Selected Topics from Comp. Science Assignment - 1 (REPORT)

Assignment submitted by the group comprising of:

1) NIKHIL JOSHI 2015A7PS0179H f20150179@hyderabad.bits-pilani.ac.in
2) ANKIT ANAND 2015A7PS0145H f20150145@hyderabad.bits-pilani.ac.in
3) KRITI GOYAL 2015A7PS0100H f20150100@hyderabad.bits-pilani.ac.in

Introduction

- The assignment comprised of 4 stages - design, implementation, documentation and testing.
- All team members contributed equally to each stage.
- Programming language used was C++ 11.
- Dataset for the assignment was retrieved from the UCI repository "banknote authentication Data Set".
- Three models of classification namely, Fisher Discriminant, Probabilistic Generative Model and Logistic Regression were implemented.

Deliverables

The directory for the Selected Topics of Comp. Science Assignment 1 consists of the following files:

1. Report.pdf (This Report)
2. code/
   - data_extraction.h
   - helper.h
   - fischer_discriminant.cpp
   - prob_gen.cpp
   - logistic_regression.cpp
   - main.cpp (the main runner file)
3. data/
   - train.txt
   - test.txt
   - ML2- Assignment 1.pdf
4. Output.txt (contains the output for the classification)

Code Execution

Since the assignment has been coded in C++
On a Linux machine, navigate to this code folder of the directory and simply type

g++ main.cpp

Followed by

./a.out

Results obtained

TASK#1: FISHER DISCRIMINANT MODEL

W_transpose is: [-0.00457795 -0.00257374 -0.00332157 -7.4492e-05]

Accuracy: 0.985437
Precision: 0.983784
Recall: 0.983784

CONFUSION MATRIX

	Predicted = 0	Predicted=1
Actual=0:	224	3
Actual=1:	3	182

TASK#2: PROBABILISTIC GENERATIVE MODEL

Accuracy: 0.973301
Precision: 0.943878
Recall: 1

CONFUSION MATRIX

	Predicted = 0	Predicted=1
Actual=0:	216	11
Actual=1:	0	185

TASK#3: LOGISTIC REGRESSION (DISCRIMINATIVE MODEL)

W_transpose is: [6.41386 -6.71389 -3.61529 -4.56107 -0.521709]

Accuracy: 0.997573
Precision: 1
Recall: 0.994595

CONFUSION MATRIX

	Predicted = 0	Predicted=1
Actual=0:	227	0
Actual=1:	1	184

NOTE: The parameters used for the Gradient Descent algorithm in the task#3 are:
No. of iterations = 11000
Learning Rate (alpha) = 0.003