CMP 466 – Machine Learning and Data Mining- Spring 2021

Team Project- Assignment 0 Due: Feb 11th, 2020

1. Farooq Mirza 80205

Omar Fayed 69702

Hamad Almaqoodi 76307

2. Breast Cancer type Prediction

3. Dataset: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>

Prediction of type of breast cancer (Benign or Malignant) using pre-existing data from patients in Wisconsin. Using this data, we can predict the type of breast cancer for patients that have not been diagnosed yet. We want to predict the type to make the process faster for doctors and patients to give them the appropriate treatment.

4. Type of data is previous records of patients

569 different cases of 357 benign, 212 malignant

Ten real-valued features are computed for each cell nucleus:

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 (perimeter^2 / 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)

All features are numbers within their specific ranges.

The labels are either Benign or Malignant.