

I would like to observe the relationship between the different types of breast cancer; malignant and benign. In this dataset, they are identified as 0 for malignant and 1 for benign as a class attribute. In the 1st figure, you can see that there are 212 cases of malignant cancer and 357 of benign cancer in this dataset. There are ten main attributes but thirty total due to the fact that mean, standard error and worst or largest of the features were computed as well. I will only look at the mean section.

I did three exploratory analysis on the dataset. For all of them, the usefulness of the relationship we can see the difference. The first one was about the different size of the cell. I look at the radius, perimeter and area between the two classes. The usefulness of this relationship is that we can see the difference in size. The radius looked at the mean of distance from center to points on the perimeter. As the figures show, the smaller the radius, perimeter and area is, the class is more likely to be 1 which is benign. The higher the numbers go, it is class 0 so the larger sizes are malignant type. For class 0, the numbers are more spread out while for class 1, it is more concentrated.

Second one was about the difference inside of the cell. I looked at the texture, smoothness, and compactness. The texture is the standard deviation of gray scale values. Smoothness is the local variation in radius lengths. Compactness is squared perimeter divided by area subtract 1.0. In this set, one can see that there is not much difference between texture and smoothness between the two classes. They are similar but with compactness, one can see the difference in range of numbers. While class 0 is more spread out from low to high but class 1 is more concentrated in the lower section.

The third one was about the difference in shapes. I looked at the symmetry, fractal dimension and compactness. The fractal dimension is the 'coastline approximation' subtract 1. The symmetry and fractal dimension did not show much difference between the two classes. For symmetry, both of them were spread out although class 1 in some case was smaller. With fractal dimension, class 1 was spread out but class 0 was concentrated in the lower numbers.

The most informative for future analysis is the first exploratory analysis. It clearly showed the difference between the two classes. I would recommend that for the most apparent identification between the two classes would be to look at the difference in sizes.