Haripriya Pandya
CS 584 Machine Learning
Assingment 03 Report

Logistic Regression

Problem Statement:
This is a type of machine learning algorithm where there is a discriminant function determines the classification of a given example. The goals is to implement different scenarios of logistic regression, like 2 class, multiple class, and non-linear combination of x vectors.

Proposed Solution:
In order to implement the logistic regression, one should implement the discriminant function, maximum likelihood function, and function calculating model parameters. Using those function, one can conduct a cross-validation operation and calculate different metrics for the efficiency of the algorithm.

Implementation details:
For this implementation I implemented the different mathematical equations for calculating model parameters like theta, maximum likelihood function and sigmoid function that predicts the class for a vector using the value of theta and max. likelihood function. I use the Iris data set to perform my analysis.

2-class:
sigmoid_function calculates the sigmoid function values for a given data matrix,
new_theta calculates the new value of theta for the new iteration using old theta value and sigmoid function values,
likelihood calculates the log likelihood values of a given x vector for a theta of a given class label

non-linear combination of input:
In this scenario, I used the polyfit_transform function from python's numpy library to create multiple combination of the input vector.

kclass:
In this scenario, the discriminant function used differs from the class classification, which is called softmax function. 'softmax_function' is used to implement such mathematical function. Similarly, in this case the log likelihood function differs distinctly from 2 class classification. log_likelihood function implements such function.

Result and discussions:

In case of 2class logistic regression my implementation is able to achieve 60% accuracy. However to the python sklearn accuracy is 100%. With model's 100% accuracy of recall I believe my implementation requires more refining. I have analyzed that with higher number of iterations to achieve the maximum theta value and a good learning rate has great effect on the accuracy of the model.

Particularly, in case of Kclass classification, with higher number of class, it is found during the analysis that it is harder to reach optimum theta values with lower learning rate and lower number of iterations. Higher rate of learning and iterations have significantly showed increase in accuracy from 60% to 92%.