Logistic Regression

Ahmad Sebaq
Nile University
a.sebaq@nu.edu.eg

Khaled Moawad
Nile University
k.abdulrazek2107@nu.edu.eg

Presented to Dr. Mostafa El Attar

CIT-651: Intro.to Machine Learning
Spring 2022
Nile University

May 23, 2022

Contents

1 Logistic Regression

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.
The loss function for logistic regression is Log Loss, which is defined as follows:

$$LogLoss = \sum_{(x,y)\in D} -y * log(y\prime) - (1-y) * log(1-y\prime)$$

where:

- $(x, y) \in D$ is the data set containing many labeled examples, which are pairs.
- y is the label in a labeled example. Since this is logistic regression, every value of must either be 0 or 1.
- $y\prime$ is the predicted value (somewhere between 0 and 1), given the set of features in .

2 Experiments

We have implemented the logistic regression algorithm and used it to classify healthy and patients from *heart.csv* dataset. We started by exploring the data, tried to get some insights from it, and found the following:

- The dataset contain 13 feature with different scale.
- Some of the features (*cp, thalach, oldpeak, exang, ca, and thal*) have good correlation with target variable that we aim to predict.

2.1 Data pre-processing

We started our experiments by ingesting all the data as it is to the model and it reached **57%** accuracy, so we **normalized the features using mean and standard divination for each feature, the accuracy increased to reach 83%**

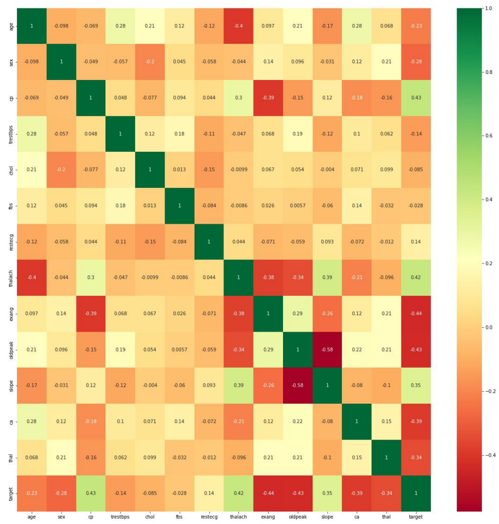

Figure 1: Correlation matrix

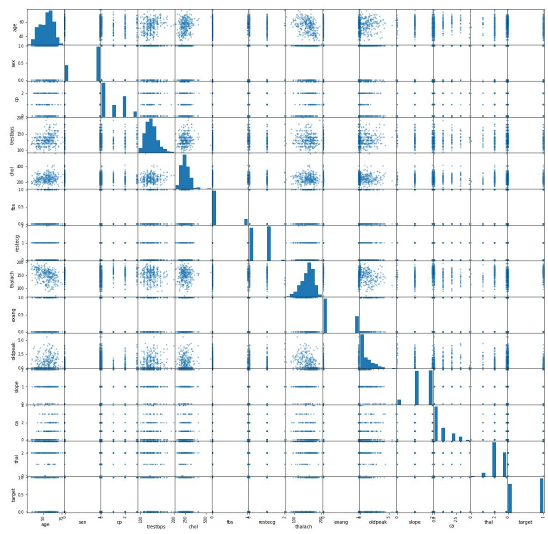

Figure 2: Scatter matrix

2.2 Learning rate

We started with a high **learning rate equal to 1** and started to reduce it a bit by bit hoping to achieve better result and we found that **the best learning rate to have is 0.1**.

In our experiment we tried multiple learning rate starting from 0.1 to 0.0001 and we found that the higher the learning rate the more stochastic the training will be and the smaller the learning rate the slower the training will be so one need to choose such hyper-parameter carefully.

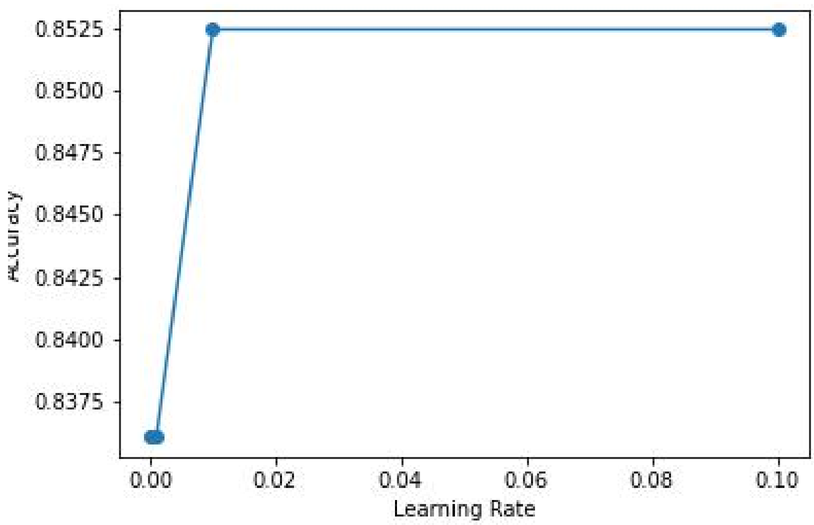

Figure 3: Learning Rates

2.3 Number of iterations

We started with a small **iteration number equal to 10** and started to increase it. We found that **1000 iteration** is good weather we used a smaller learning rate or not, and we came to the conclusion the more you train your model the more accurate and stable it becomes.

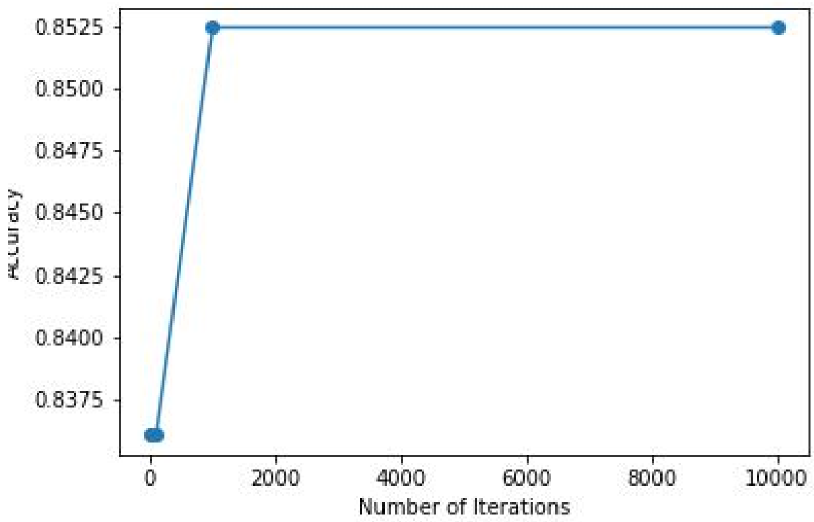

Figure 4: Number of iterations

2.4 Features

As we mentioned that we started with all the features and we reached **83 %** accuracy, then we started to take pairs from our features at random.
Also, we selected some good pairs base on the correlation between the features and our target, we found that features like: *cp, thalach, oldpeak, exang, ca, and thal*, which have high correlation with our target achieves high accuracy.