CEBD 1260 Final Programming Assignment

Part 1

1) First task to is to classify data from a cancer diagnostic database. In this database are patients with tumors, characteristics of those tumors, and biospy results indicating whether the tumor is Malignant or Benign.

About the dataset

In cancer_data.txt you will find the following variables:

- 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)
- cancer (0 = Benign, 1 = Malignant) *target*

Use any machine learning algorithm you wish. In your answer include a short description of your algorithm of choice and predicted category of a new patient with a tumor with the following features:

- radius: 14
- texture: 14
- perimeter: 88
- area: 566
- smoothness: 1
- compactness: 0.08
- concavity: 0.06
- concae points: 0.04
- symmetry: 0.18
- fractal dimension: 0.05

In [ ]:

Part 2