

A REPORT ON MINI-PROJECT NEWTON'S METHOD MACHINE LEARNING CS(667)

DIPAK P DULAL
SUBMITTED TO
DR. DA YAN
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ALABAMA AT BIRMINGHAM

INTRODUCTION

Breast cancer is one of the most common cancer in the world and the analysis of the available data to predict and get decision wisely regarding different factors and results on breast cancer is the integral part to help in future to handle cases related to it.

Here, I used Logistic Regression classifier with Newtons method on Wisconsin breast cancer data set to classify the patients as malignant and benign and predict with the model.

Checklists. .

Basically the following library are used and we define some classes to observe the performance:

```
import pandas as pd
import numpy as np
```

Data Preparation. .

We will use the Breast Cancer Wisconsin data set for UCI machine learning repository:
<http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+28Diagnostic29>.
Data File: breast-cancer-wisconsin.data(class2 for benign,4 for malignant).

From the data we excluded the ID and class for input X and the class is for the prediction y(2 for benign, and 4 for malignant). The features are as usual in the given name files. There are total 558 samples and 9 features of input matrix X.

Ideas and Validation Result of the Model. .

Project was simply accomplished through the usage of precomputed final formula for Newton-Raphson method. Firstly the data was extracted and normalized. Afterwards a column was added that reflected constant term corresponding to $x_1 = 1$ restriction. Afterwards function *Newton* was created that implemented learning part as follows: * Start with zero vector * For each loop cycle compute μ and S and use them in $\theta \leftarrow \theta - H^{-1}g = \theta + (X^T S X)^{-1}(X^T(\mu - y))$ * Terminate when θ variance is too small For already trained model a.k.a. θ vector *Predict* function can be called to predict correspondent result. Afterwards *Validate* function performs 80

As we can see, model that we get this way is very accurate and fails 1 or 2 times for ≈ 1 test example in 20 runs..

Finally, we got Average success ratio is 0.9996428571428572

Conclusion. Thus when use the effective model logistic regression with Newton's Method which gives the quite accurate result in less iterations.

The End