Project2 :Breast Cancer Prediction Model

Group 7: Melanie Mayer, Yuqi Miao, Sibei Liu, Xue Jin

March 31, 2020

Introduction

Breast cancer is the most common invasive cancer and the second leading cause of cancer death in women worldwide, marked by the uncontrolled growth of breast cells. Non-cancerous breast tumors do not metastasize and are usually not life-threatening, while malignant tumors are cancerous, aggressive and deadly. Therefore, it's important to have breast lumps accurately diagnosed so that decision with regard to medical treatment, rehabilitation and personal matters can be made appropriately.

Objectives

The main objective of this project is to build a predictive model based on binary logistic regression that classifies between malignant and benign cases. Using the Breast Cancer Diagnosis dataset, logistics model and logistic-LASSO model will be implemented to predict the diagnosis. As two important methods for numerical optimization, Newton-Raphson algorithm and Pathwise Coordinate optimization will be developed to estimate the logistic model and the lasso model respectively.

Breast Cancer Diagnosis Dataset

The Breast Cancer Diagnosis dataset contains the diagnosis and a set of 30 features capturing the characteristics of the cell nuclei present in the digitized image of breast mass. Ten features are collected for each cell nucleus:

- radius (mean of distances from center to points on the perimeter)
- texture (standard deviation of gray-scale values)
- perimeter
- area
- smoothness (local variation in radius lengths)
- compactness (perimeter^2 / area - 1.0)
- concavity (severity of concave portions of the contour)
- concave points (number of concave portions of the contour)
- symmetry
- fractal dimension ("coastline approximation" - 1)

The mean, standard error (SE) and largest values of these features are computed for each image, resulting in 30 features, which will be analyzed to understand the predictive value for diagnosis of cancer(malignant) or benign cases.

Methods

Results

Discussion