.09180470281457291

Final β_0 and β_1 values after 10 iterations are respectively -0.0652693425195832 and 0.09180470281457291.