

Pravin Pidikiti
Benjamin Alterman
CSC 2730
Final Project
6 December 2019

1. For the final project, we considered predicting whether a patient has a certain disease or not based on their symptoms. Through kaggle, we found a dataset that entails the data from Breast Cancer patients predicting whether their breast cancer is benign or malignant based off of the features of the cancer mass.

Attribute Information: ID number and Diagnosis.

Features : a) radius (mean of distances from center to points on the perimeter) b) texture (standard deviation of gray-scale values) c) perimeter d) area e) smoothness (local variation in radius lengths) f) compactness ($\text{perimeter}^2 / \text{area} - 1.0$) g) concavity (severity of concave portions of the contour) h) concave points (number of concave portions of the contour) i) symmetry j) fractal dimension ("coastline approximation" - 1). The mean, standard error and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. All feature values are recoded with four significant digits.

Missing attribute values: none. Class distribution: 357 benign, 212 malignant

Target Prediction: Diagnosis

2. We used LinearRegression from sklearn because all the data was already in numerical data. Parameters were some of the features from the data.
- 3.

Parameter:	Mean Squared Error:
Radius Mean	0.027295478663477157
Texture Mean	0.05071584085645591
Perimeter Mean	0.38127675508736364
Area Mean	0.15539934330259977
Smoothness Mean	7955.383595647786
All Parameters (31)	0.013268685876730154

4. In the predictors from the dataset used for the prediction were the mean of the following: radius, texture, perimeter, area, and smoothness size. Every feature except smoothness mean gave a mean squared error of 0.1, indicating they are good predictors to predict whether the cancer is benign or malignant. The smoothness feature gave a mean squared error off by approximately 1000 due to the values being extremely small. This is indicating smoothness mean is bad predictor