Summer 2018 EE690 Machine Learning Dr. L. Jololian

Assignment #2

Logistic Regression

**Due Date: Thursday, July 12**

For this assignment, you are asked to implement the logistic regression algorithm for classifying data. The implementation will be done in two different ways: First, you will use the Scikit-Learn library to classify the data, then you will use your own Python implementation of the algorithm to classify the same data a second time.

For each of the two implementations, you will split the dataset into two datasets: training and testing datasets. You will use the training dataset to develop the logistic regression model, then you will validate the model using the testing dataset.

The dataset consists of three classes of points, with each point labeled as 0, 1, or 2. Each point has 4 features, .

In your own Python implementation of the algorithm, you will use the One-vs-Rest strategy in predicting the class of a data point.

The distance between a point and a hyperplane is given by the formula  . We can also write the distance using vector notation: .

Output:

* Print the W vector for each of the three classes
* While computing the W vector for each class, show a plot of the cost as a function of the number of iterations.
* Show the accuracy in the predictions for each of the three binary classifiers.
* Show the accuracy of the overall model, in predicting the testing dataset.