

code.fun.do

Team: name.random()

Project: Pink Ribbon

December 2017

Contents

1	Introduction	3
2	Machine Learning Involved:	3
2.1	Logistic Regression	3
2.2	k-nearest-neighbours Algorithm	3
3	How to use?	4

1 Introduction

Pink Ribbon, is a symbol for Breast Cancer, and that is what our Project is about. The user needs to enter the details (explained ahead) and we will tell if the Breast Cancer is malignant or benign. We tested a few Machine Learning and chose the one that gave the best results. The analysis is shown in further sections. Following are the parameters that the user needs to enter:

1. Clump Thickness
2. Uniformity of Cell Size
3. Uniformity of Cell Shape
4. Marginal Adhesion
5. Single Epithelial Cell Size
6. Bare Nuclei
7. Bland Chromatin
8. Normal Nucleoli
9. Mitoses

2 Machine Learning Involved:

We are trying to solve a basic Classification Problem in which the task is to label the input data point as either Malign or Benign. The training data was obtained by the courtesy of UCI Machine Learning Repository. The link for the training data used is <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>. As you can see there are a total of 699 data points. To choose the best model, we trained our model on first 400 points and tested on the remaining 299 points. Finally, we chose the algorithm which gave best results. To deal with the missing points, we replaced them by the average of the remaining points. Following are the analysis of the 2 models we used:

2.1 Logistic Regression

We implemented the Logistic Regression Algorithm and checked it's accuracy. We trained our model on 600 points and then applied it on remaining 99 points. We got accuracy of 92.9292

2.2 k-nearest-neighbours Algorithm

We implemented the k-nearest-neighbours Algorithm and checked it's accuracy. As expected, it outperformed the Linear Logistic Regression Model for various values of k. We finally used $k = 20$, which gave the best results. For 299 test cases, it gave correct results for 296, with an impressive accuracy of 98.9

3 How to use?

As you can see, we have implemented our model and presented it using a simple GUI to make it more user freindly. There are 9 input fields one below other. Users have to choose an integer ranging from 1 to 10 and click on submit button. The model will predict your cancer in the last text field at the bottom.