Project 3

Logistic Regression

**Problem Description:**

Given a dataset representing body length and fin length of two new species of fish (i.e. Tiger0 and Tiger1), develop a logistic regression algorithm that will predict whether the fish belongs to Tiger0 or Tiger1 category.

A screenshot of a cell phone

Description automatically generated**Data Description:**

The training data consisted of 300 records representing two features of either Tiger0 fish or Tiger1 fish. Each row has three-tab separated entries. The first is a floating-point representation of the body length in centimeters, second one is the fin length in centimeters and the third column indicates the class of fish it belongs to. A plot of the data is shown in Figure 1 where green is represented by Tiger1 fish and red is represented by Tiger0 fish.

Figure 1. The Initial Data Set

**Initial Values:**

* The initial values that have been chosen for weights is [0.0,0.0,0.0,0.0,0.0,0.0]
* The initial values that have been chosen for alpha is ﻿0.5e-3
* The initial value of J is ﻿0.6931316092428014

**Final Values:**

* The final value for alpha is ﻿0.5e-3
* The final value of weights chosen is [0,0,0,0,0,0] and weights predicted by the program is

weight0 -0.4372769763266194 weight1 0.1882636823614069 weight2 -1.1535986503627857 weight3 -1.6304908326375558 weight4 -3.5558661788545614

* My Logistic Regression goes through 1000 iterations.
* The final value of J on training data is 0.﻿2974856434106421 and on test data is ﻿0.2476982833132585

**Plot of the data:**

Plotting the graph for number of iterations on the vertical axis and Cost function J on the horizontal axis.

A picture containing screenshot

Description automatically generated

**Standardization of data:**

I have calculated mean of the column first and then variance. Then with the help of variance, we calculate standard deviation. Finally, to reduce the values we subtract it by mean and divide it by standard deviation.

**Results:**

In the test set, 45 records were true positives, 32 records were true negatives, 12 records were false positives and 1 record is false negative. The results achieved an accuracy of 0.85 and precision of 0.78. Recall that is how many of the found were correct hits was equal to 0.94. The overall F1 score of Project1 is **0.86**.