Final Project Report

On

Logistic Regression with Gradient Descent

By: Subhay Manandhar

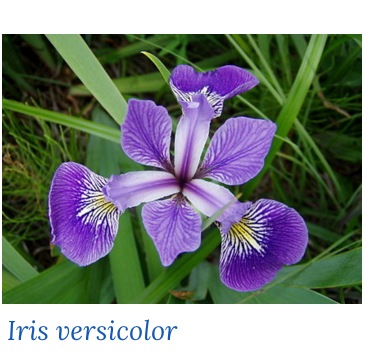
@02734858

Date: 04/15/2016

About the data:

Link to the data: <https://archive.ics.uci.edu/ml/datasets/Iris>

The data that I used for the logistic regression is an Iris flower data set. It is very common and widely popular data set used for machine learning purposes.

There are total 3 types of possible Iris flower. Iris Setosa, Iris Versicolour or Iris Virginica. 

These species of Iris plant are very similar and are hard to determine from each other through observation.

So, I took the data set containing the physical measurement of the random Iris plants.

I had 4 features data. The petal length, petal width, sepal length, sepal breadth.

With this 4 data, I wanted to use machine learning to predict a random Iris plant classification.

Using the data:

The dataset had 150 instances of data. 50 for each type of plant. I took 25 dataset from each plant for learning and the other data for testing the algorithm.

So 77 for training and 73 for testing.

Finding the regression fit:

I used the sklearn library so that I can use the regression fit library and to predict the test data using the fit algorithm.

Result:

After finding the regression fit and testing with the test data, I found that among the 73 test data, there was correct prediction in 69 cases which made my prediction correct for 94.5 percent of the time.