**UNIVERSITY OF VICTORIA**

**Department of Electrical and Computer Engineering**

**ECE 403/503 Optimization for Machine Learning**

**LABORATORY REPORT**

Experiment No: 4

Title: Breast Cancer Diagnosis via Logistic Regression

Date of Experiment: 16 July 2019

Report Submitted on: 23 July 2019

To: Jinlong Zhan - [jinlongzhan@uvic.ca](mailto:jinlongzhan@uvic.ca)

Laboratory Group No.: B01 - 6

Name(s):

Philippe Larocque - V00903595

Alexander Fung - V00904094

## Objectives

The objective of the lab is to develop a computer program for automatic diagnosis of breast cancer based on logistic regression. [1]

## Introduction

In order to identify the breast cancer, we use a large training data set (D\_bc\_tr.mat) of 31x480. This training set is normalized, and then passed into functions to evaluate the objective function and its gradient. The results are then passed through a gradient descent to obtain the minimum point.

From the training data above, we will obtain w\* and b\* that can be used to classify a binary result +1 (malignant) or -1 (benign) for predicting breast cancer from a testing set (D\_bc\_te.mat). We can modify the number of iterations to improve the accuracy of the results.

## Results

See MATLAB code in Appedix A.

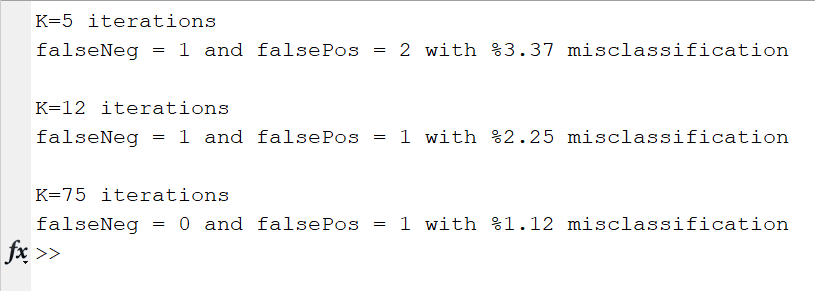


Figure 1 - Output from command line

Table 1 - Summary of results for different K iterations

| **Item** | **K=5** | **K=12** | **K=75** |
| --- | --- | --- | --- |
| False Positive | 2 | 1 | 1 |
| False Negative | 1 | 1 | 0 |
| Total Errors | 3 | 2 | 1 |
| Error rate | 3.3708% | 2.2472% | 1.1236% |

## Discussion

The results of the MATLAB code takes the testing data to calculate the number of positive or negative breast cancer diagnoses. By increasing the number of iterations performed, we approach a more accurate minimum for the function. With 5 iterations, we get 3 errors for 3.3708%, with 12 iterations, we get 2 errors for 2.2472%, and with 75 iterations we get 1 error for 1.1236%. The cause for the remaining misclassification even after 75 iterations could be due to an outlier. To correct this, more iterations may improve the results or more training data would be required to create a better model.

## Conclusion

The objective of this experiment was to build a program to diagnose breast cancer. Based on the results, the model can predict breast cancer with a 1.1% to 3.3% error, depending on the number of iterations.

## References

[1] Lu, Wu-Sheng. (May 2019). Experiment 1 - Laboratory Manual ECE 403/504 Optimization for Machine Learning. [Online]. Accessed May 2019.

<https://ece.uvic.ca/~wslu/403/403pass/Trans/LabManual-ECE403-503-2019.pdf>

## Appendix A - MATLAB code

%Initialization%

close all

clear all

clc

load D\_bc\_tr.mat

load D\_bc\_te.mat

global Xtrain

global ytrain

Xtrain = zeros(30,480);

for i = 1:30

xi = D\_bc\_tr(i,:);

mi = mean(xi);

vi = sqrt(var(xi));

Xtrain(i,:) = (xi - mi)/vi;

end

Xtest = zeros(30,89);

for i = 1:30

xi = D\_bc\_te(i,:);

mi = mean(xi);

vi = sqrt(var(xi));

Xtest(i,:) = (xi - mi)/vi;

end

ytrain = D\_bc\_tr(31,:);

ytest = D\_bc\_te(31,:);

w\_zero = zeros(31,1);

[xs1,fsGD, k1] = grad\_desc('f\_hmin','f\_gmin', w\_zero, 3e-9, 5);

[xs2,fsGD, k2] = grad\_desc('f\_hmin','f\_gmin', w\_zero, 3e-9, 12);

[xs3,fsGD, k3] = grad\_desc('f\_hmin','f\_gmin', w\_zero, 3e-9, 75);

result(xs1,Xtest,ytest, k1)

result(xs2,Xtest,ytest, k2)

result(xs3,Xtest,ytest, k3)

Functions:

