Multi-Variate Logistic Regression

Subhash,Ishwar,Atul,Shubhangi

07, August 2020

1 Introduction

In the previous session you have seen that the method of logistic regression is similar to linear regression except that the final outcome is binary ie. success/failure. We have also seen the features of the sigmoid function. In this session we will see how to implement logistic regression in more than one variable for classification problems .

2 Intuition on Multi-Class Classification

We'll talk about how to get logistic regression to work for multi-class classification problems. And in particular I want to tell you about an algorithm called one-versus-all classification.
What's a multiclass classification problem?
Here are some examples.
Lets say you want a learning algorithm to automatically put your email into different folders or to automatically tag your emails soyou might have different folders or different tags for work email,email from your friends, email from your family, and emails about your hobby.And so here we have a classification problem with four classes which we mightassign to the classes $y = 1$, y =2, y =3, and $y = 4$ too. And another example, for medical diagnosis,if a patient comes into your office with maybe a stuffy nose,the possible diagnosis could be that they're not ill.Maybe that's $y = 1$. Or they have a cold, 2. Or they have a flu. And a third and final example if you are using machine learning to classify the weather, you know maybe you want to decide that the weather is sunny, cloudy,rainy, or snow, or if it's gonna be snow, and so in all of these examples,y can take on a small number of values, maybe one to three, one to four and so on, and these are multiclass classification problems.

3 Logistic Regression

Now we are going to use the same method of linear regression in multiple variable.

$$\bar{z} = w_0 + w_1 * x_1 + w_2 * x_2 \dots \tag{1}$$

$$\bar{y} = \sigma(\bar{z}) = \frac{1}{e^{\bar{z}} + 1} \tag{2}$$

Now the idea that we have to implement gradient descent for this certain process come in to our mind. But first we have to define the loss function for this method.

4 Loss Function

The loss function we are going to use is called **Binary Cross-Entropy** and its defines as :

$$L(y, \bar{y}) = -[y * log(\bar{y}) + (1 - y) * log(1 - \bar{y})] \tag{3}$$

where,

$\bar{y}$ is predicted output which lies in the range $0 \leq \bar{y} \leq 1$ and y is 0 or 1 depending on the classification.

5 Cost Function and Gradient Descent

5.1 Cost Function

Cost function takes two parameters as input y_i and $\bar{y_i}$,you can think of it as the cost the algorithm has to pay if it makes a prediction $\bar{y_i}$ while the actual label was y_i.

$$\mathrm{J} = 1/\mathrm{m} \sum_{i=1}^{m} L\left(y\hat{}(i), y(i)\right) = -1/\mathrm{m} \sum_{i=1}^{m} \left(y(i) * \log y\hat{}(i) + (1 - y(i)) * \log\left(1 - y\hat{}(i)\right)\right) \tag{4}$$

5.2 Gradient Descent

The objective of training a machine learning model is to minimize the loss or error between ground truths and predictions by changingthe trainable parameters.
And gradient, which is the extension of derivative in multi –dimensionalspace, tells the direction along which the loss or error is optically minimized. If you recall from vector calculus class, gradient is defined as the maximum rate of change.
Therefore,theformula for gradient descent is given as :

Repeat Until Convergence:-

$$w_j := w_j - \alpha * \frac{dJ}{dw_j} \tag{5}$$

where,
α is the learning rate and J is the cost function

6 Example

Now lets make use of multi-variate logistic regression in the dataset as shown below:

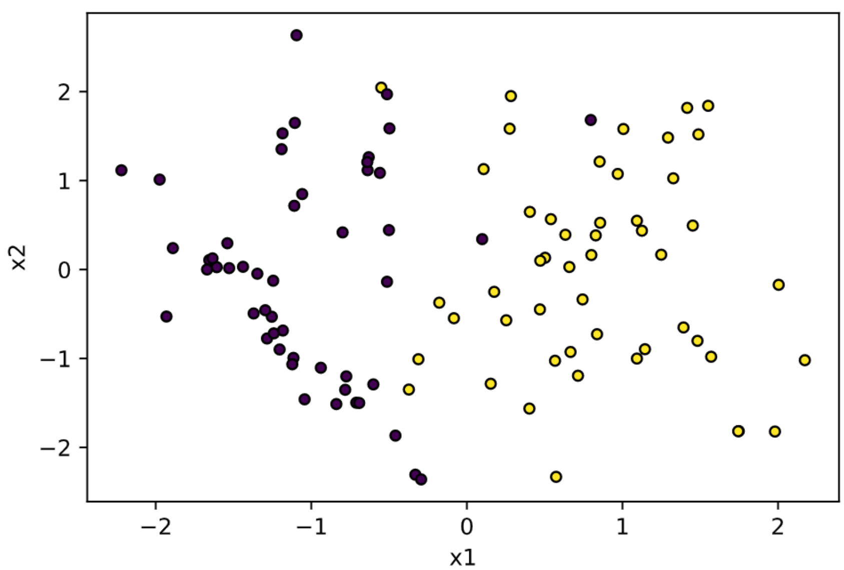

Figure 1: 2-D data having 2 variables x1 and x2

So now we define three parameters for the logistic regression such as w_0 , w_1, w_2

$$\bar{y_i} = \frac{1}{1+e^{-(w_0+w_1*x_1+w_2*x_2)}} \tag{6}$$

after 100 iterations we get the Decision Boundary of our Logistic Regressor

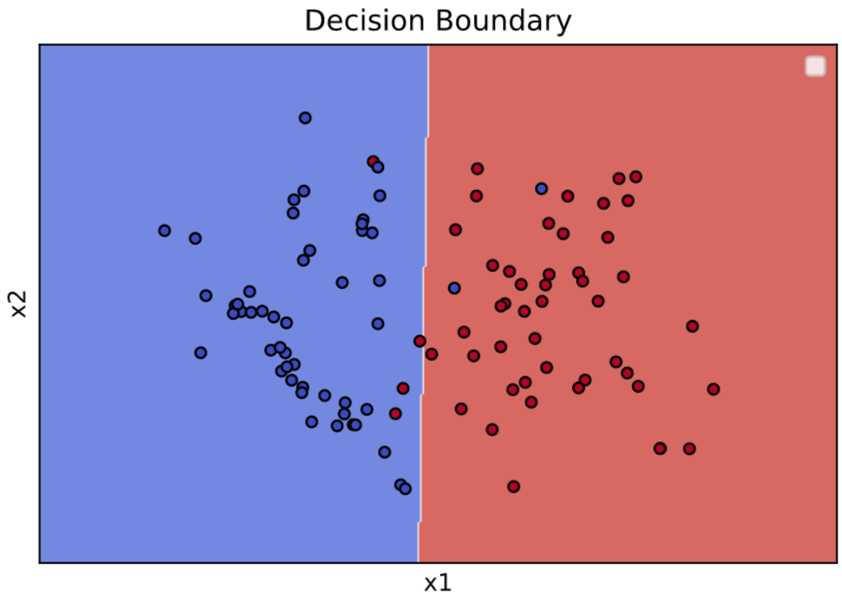

Figure 2: Our Decision boundary

final model parameters are :
$\vec{w} = (0.44589015, 3.0173601986106813, -0.034951518186683156)$